Annals of the Institute of Statistical Mathematics (2020) 72:1371–1396
https://doi.org/10.1007/s10463-019-00728-0

Robust estimation for general integer-valued time series models

Byungsoo Kim[1] · Sangyeol Lee[2]

Received: 17 October 2018 / Revised: 8 April 2019 / Published online: 22 July 2019

Abstract
In this study, we consider a robust estimation method for general integer-valued time series models whose conditional distribution belongs to the one-parameter exponential family. As a robust estimator, we employ the minimum density power divergence estimator, and we demonstrate this is strongly consistent and asymptotically normal under certain regularity conditions. A simulation study is carried out to evaluate the performance of the proposed estimator. A real data analysis using the return times of extreme events of the Goldman Sachs Group stock is also provided as an illustration.

Keywords Robust estimation · Minimum density power divergence estimator · General integer-valued time series · One-parameter exponential family · INGARCH models

1 Introduction

In recent years, integer-valued time series models have received considerable attention from researchers in diverse research areas. Since the work of McKenzie (1985) and Al-Osh and Alzaid (1987), integer-valued autoregressive (INAR) models based on a binomial thinning operator have been widely employed to analyze correlated time series of counts. See Weiß (2008) for a review. Although INAR models are useful in many cases, the equidispersion property that arises in the INAR model with Poisson innovations can lead to a serious problem, because many real datasets exhibit over-dispersion. To remedy this, Ferland et al. (2006) proposed to use Poisson integer-valued generalized autoregressive conditional heteroscedasticity (INGARCH) models, and

✉ Sangyeol Lee
sylee@stats.snu.ac.kr

Byungsoo Kim
bkim@yu.ac.kr

1 Department of Statistics, Yeungnam University, Kyungsan 38541, Korea

2 Department of Statistics, Seoul National University, Seoul 08826, Korea

later, Fokianos et al. (2009) developed Poisson autoregressive (Poisson AR) models, including nonlinear specifications for their intensity processes. These models not only merit keeping the Poisson distribution as their underlying distributions but also capture the over-dispersion phenomenon effectively.

Researchers invested considerable efforts to relax the Poisson assumption in INGARCH models and extended the Poisson INGARCH model to other distributional models. Examples include negative binomial INGARCH (NB-INGARCH) models (Davis and Wu 2009; Christou and Fokianos 2014) and zero-inflated generalized Poisson INGARCH models (Zhu 2012a, b; Lee et al. 2016). Davis and Liu (2016) recently considered one-parameter exponential family AR models, called general integer-valued time series models. Diop and Kengne (2017) and Lee and Lee (2018) then utilized this framework to handle the problem of detecting a change point. In these articles, the conditional maximum likelihood estimator (CMLE) is employed for parameter estimation. However, the CMLE is sensitive to outliers and a model bias when an exponential family AR model is fitted to datasets. Thus, robust estimation in general integer-valued time series models is crucial and deserves a special investigation.

In this study, we adopt the minimum density power divergence estimator (MDPDE), proposed by Basu et al. (1998) (BHHJ), as a robust estimator, because this method is well known to consistently offer robust estimators in various situations and enjoys both efficiency and robustness as a trade-off through controlling the tuning parameter in the MDPDE. For example, see Mihoko and Eguchi (2002), who used the density power divergence to recover the original independent signals when their linear mixtures were observed. Later, Lee and Song (2009), Kim and Lee (2013), Kang and Lee (2014a), and Kim and Lee (2017) studied the MDPDE for GARCH models, the covariance matrix of multivariate times series, Poisson AR models, and zero-inflated Poisson AR models, respectively. In those studies, the MDPDE was proven to have strong robust properties, with little loss in asymptotic efficiency relative to the MLE. Our analysis also confirms the same conclusion for general integer-valued time series models.

The organization of this paper is as follows: Sect. 2 constructs the MDPDE for general integer-valued time series models. Section 3 demonstrates the asymptotic properties of the MDPDE. Section 4 presents a simulation study and real data analysis. Finally, Sect. 5 provides a conclusion. Proofs are provided in "Appendix."

2 MDPDE for general integer-valued time series models

BHHJ proposed the density power divergence d_α between two density functions f and g as

$$d_\alpha(g,f) := \begin{cases} \int \{f^{1+\alpha}(y) - \left(1+\frac{1}{\alpha}\right) g(y) f^\alpha(y) + \frac{1}{\alpha} g^{1+\alpha}(y)\}\mathrm{d}y, & \alpha > 0, \\ \int g(y)(\log g(y) - \log f(y))\mathrm{d}y, & \alpha = 0. \end{cases}$$

For a parametric family $\{F_\theta,\ \theta \in \Theta\}$ possessing densities $\{f_\theta\}$ and a distribution G with density g, they defined the minimum density power divergence functional $T_\alpha(G)$ by $d_\alpha(g, f_{T_\alpha(G)}) = \min_{\theta\in\Theta} d_\alpha(g, f_\theta)$. In particular, if $G = F_{\theta_0} \in \{F_\theta\}$, then

$T_\alpha(F_{\theta_0}) = \theta_0$. Based on the above, given a random sample $Y_1, \ldots, Y_n$ with unknown density g, the MDPDE is defined by

$$\hat{\theta}_{\alpha,n}^{\text{MDPDE}} = \underset{\theta \in \Theta}{\operatorname{argmin}}\, H_{\alpha,n}(\theta),$$

where $H_{\alpha,n}(\theta) = \frac{1}{n}\sum_{t=1}^{n} V_\alpha(\theta; Y_t)$ and

$$V_\alpha(\theta; Y_t) = \begin{cases} \int f_\theta^{1+\alpha}(y)\mathrm{d}y - \left(1 + \frac{1}{\alpha}\right) f_\theta^\alpha(Y_t), & \alpha > 0, \\ -\log f_\theta(Y_t), & \alpha = 0. \end{cases}$$

When $\alpha = 0$ and 1, the MDPDE is the same as the MLE and L_2-distance estimator, respectively. BHHJ showed that $\hat{\theta}_{\alpha,n}^{\text{MDPDE}}$ is consistent for $T_\alpha(G)$ and asymptotically normal, and demonstrated that the estimator is robust against outliers, but still retains high efficiency when the true distribution belongs to a parametric family $\{F_\theta\}$ and α is close to zero.

To apply the BHHJ's procedure to general integer-valued time series models, we require the conditional version of the MDPDE. Let $\{g_\theta(\cdot|\mathcal{F}_{t-1})\}$ be the parametric family of autoregressive models indexed by the parameter θ, and let $g_{\theta_0}(\cdot|\mathcal{F}_{t-1})$ be the true conditional density of the time series Y_t given $\mathcal{F}_{t-1}$, where $\mathcal{F}_{t-1}$ is a σ-field generated by $Y_{t-1}, Y_{t-2}, \ldots$. Then, the MDPDE is defined by

$$\hat{\theta}_{\alpha,n}^{\text{MDPDE}} = \underset{\theta \in \Theta}{\operatorname{argmin}}\, H_{\alpha,n}(\theta),$$

where $H_{\alpha,n}(\theta) = \frac{1}{n}\sum_{t=1}^{n} V_\alpha(\theta; \mathcal{F}_{t-1}, Y_t)$ and

$$V_\alpha(\theta; \mathcal{F}_{t-1}, Y_t) = \begin{cases} \int g_\theta^{1+\alpha}(y|\mathcal{F}_{t-1})\mathrm{d}y - \left(1 + \frac{1}{\alpha}\right) g_\theta^\alpha(Y_t|\mathcal{F}_{t-1}), & \alpha > 0, \\ -\log g_\theta(Y_t|\mathcal{F}_{t-1}), & \alpha = 0 \end{cases} \tag{1}$$

(cf. Section 2 of Kang and Lee 2014a).

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

$$Y_t|\mathcal{F}_{t-1} \sim p(y|\eta_t), \quad X_t := E(Y_t|\mathcal{F}_{t-1}) = f_\theta(X_{t-1}, Y_{t-1}), \tag{2}$$

where $f_\theta(x, y)$ is a nonnegative bivariate function defined on $[0, \infty) \times \mathbb{N}_0$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, depending on the parameter $\theta \in \Theta \subset \mathbb{R}^d$, which satisfies $\inf_{\theta \in \Theta} f_\theta(x, y) \geq x^*$ for some $x^* > 0$ and 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 contains many famous discrete distributions, including the Poisson, negative binomial, and binomial distributions.

We set $B(\eta) = A'(\eta)$. Then, $B(\eta_t)$ and $B'(\eta_t)$ are the conditional mean and variance of Y_t, respectively, and $X_t = B(\eta_t)$. The derivative of $A(\eta)$ generally exists for the exponential family (cf. Lehmann and Casella 1998). We note that since $B'(\eta_t) = Var(Y_t|\mathcal{F}_{t-1}) > 0$, $B(\eta)$ is strictly increasing. Because the support of Y_t is assumed to include nonnegative integers, we have $B(\eta_t) = E(Y_t|\mathcal{F}_{t-1}) > 0$, which implies that $A(\eta)$ is also strictly increasing. To emphasize the role of θ, we also use notation $X_t(\theta)$ and $\eta_t(\theta) = B^{-1}(X_t(\theta))$ to denote X_t and η_t.

Davis and Liu (2016) showed that the following assumption 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 $\omega_1, \omega_2 \geq 0$ satisfies $\omega_1 + \omega_2 < 1$.

Furthermore, they demonstrated that the conditional mean X_t can be expressed as a function of the observations Y_t. That is, there exists a measurable function $f_\infty^\theta : \mathbb{N}_0^\infty \to [0, \infty)$ such that $X_t(\theta) = f_\infty^\theta(Y_{t-1}, Y_{t-2}, \ldots)$ a.s.

Given $Y_1, \ldots, Y_n$ generated by (2), from (1), we define the MDPDE $\hat{\theta}_{\alpha,n}$ for general integer-valued time series models as

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

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 recursively updated through the following equations:

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

with an arbitrarily chosen initial value $\widetilde{X}_1$. From (3), we can see that the MDPDE with $\alpha = 0$ corresponds to the CMLE.

3 Asymptotic properties of the MDPDE

3.1 Consistency and asymptotic normality of the MDPDE

In this subsection, we verify the consistency and asymptotic normality of the MDPDE under the regularity conditions given below. Some of these are found in Lee and Lee (2018), whereas others are newly considered to handle the MDPDE. Throughout this study, V and $\rho \in (0, 1)$ denote a generic integrable random variable and a constant, respectively; the symbol $\|\cdot\|$ denotes the L^1 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) $\inf_{\theta\in\Theta}\inf_{0\le\delta\le 1} B'((1-\delta)\eta_t(\theta)+\delta\tilde{\eta}_t(\theta)) \ge \underline{c}$ for some $\underline{c} > 0$.
(A4) $EY_1^4 < \infty$.
(A5) If there exists $t \ge 1$ such that $X_t(\theta) = X_t(\theta_0)$ a.s., then $\theta = \theta_0$.
(A6) $\sup_{\theta\in\Theta}\sup_{0\le\delta\le 1}\left|\frac{B''((1-\delta)\eta_t(\theta)+\delta\tilde{\eta}_t(\theta))}{B'((1-\delta)\eta_t(\theta)+\delta\tilde{\eta}_t(\theta))^3}\right| \le K$ for some $K > 0$.
(A7) The mapping $\theta \mapsto f_\infty^\theta$ is twice continuously differentiable with respect to θ and satisfies

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

(A8) $\sup_{\theta\in\Theta}\left\|\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta} - \frac{\partial X_t(\theta)}{\partial\theta}\right\| \le V\rho^t$ a.s.
(A9) $\nu^{\mathrm{T}}\frac{\partial X_t(\theta_0)}{\partial\theta} = 0$ a.s. implies $\nu = 0$.

Remark 1 The conditions similar to **(A1)–(A9)** can be found in many studies (cf. Kang and Lee 2014a, b; Cui and Zheng 2017; Lee and Lee 2018). **(A2)** and **(A7)** are conditions related to the moments of conditional mean of Y_t and its first and second derivatives, which are essential for proving the asymptotic properties of the MDPDE. Certain integer-valued time series models belonging to general integer-valued time series models require additional assumption on parameters to satisfy **(A4)**. As an example, the case of NB-INGARCH model is provided in Sect. 3.2 below. In practice, since the past history of observations and conditional mean process is unknown, **(A8)** is imposed to approximate $\partial X_t(\theta)/\partial\theta$ by $\partial\widetilde{X}_t(\theta)/\partial\theta$ exponentially fast, which is needed to prove Lemma 6 in "Appendix." A group of most popular INGARCH models, such as Poisson, negative binomial, and binomial INGARCH models, satisfies **(A1)–(A9)** as seen in Sect. 3.2.

Under these conditions, we obtain the following asymptotic results whose proofs are provided in "Appendix."

Theorem 1 *Suppose that conditions* **(A0)–(A5)** *hold. Then,*

$$\hat{\theta}_{\alpha,n} \xrightarrow{a.s.} \theta_0 \ \ as \ \ n\to\infty.$$

Theorem 2 *Suppose that conditions* **(A0)–(A9)** *hold. Then,*

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

where

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

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

Remark 2 The tuning parameter α controls the trade-off between the robustness and asymptotic efficiency. That is, adopting a relatively large α is favorable when the robustness is more emphasized, whereas a small α is suitable if the efficiency is the primary concern. Fujisawa and Eguchi (2006), Toma and Broniatowski (2011), and Durio and Isaia (2011) developed a procedure for choosing an optimal α. Here, we adopt the method of Warwick (2005) and Warwick and Jones (2005) to choose α that minimizes the trace of the estimated asymptotic mean squared error ($\widehat{\text{AMSE}}$):

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

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

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

This criterion is applied to our empirical study in Sect. 4.2.

3.2 INGARCH models

As an example, we consider the INGARCH(1,1) models defined by

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

where $X_t = B(\eta_t) = E(Y_t|\mathcal{F}_{t-1})$, $\theta = (d, a, b)^{\mathrm{T}} \in \Theta \subset (0, \infty) \times [0, \infty)^2$ with $a + b < 1$, and Θ is compact. Then, condition **(A0)** holds, and thus, the process $\{(X_t, Y_t),\ t \geq 1\}$ admits a strictly stationary and ergodic solution. To obtain the asymptotic results for the INGARCH(1,1) models, **(A1)** is replaced by the following:

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

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

Moreover, by recursion, we have

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

where the initial value $\widetilde{X}_1$ is taken as $d/(1-a)$ for simplicity. Based on the above and **(A4)**, conditions **(A2), (A5)**, and **(A7)–(A9)** are satisfied for the INGARCH(1,1) models, the proof of which is essentially the same as that of Theorem 3 in Kang and Lee (2014b). Because sufficient conditions for **(A4)** would depend on the conditional distribution of Y_t, we provide those conditions for Poisson, negative binomial, and binomial INGARCH(1,1) models. We also show that **(A3)** and **(A6)** hold for these models.

1. Consider the Poisson INGARCH(1,1) model given by

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

Owing to Proposition 1 and Remark 2 of Cui and Zheng (2017), under **(A0)**, Y_t has finite moments of any order. Hence, **(A4)** holds. In this model, $\eta_t(\theta) = \log(X_t(\theta))$ and $A(\eta_t(\theta)) = \mathrm{e}^{\eta_t(\theta)}$. Since $B'(\eta_t(\theta)) = X_t(\theta) \geq d_L$, $B'(\tilde{\eta}_t(\theta)) = \widetilde{X}_t(\theta) \geq d_L$, and $B'(\eta) = \mathrm{e}^{\eta}$ is strictly increasing, **(A3)** holds. Finally, **(A6)** is satisfied owing to **(A3)** and the fact that $B'(\eta) = B''(\eta)$.

2. Consider the NB-INGARCH(1,1) model, defined by

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

where $\text{NB}(r, p)$ denotes the negative binomial distribution with parameters $r \in \mathbb{N}$ and $p \in (0, 1)$; namely, it counts the number of failures in a sequence of Bernoulli trials with the success probability p before the rth success occurs. This model is considered by assuming that r is fixed and supposed to be known. Unlike in the Poisson INGARCH(1,1) model, Y_t does not have finite moments of any order under **(A0)** (see Proposition 1 and Remark 3 of Cui and Zheng (2017) for more details). In particular, Ahmad and Francq (2016) revealed that $EY_t^4 < \infty$ if and only if $(a+b)^4 + 6b^2(a+b)^2/r + b^3(8a+11b)/r^2 + 6b^4/r^3 < 1$. Hence, **(A4)** holds if the latter condition is satisfied. In this model, $\eta_t(\theta) = \log(X_t(\theta)/(X_t(\theta)+r))$ and $A(\eta_t(\theta)) = r\log(r/(1-\mathrm{e}^{\eta_t(\theta)}))$. Since $B'(\eta_t(\theta)) = r\mathrm{e}^{\eta_t(\theta)}/(1-\mathrm{e}^{\eta_t(\theta)})^2 = X_t(\theta)(X_t(\theta)+r)/r \geq d_L(d_L+r)/r$, $B'(\tilde{\eta}_t(\theta)) = \widetilde{X}_t(\theta)(\widetilde{X}_t(\theta)+r)/r \geq d_L(d_L+r)/r$, and $B'(\eta) = r\mathrm{e}^{\eta}/(1-\mathrm{e}^{\eta})^2$ is strictly increasing on $\eta < 0$, it follows that **(A3)** holds. Next, using the fact that $B''(\eta) = r\mathrm{e}^{\eta}(1+\mathrm{e}^{\eta})/(1-\mathrm{e}^{\eta})^3$, we have $B''(\eta)/B'(\eta)^3 = (1-\mathrm{e}^{\eta})^3(1+\mathrm{e}^{\eta})/r^2\mathrm{e}^{2\eta}$ and it is positive and strictly decreasing on $\eta < 0$. Furthermore, owing to the fact that $d_L/(d_L+r) \leq \mathrm{e}^{\eta_t(\theta)} < 1$, it holds

$$\frac{B''(\eta_t(\theta))}{B'(\eta_t(\theta))^3} = \frac{1}{B'(\eta_t(\theta))^{\frac{3}{2}}} \frac{1+\mathrm{e}^{\eta_t(\theta)}}{(r\mathrm{e}^{\eta_t(\theta)})^{\frac{1}{2}}} \leq \frac{r^{\frac{3}{2}}}{(d_L(d_L+r))^{\frac{3}{2}}} \frac{2(d_L+r)^{\frac{1}{2}}}{(rd_L)^{\frac{1}{2}}} = \frac{2r}{d_L^2(d_L+r)},$$

and $B''(\tilde{\eta}_t(\theta))/B'(\tilde{\eta}_t(\theta))^3$ also has the same upper bound. Therefore, **(A6)** is satisfied.

3. Consider the binomial INGARCH(1,1) model given by

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

where $d > 0, a \geq 0, b \geq 0$, $d + am + bm \leq m$ is assumed to ensure that $p_t \in (0, 1)$, and m is supposed to be known. Hence, the parameter space Θ in **(A1)**$'$ for the binomial INGARCH(1,1) model becomes

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

From Proposition 1 and Remark 3 of Cui and Zheng (2017), Y_t has finite moments of any order under **(A0)**. Thus, **(A4)** holds. In this model, $\eta_t(\theta) = \log(X_t(\theta)/(m - X_t(\theta)))$ and $A(\eta_t(\theta)) = m\log(1 + e^{\eta_t(\theta)})$. Since $X_t(\theta) \le d + m(a+b) \le d_U + m(1-\epsilon)$, we have that $B'(\eta_t(\theta)) = me^{\eta_t(\theta)}/(1+e^{\eta_t(\theta)})^2 = X_t(\theta)(m - X_t(\theta))/m \ge d_L(\epsilon - d_U/m)$, and similarly $B'(\tilde{\eta}_t(\theta)) \ge d_L(\epsilon - d_U/m)$. Because $B'(\eta)$ is strictly increasing on $\eta \le 0$ and strictly decreasing on $\eta \ge 0$, it holds $B'((1-\delta)\eta_t(\theta) + \delta\tilde{\eta}_t(\theta)) \ge \min(B'(\eta_t(\theta)), B'(\tilde{\eta}_t(\theta))) \ge d_L(\epsilon - d_U/m)$, and thus **(A3)** is satisfied. Note that

$$\left|\frac{B''(\eta_t(\theta))}{B'(\eta_t(\theta))^3}\right| = \left|\frac{(1+e^{\eta_t(\theta)})^3(1-e^{\eta_t(\theta)})}{(me^{\eta_t(\theta)})^2}\right| \le \frac{(1+e^{\eta_t(\theta)})^4}{(me^{\eta_t(\theta)})^2} = \frac{1}{B'(\eta_t(\theta))^2} \le \frac{1}{d_L^2(\epsilon - d_U/m)^2},$$

and $|B''(\tilde{\eta}_t(\theta))/B'(\tilde{\eta}_t(\theta))^3|$ also has the same upper bound. Owing to the fact that $|B''(\eta)/B'(\eta)^3|$ is strictly decreasing on $\eta \le 0$ and strictly increasing on $\eta \ge 0$, we have $|B''((1-\delta)\eta_t(\theta) + \delta\tilde{\eta}_t(\theta))/B'((1-\delta)\eta_t(\theta) + \delta\tilde{\eta}_t(\theta))^3| \le \max(|B''(\eta_t(\theta))/B'(\eta_t(\theta))^3|, |B''(\tilde{\eta}_t(\theta))/B'(\tilde{\eta}_t(\theta))^3|) \le 1/(d_L^2(\epsilon - d_U/m)^2)$. Therefore, **(A6)** is established.

Remark 3 One may consider nonlinear models such as the threshold Poisson autoregressive model (INTGARCH(1,1)) studied by Doukhan and Kengne (2015) and Diop and Kengne (2017). The INTGARCH(1,1) model is defined by

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

where $\theta = (d, a, b_1, b_2)^{\mathrm{T}}$ belongs to a compact subset Θ of $(0,\infty) \times [0,\infty)^3$ and satisfies $a + \max(b_1, b_2) < 1$, and l is a nonnegative integer value, called the threshold parameter of the model. In this model, **(A0)** is satisfied. The parameter space Θ is given by

$$\Theta = \{\theta = (d, a, b_1, b_2)^{\mathrm{T}} \in \mathbb{R}_+^4 : 0 < d_L \le d \le d_U,\ \epsilon \le a + \max(b_1, b_2) \le 1 - \epsilon\}$$

for some $d_L, d_U, \epsilon > 0$. Similar to the INGARCH(1,1) model, we can obtain

$$X_t(\theta) = \frac{d}{1-a} + \sum_{k=0}^{\infty} a^k\left(b_1\max(Y_{t-k-1} - l, 0) + b_2\min(Y_{t-k-1}, l)\right)$$

and

$$\widetilde{X}_t(\theta) = \frac{d}{1-a} + \sum_{k=0}^{t-2} a^k\left(b_1\max(Y_{t-k-1} - l, 0) + b_2\min(Y_{t-k-1}, l)\right),$$

where $\widetilde{X}_1$ is taken as $d/(1-a)$. Based on the above, the conditions in Sect. 3.1 can be verified by following the arguments similar to those for the INGARCH(1,1) models.

Table 1 Sample mean (variance × 10^2/MSE × 10^2) of estimators for the NB-INGARCH(1,1) model when no outliers exist

α	$\hat{d}$	$\hat{a}$	$\hat{b}$
MDPDE			
0 (CMLE)	1.079(13.22/13.82)*	0.367(2.675/2.780)*	0.201(0.218/0.218)*
0.1	1.077(13.57/14.15)	0.368(2.740/2.839)	0.201(0.224/0.224)
0.2	1.075(14.05/14.60)	0.369(2.832/2.926)	0.201(0.233/0.233)
0.3	1.074(14.73/15.26)	0.370(2.974/3.064)	0.201(0.247/0.247)
0.4	1.074(15.30/15.83)	0.370(3.091/3.181)	0.201(0.259/0.259)
0.5	1.074(16.05/16.58)	0.369(3.240/3.330)	0.201(0.273/0.273)
0.75	1.074(17.30/17.82)	0.369(3.486/3.577)	0.202(0.304/0.304)
1	1.074(18.46/19.00)	0.369(3.725/3.820)	0.202(0.336/0.336)

4 Empirical studies

4.1 Simulation

In this subsection, we compare the performance of the MDPDE ($\alpha > 0$) with that of the CMLE ($\alpha = 0$). To this end, we employ the NB-INGARCH(1,1) and INTGARCH(1,1) models. Simulation results for the Poisson INGARCH(1,1) model can be found in Kang and Lee (2014a). First, we consider the NB-INGARCH(1,1) model:

$$Y_t|\mathcal{F}_{t-1} \sim \text{NB}(r, p_t), \quad X_t = \frac{r(1-p_t)}{p_t} = d + aX_{t-1} + bY_{t-1}, \tag{4}$$

where X_1 is assumed to be 0 for the data generation and $\widetilde{X}_1$ is set to be the sample mean of the data. In this model, we consider the case with $r = 10$ and $\theta = (d, a, b)^{\mathrm{T}} = (1, 0.4, 0.2)^{\mathrm{T}}$. For comparison, we investigate the sample mean, variance, and mean squared error (MSE) of the estimators. The sample size under consideration is $n = 500$, and the number of repetitions for each simulation is 1000.

Table 1 presents the results when the data are not contaminated by outliers. In Tables 1, 2, 3, 4, 5, 6, and 7, the symbol * stands for the minimal MSE and the italic values represent MDPDEs with smaller MSEs than the CMLE. Table 1 shows that when the data are not contaminated by outliers, the CMLE has minimal MSEs for all parameters and the MSEs of the MDPDEs with small α are similar to those of the CMLE. As α increases, the MSEs of the MDPDEs also increase, which confirms that an MDPDE with large α results in a loss of efficiency and that the CMLE outperforms the MDPDE when no outliers are present in the data.

Next, to evaluate the robustness of the estimators, we observe the contaminated data $Y_{c,t}$ as follows (cf. Fried et al. 2015):

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

where Y_t are generated from (4), P_t are iid Bernoulli random variables with success probability p, and $Y_{o,t}$ are iid NB$(10, \kappa)$ random variables. We assume that Y_t, P_t, and

Table 2 Sample mean (variance × 10^2/MSE × 10^2) of estimators for the NB-INGARCH(1,1) model when $p = 0.01$

α	$\kappa = 0.5$		
	$\hat{d}$	$\hat{a}$	$\hat{b}$
MDPDE			
0 (CMLE)	1.177(22.29/25.40)	0.376(3.979/4.031)	0.172(0.290/0.371)
0.1	*1.135(18.87/20.68)*	*0.377(3.544/3.593)**	*0.174(0.260/0.328)**
0.2	*1.122(18.88/20.35)**	*0.376(3.607/3.660)*	*0.175(0.267/0.330)*
0.3	*1.117(19.27/20.62)*	*0.375(3.727/3.784)*	*0.176(0.283/0.342)*
0.4	*1.119(19.87/21.28)*	*0.372(3.845/3.917)*	*0.177(0.295/0.349)*
0.5	*1.121(20.40/21.84)*	0.370(3.968/4.053)	*0.178(0.310/0.359)*
0.75	*1.123(21.23/22.74)*	0.366(4.216/4.326)	0.181(0.352/0.389)
1	*1.128(22.32/23.93)*	0.362(4.441/4.582)	0.184(0.389/0.415)
α	$\kappa = 0.4$		
	$\hat{d}$	$\hat{a}$	$\hat{b}$
MDPDE			
0 (CMLE)	1.239(28.92/34.59)	0.377(4.909/4.956)	0.155(0.364/0.565)
0.1	*1.185(20.97/24.36)*	*0.369(3.849/3.939)**	*0.158(0.264/0.437)*
0.2	*1.183(20.93/24.26)**	*0.363(3.853/3.988)*	*0.160(0.264/0.425)**
0.3	*1.182(20.99/24.28)*	*0.360(3.906/4.063)*	*0.161(0.277/0.426)*
0.4	*1.183(21.58/24.92)*	*0.357(4.047/4.226)*	*0.163(0.296/0.431)*
0.5	*1.183(21.77/25.10)*	*0.355(4.135/4.329)*	*0.165(0.318/0.438)*
0.75	*1.192(22.91/26.58)*	*0.347(4.399/4.680)*	*0.171(0.363/0.446)*
1	*1.197(23.44/27.30)*	*0.340(4.593/4.951)*	*0.177(0.410/0.463)*

$Y_{o,t}$ are all independent. In this simulation, we consider the cases of $p = 0.01,\ 0.03$ and $\kappa = 0.5,\ 0.4$. Tables 2 and 3 present the results when $p = 0.01$ and 0.03, respectively. In most cases, the MDPDE outperforms the CMLE; that is, the MDPDE has smaller MSEs than the CMLE. As p increases or κ decreases, the MSE of the CMLE increases to a greater extent than that of the MDPDE in the italic values. This indicates that as the data get more contaminated by outliers, the MDPDE performs better than the CMLE. Moreover, when either p increases or κ decreases, the italic values become wider and the symbol * tends to move downward, which indicates that if the data are severely contaminated by outliers, an MDPDE with large α performs better. Note that the symbol * for $\hat{b}$ moves significantly further downward than those for $\hat{d}$ and $\hat{a}$. When $p = 0.03$ and $\kappa = 0.4$, the smallest MSE of $\hat{b}$ is achieved at $\alpha = 1$. This result indicates that outliers damage $\hat{b}$ more severely than $\hat{d}$ and $\hat{a}$.

In Table 4, we additionally investigate the performance of the parameter estimators by considering the cases of $p = 0$ (no outliers) and $p = 0.01,\ \kappa = 0.4$, based on the asymptotic variance $\widehat{\text{As.var}}(\hat{\theta}_{\alpha,n})$ in Remark 2. Here, we use $n = 2000$, because this consistently yields the estimated asymptotic variance stably, and compare the averaged estimated asymptotic variances (AEAVs) calculated from the 1000 values

Table 3 Sample mean (variance $\times 10^2$/MSE $\times 10^2$) of estimators for the NB-INGARCH(1,1) model when $p = 0.03$

α	$\kappa = 0.5$		
	$\hat{d}$	$\hat{a}$	$\hat{b}$
MDPDE			
0 (CMLE)	1.341(38.64/50.21)	0.394(5.857/5.855)	0.128(0.320/0.839)
0.1	*1.231(27.88/33.21)*	*0.394(4.737/4.737)*	*0.133(0.254/0.701)*
0.2	*1.186(26.04/29.49)*	*0.395(4.681/4.679)**	*0.134(0.254/0.684)*
0.3	*1.175(26.13/29.18)**	*0.392(4.816/4.818)*	*0.135(0.263/0.682)**
0.4	*1.172(26.61/29.54)*	*0.389(4.972/4.980)*	*0.136(0.277/0.683)*
0.5	*1.173(27.29/30.25)*	*0.385(5.145/5.162)*	*0.138(0.296/0.684)*
0.75	*1.175(28.15/31.17)*	*0.379(5.488/5.529)*	*0.142(0.353/0.686)*
1	*1.176(27.95/31.02)*	*0.373(5.595/5.660)*	*0.148(0.417/0.691)*
α	$\kappa = 0.4$		
	$\hat{d}$	$\hat{a}$	$\hat{b}$
MDPDE			
0 (CMLE)	1.422(53.79/71.58)	0.421(7.662/7.699)	0.096(0.369/1.454)
0.1	*1.303(38.00/47.14)*	*0.395(6.157/6.154)*	*0.099(0.222/1.243)*
0.2	*1.267(34.89/41.98)*	*0.390(5.898/5.902)**	*0.099(0.211/1.232)*
0.3	*1.257(34.37/40.95)**	*0.388(5.960/5.969)*	*0.099(0.222/1.233)*
0.4	*1.258(34.34/40.96)*	*0.384(6.036/6.057)*	*0.101(0.239/1.221)*
0.5	*1.266(34.83/41.84)*	*0.378(6.151/6.195)*	*0.103(0.264/1.203)*
0.75	*1.284(35.05/43.09)*	*0.362(6.360/6.495)*	*0.111(0.345/1.134)*
1	*1.287(34.32/42.54)*	*0.353(6.461/6.677)*	*0.121(0.445/1.070)**

of the parameter estimates. The upper table of Table 4 ($p = 0$) shows that the CMLE produces minimal AEAVs, and the AEAV increases as α increases as anticipated. On the other hand, the lower table of Table 4 ($p = 0.01,\ \kappa = 0.4$) shows that the AEAVs of the MDPDE are smaller than those of the CMLE, except for $\hat{b}$ with $\alpha = 1$. All these results confirm that the MDPDE outperforms the CMLE in terms of efficiency when the data are contaminated by outliers.

Now, we consider the INTGARCH(1,1) model:

$$Y_t|\mathcal{F}_{t-1} \sim \text{Poisson}(X_t), \quad X_t = d + aX_{t-1} + b_1 \max(Y_{t-1} - l, 0) + b_2 \min(Y_{t-1}, l), \tag{5}$$

where X_1 and $\widetilde{X}_1$ are 0 and the sample mean of data, respectively. We set $l = 3$ and $\theta = (d, a, b_1, b_2)^{\mathrm{T}} = (1, 0.4, 0.1, 0.3)^{\mathrm{T}}$. To compare the robustness of the estimators, we consider the contaminated data $Y_{c,t}$ as

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

Table 4 Sample mean (variance $\times 10^2$/MSE $\times 10^2$/AEAV $\times 10^2$) of estimators for the NB-INGARCH(1,1) model when $p = 0$ and $p = 0.01$, $\kappa = 0.4$

α	$p = 0$ (no outliers exist)		
	$\hat{d}$	$\hat{a}$	$\hat{b}$
MDPDE			
0 (CMLE)	1.017(3.193/3.220/3.373)*	0.394(0.659/0.662/0.702)*	0.199(0.053/0.053/0.053)*
0.1	1.018(3.226/3.254/3.440)	0.394(0.666/0.669/0.716)	0.199(0.054/0.054/0.054)
0.2	1.018(3.329/3.359/3.579)	0.393(0.687/0.690/0.745)	0.199(0.055/0.055/0.056)
0.3	1.019(3.469/3.501/3.757)	0.393(0.715/0.719/0.782)	0.199(0.058/0.058/0.058)
0.4	1.019(3.628/3.663/3.961)	0.393(0.747/0.752/0.824)	0.199(0.060/0.060/0.061)
0.5	1.020(3.799/3.835/4.180)	0.393(0.782/0.787/0.870)	0.199(0.063/0.063/0.064)
0.75	1.022(4.248/4.291/4.772)	0.392(0.874/0.879/0.993)	0.199(0.070/0.070/0.072)
1	1.023(4.721/4.770/5.396)	0.392(0.970/0.976/1.123)	0.199(0.077/0.077/0.080)
α	$p = 0.01,\ \kappa = 0.4$		
	$\hat{d}$	$\hat{a}$	$\hat{b}$
MDPDE			
0 (CMLE)	1.194(10.24/13.98/13.75)	0.398(1.745/1.743/2.404)	0.152(0.094/0.325/0.099)
0.1	*1.123(5.745/7.248/6.580)*	*0.396(1.089/1.090/1.316)**	*0.157(0.067/0.256/0.070)*
0.2	*1.116(5.591/6.926/6.239)**	*0.392(1.090/1.096/1.282)*	*0.158(0.068/0.247/0.071)*
0.3	*1.119(5.708/7.111/6.439)*	*0.387(1.127/1.142/1.334)*	*0.159(0.072/0.239/0.074)*
0.4	*1.124(5.880/7.406/6.744)*	*0.383(1.171/1.199/1.402)*	*0.161(0.076/0.229/0.079)*
0.5	*1.129(6.065/7.733/7.076)*	*0.378(1.217/1.263/1.472)*	*0.163(0.081/0.218/0.084)*
0.75	*1.143(6.529/8.554/8.068)*	*0.368(1.329/1.431/1.673)*	*0.169(0.093/0.192/0.097)*
1	*1.153(6.979/9.309/9.425)*	*0.359(1.434/1.602/1.943)*	*0.174(0.106/0.172/0.110)**

where Y_t are generated from (5), P_t are iid Bernoulli random variables with success probability p, and $Y_{o,t}$ are iid Poisson random variable with intensity γ. We consider the cases of $p = 0.01,\ 0.03$ and $\gamma = 5, 10$. The results are presented in Tables 5, 6, and 7 for the cases of no outliers, $p = 0.01$, and $p = 0.03$, respectively. These tables show results similar to those in Tables 1, 2, and 3. In Table 7, we observe that $\hat{b}_1$ with $\alpha = 1$ has the smallest MSE when $p = 0.03$ and $\gamma = 10$. This result indicates that when the data are severely contaminated by outliers, $\hat{b}_1$ is more severely affected than the other parameters. Overall, our findings strongly support the MDPDE as a promising robust estimator for general integer-valued time series models.

4.2 Real data analysis

In this subsection, we apply the MDPDE to the analysis of return times of extreme events for Goldman Sachs Group (GS) stock, based on the daily log-returns between May 5, 1999, and March 15, 2012. Davis and Liu (2016) studied this dataset and considered the geometric distribution as a conditional distribution according to the

Table 5 Sample mean (variance × 10^2/MSE × 10^2) of estimators for the INTGARCH(1,1) model when no outliers exist

α	$\hat{d}$	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$
MDPDE				
0 (CMLE)	1.073(17.73/18.24)*	0.368(2.654/2.754)*	0.096(0.481/0.482)*	0.310(0.613/0.623)*
0.1	1.077(17.83/18.41)	0.367(2.670/2.778)	0.097(0.486/0.487)	0.310(0.624/0.633)
0.2	1.080(18.33/18.96)	0.366(2.752/2.866)	0.097(0.498/0.498)	0.309(0.648/0.657)
0.3	1.083(18.78/19.46)	0.365(2.839/2.960)	0.098(0.510/0.510)	0.309(0.681/0.689)
0.4	1.085(19.68/20.39)	0.364(3.000/3.125)	0.098(0.526/0.526)	0.309(0.724/0.731)
0.5	1.087(20.38/21.11)	0.364(3.120/3.246)	0.099(0.542/0.541)	0.308(0.766/0.772)
0.75	1.087(22.11/22.85)	0.364(3.408/3.533)	0.101(0.587/0.587)	0.307(0.873/0.878)
1	1.090(24.61/25.40)	0.364(3.768/3.896)	0.102(0.640/0.640)	0.307(0.985/0.989)

Table 6 Sample mean (variance × 10^2/MSE × 10^2) of estimators for the INTGARCH(1,1) model when $p = 0.01$

α	$\gamma = 5$			
	$\hat{d}$	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$
MDPDE				
0 (CMLE)	1.088(19.67/20.42)	0.373(2.924/2.994)	0.081(0.400/0.434)	0.312(0.737/0.752)
0.1	*1.081(18.74/19.37)**	*0.372(2.796/2.871)**	*0.082(0.397/0.430)**	*0.313(0.704/0.719)**
0.2	*1.081(18.81/19.45)*	*0.370(2.809/2.896)*	0.082(0.406/0.438)	*0.313(0.705/0.722)*
0.3	*1.083(19.25/19.92)*	0.368(2.900/3.001)	0.082(0.417/0.448)	*0.313(0.725/0.742)*
0.4	1.088(20.13/20.89)	0.365(3.044/3.163)	0.083(0.432/0.461)	0.314(0.750/0.768)
0.5	1.090(20.64/21.42)	0.364(3.144/3.270)	0.083(0.448/0.476)	0.314(0.778/0.796)
0.75	1.097(22.33/23.24)	0.361(3.432/3.584)	0.084(0.488/0.512)	0.314(0.859/0.877)
1	1.105(24.04/25.13)	0.357(3.725/3.908)	0.086(0.535/0.554)	0.314(0.947/0.966)
α	$\gamma = 10$			
	$\hat{d}$	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$
MDPDE				
0 (CMLE)	1.098(25.55/26.49)	0.377(3.623/3.671)	0.052(0.297/0.527)	0.323(0.871/0.925)
0.1	*1.064(20.53/20.92)*	*0.376(2.985/3.041)**	*0.053(0.283/0.506)**	*0.324(0.707/0.766)*
0.2	*1.053(20.14/20.40)**	*0.376(2.986/3.042)*	*0.053(0.287/0.506)*	*0.324(0.706/0.762)**
0.3	*1.052(20.52/20.78)*	*0.375(3.049/3.109)*	*0.054(0.295/0.509)*	*0.323(0.719/0.771)*
0.4	*1.053(21.16/21.41)*	*0.375(3.154/3.216)*	*0.054(0.305/0.515)*	*0.322(0.742/0.791)*
0.5	*1.053(21.63/21.89)*	*0.374(3.236/3.298)*	*0.055(0.318/0.522)*	*0.322(0.767/0.813)*
0.75	*1.055(23.68/23.96)*	*0.375(3.578/3.637)*	0.057(0.359/0.543)	*0.319(0.856/0.893)*
1	*1.057(25.35/25.65)*	0.376(3.870/3.926)	0.060(0.412/0.573)	0.317(0.958/0.986)

Table 7 Sample mean (variance × 10^2/MSE × 10^2) of estimators for the INTGARCH(1,1) model when $p = 0.03$

α	$\gamma = 5$			
	$\hat{d}$	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$
MDPDE				
0 (CMLE)	1.122(25.31/26.79)	0.379(3.538/3.580)	0.060(0.306/0.467)	0.319(0.849/0.886)
0.1	*1.098(21.95/22.89)*	*0.377(3.135/3.186)*	*0.060(0.290/0.447)*	*0.321(0.757/0.801)*
0.2	*1.088(21.35/22.11)**	*0.374(3.089/3.152)**	*0.061(0.291/0.445)**	*0.322(0.742/0.789)**
0.3	*1.087(21.64/22.37)*	*0.371(3.141/3.221)*	*0.061(0.299/0.448)*	*0.322(0.749/0.796)*
0.4	*1.085(22.04/22.74)*	*0.370(3.239/3.328)*	*0.062(0.309/0.455)*	*0.322(0.771/0.817)*
0.5	*1.087(22.96/23.69)*	*0.368(3.388/3.488)*	*0.062(0.322/0.463)*	*0.321(0.795/0.840)*
0.75	*1.093(24.63/25.46)*	0.364(3.682/3.808)	0.064(0.353/0.484)	0.321(0.870/0.913)
1	1.099(26.51/27.46)	0.361(4.017/4.165)	0.065(0.391/0.511)	0.321(0.961/1.002)
α	$\gamma = 10$			
	$\hat{d}$	$\hat{a}$	$\hat{b}_1$	$\hat{b}_2$
MDPDE				
0 (CMLE)	1.171(40.71/43.60)	0.389(5.180/5.186)	0.029(0.163/0.668)	0.333(1.383/1.489)
0.1	*1.073(24.52/25.02)*	*0.382(3.451/3.480)*	*0.027(0.122/0.651)*	*0.338(0.813/0.960)*
0.2	*1.039(21.94/22.07)*	*0.382(3.198/3.227)*	*0.027(0.116/0.649)*	*0.337(0.736/0.870)*
0.3	*1.027(21.31/21.36)**	*0.383(3.170/3.195)**	*0.027(0.117/0.650)*	*0.334(0.731/0.848)**
0.4	*1.025(21.88/21.92)*	*0.383(3.297/3.324)*	*0.027(0.119/0.649)*	*0.333(0.755/0.860)*
0.5	*1.028(22.61/22.67)*	*0.381(3.429/3.461)*	*0.028(0.122/0.646)*	*0.331(0.780/0.877)*
0.75	*1.037(25.05/25.16)*	*0.379(3.768/3.807)*	*0.029(0.136/0.639)*	*0.329(0.851/0.934)*
1	*1.043(27.03/27.19)*	*0.379(4.085/4.126)*	*0.031(0.160/0.637)**	*0.327(0.944/1.016)*

result in Chang (2010). We first calculate the hitting times, $\tau_1, \tau_2, \ldots$, for which the log-returns of GS stock falls outside of the 0.05 and 0.95 quantiles of the data. Then, the return times of extreme events are obtained by $Y_t = \tau_t - \tau_{t-1}$. Figure 1 plots the return times of GS stock and their autocorrelation function (ACF), showing that the data are serially autocorrelated and have some aberrant observations. The sample mean and variance are 10.01 and 1106, respectively. A significantly large value of the sample variance is expected to be influenced by some extraordinarily large observations.

To illustrate the behavior of the MDPDE in the presence of outliers, we fit the geometric INGARCH(1,1) model (NB-INGARCH(1,1) model with $r = 1$) to the data using both the MDPDE and CMLE. Because $Y_t \geq 1$, we employ a version of the geometric distribution that counts the total number of trials instead of the failures. More precisely, we fit the following geometric INGARCH(1,1) model to the data:

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

where $\widetilde{X}_1$ is the sample mean of the data and $\partial\tilde{\eta}_1(\hat{\theta}_{\alpha,n})/\partial\theta$ and $\partial^2\tilde{\eta}_1(\hat{\theta}_{\alpha,n})/\partial\theta\partial\theta^{\mathrm{T}}$ are set to be zero vector and matrix, respectively, for computing the standard errors

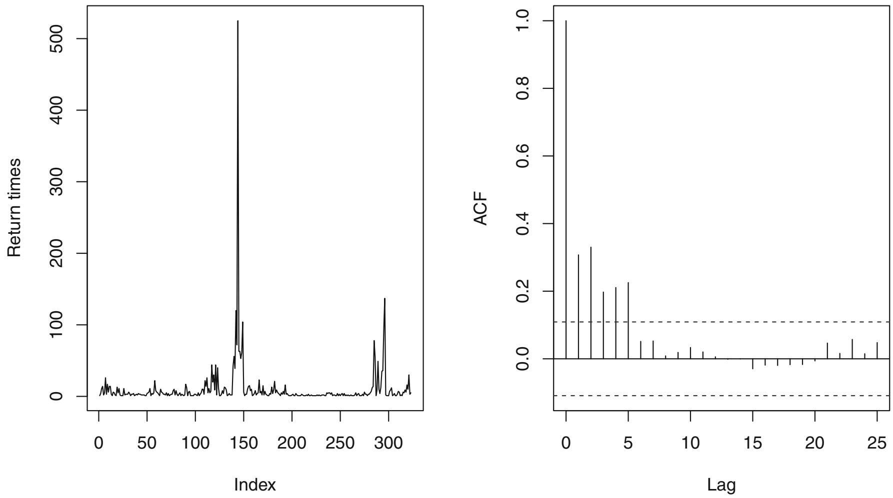

Fig. 1 Return times of GS stock and their autocorrelation function

Table 8 Parameter estimates for the geometric INGARCH(1,1) model

α	$\hat{d}$	$\hat{a}$	$\hat{b}$	$\widehat{\text{AMSE}}$
MDPDE				
0 (CMLE)	0.526(0.406)	0.490(0.175)	0.483(0.156)	0.623
0.05	0.420(0.288)	0.502(0.141)	0.484(0.135)	0.476
0.1	0.416(0.250)	0.506(0.127)	0.464(0.121)	0.430
0.15	0.420(0.238)	0.509(0.123)	0.447(0.115)	0.410
0.2	0.426(0.237)	0.513(0.125)	0.432(0.114)	0.401
0.25	0.432(0.242)	0.518(0.129)	0.418(0.115)	0.398$^{\bullet}$
0.3	0.437(0.249)	0.523(0.135)	0.406(0.118)	0.399
0.35	0.439(0.258)	0.529(0.141)	0.395(0.121)	0.402
0.4	0.438(0.267)	0.535(0.148)	0.385(0.125)	0.405
0.45	0.435(0.275)	0.543(0.154)	0.375(0.128)	0.409
0.5	0.429(0.282)	0.551(0.159)	0.366(0.131)	0.411

of the estimators and $\widehat{\text{AMSE}}$ in Remark 2. Because the MDPDE with a too large α can result in a significant loss of efficiency (Basu et al. 1998), we consider α from 0.05 to 0.5 with increase of 0.05. Table 8 presents the parameter estimates for the geometric INGARCH(1,1) model. In Table 8, the figures in parentheses denote the standard errors of the corresponding estimators and the symbol $^{\bullet}$ stands for the minimal $\widehat{\text{AMSE}}$. In the table, we observe that the MDPDE is more stable than the CMLE. In other words, the standard errors of the MDPDE are smaller than those of the CMLE for all α and parameters. This result indicates that the MDPDE performs better than the CMLE in terms of the stability when outliers are present. The optimal α is chosen to be $\alpha = 0.25$, and the corresponding MDPDE appears to be quite different from the

CMLE. This indicates that outliers affect the parameter estimates when the geometric INGARCH(1,1) model is employed. Note that for all α, $\hat{a} + \hat{b}$ are close to 1, which resembles the behavior of integrated GARCH models. Overall, our findings confirm that the MDPDE can provide a functional robust alternative to the CMLE in the presence of outliers.

5 Concluding remarks

In this study, we developed an MDPDE-based robust estimation method for general integer-valued time series models with a conditional distribution belonging to the one-parameter exponential family. We verified that under certain regularity conditions, the MDPDE is consistent and asymptotically normal. Our simulation study and real data analysis confirmed the validity of the proposed estimator.

Acknowledgements We thank the Editor, an AE, and one referee for their consideration and valuable comments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2015R1C1A1A01052330) (B. Kim) and (No. 2018R1A2A2A05019433) (S. Lee).

Appendix

In this appendix, we provide the proofs of Theorems 1 and 2. Because Lee and Lee (2018) verified the strong consistency and asymptotic normality of the CMLE under similar conditions, we focus on the MDPDE with $\alpha > 0$. The asymptotic results for the CMLE also can be found in Davis and Liu (2016) and Cui and Zheng (2017). The following properties of the probability mass function of the nonnegative integer-valued exponential family are useful for proving the theorems. 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) $\sum_{y=0}^{\infty} y^2 p(y|\eta) = B'(\eta) + B(\eta)^2$.

In what follows, we denote $H_{\alpha,n}(\theta) = n^{-1} \sum_{t=1}^{n} l_{\alpha,t}(\theta)$ and employ the notations $\eta_t = \eta_t(\theta)$, $\tilde{\eta}_t = \tilde{\eta}_t(\theta)$ and $\eta_t^0 = \eta_t(\theta_0)$ for brevity.

Lemma 1 *Suppose that the conditions* **(A0)–(A3)** *hold. Then, we have*

$$\sup_{\theta \in \Theta} |\widetilde{X}_t(\theta) - X_t(\theta)| \leq V\rho^t \ \ \textit{and} \ \ \sup_{\theta \in \Theta} |\tilde{\eta}_t - \eta_t| \leq V\rho^t \ \ \textit{a.s.}$$

Proof From **(A0)**, we have

$$\begin{aligned} |\widetilde{X}_t(\theta) - X_t(\theta)| &= |f_\theta(\widetilde{X}_{t-1}(\theta), Y_{t-1}) - f_\theta(X_{t-1}(\theta), Y_{t-1})| \\ &\leq \omega_1 |\widetilde{X}_{t-1}(\theta) - X_{t-1}(\theta)| \\ &\leq \omega_1^{t-1} |\widetilde{X}_1 - X_1(\theta)|. \end{aligned}$$

Then, due to the mean value theorem (MVT) and **(A3)** with the fact that B^{-1} is strictly increasing, it holds that

$$\begin{aligned}|\tilde{\eta}_t - \eta_t| &= |B^{-1}(\widetilde{X}_t(\theta)) - B^{-1}(X_t(\theta))| \\ &= \frac{1}{B'(B^{-1}(X_t^*(\theta)))}|\widetilde{X}_t(\theta) - X_t(\theta)| \\ &\le \frac{\omega_1^{t-1}}{B'(\eta_t^*)}|\widetilde{X}_1 - X_1(\theta)| \\ &\le \frac{\omega_1^{t-1}}{\underline{c}}|\widetilde{X}_1 - X_1(\theta)|,\end{aligned}$$

where $\eta_t^* = B^{-1}(X_t^*(\theta))$ and $X_t^*(\theta)$ is an intermediate point between $\widetilde{X}_t(\theta)$ and $X_t(\theta)$. Hence, the proof is completed by **(A2)**. □

Lemma 2 *Suppose that conditions* **(A0)–(A4)** *hold. Then, we have*

$$\sup_{\theta\in\Theta}|H_{\alpha,n}(\theta) - \widetilde{H}_{\alpha,n}(\theta)| \xrightarrow{a.s.} 0 \ \ as \ \ n \to \infty.$$

Proof It suffices to show that

$$\sup_{\theta\in\Theta}|l_{\alpha,t}(\theta) - \tilde{l}_{\alpha,t}(\theta)| \xrightarrow{a.s.} 0 \ \text{ as } \ t \to \infty.$$

Note that $|l_{\alpha,t}(\theta) - \tilde{l}_{\alpha,t}(\theta)| \le I_t(\theta) + II_t(\theta)$, where

$$\begin{aligned}I_t(\theta) &= \left|\sum_{y=0}^{\infty} p(y|\eta_t)^{1+\alpha} - \sum_{y=0}^{\infty} p(y|\tilde{\eta}_t)^{1+\alpha}\right|, \\ II_t(\theta) &= \left(1 + \frac{1}{\alpha}\right)\left|p(Y_t|\eta_t)^{\alpha} - p(Y_t|\tilde{\eta}_t)^{\alpha}\right|.\end{aligned}$$

First, due to the MVT, (E1)–(E3), and the fact that B is strictly increasing, it holds that

$$\begin{aligned}I_t(\theta) &\le (1+\alpha)|\eta_t - \tilde{\eta}_t|\sum_{y=0}^{\infty} p(y|\eta_t^*)^{1+\alpha}|y - B(\eta_t^*)| \\ &\le 2(1+\alpha)|\eta_t - \tilde{\eta}_t|B(\eta_t^*) \\ &\le 2(1+\alpha)|\eta_t - \tilde{\eta}_t|\left(B(\eta_t) + |B(\tilde{\eta}_t) - B(\eta_t)|\right) \\ &= 2(1+\alpha)|\eta_t - \tilde{\eta}_t|(X_t(\theta) + |\widetilde{X}_t(\theta) - X_t(\theta)|)\end{aligned}$$

for some intermediate point η_t^* between η_t and $\tilde{\eta}_t$. Hence,

$$\sup_{\theta\in\Theta} I_t(\theta) \le 2(1+\alpha)\sup_{\theta\in\Theta}|\eta_t - \tilde{\eta}_t|\left(\sup_{\theta\in\Theta} X_t(\theta) + \sup_{\theta\in\Theta}|\widetilde{X}_t(\theta) - X_t(\theta)|\right).$$

According to Lemma 2.1 of Straumann and Mikosch (2006) together with Lemma 1 and **(A2)**, $\sup_{\theta\in\Theta} X_t(\theta) \sup_{\theta\in\Theta} |\eta_t - \tilde{\eta}_t| \to 0$ a.s. as $t \to \infty$. Therefore, $\sup_{\theta\in\Theta} I_t(\theta)$ converges to 0 a.s. by Lemma 1.

Next, from the MVT, (E1), and the fact that B is strictly increasing, we have

$$\begin{aligned} II_t(\theta) &= (1+\alpha)p(Y_t|\eta_t^*)^\alpha |Y_t - B(\eta_t^*)||\eta_t - \tilde{\eta}_t| \\ &\le (1+\alpha)|\eta_t - \tilde{\eta}_t|(Y_t + B(\eta_t^*)) \\ &\le (1+\alpha)|\eta_t - \tilde{\eta}_t|(Y_t + B(\eta_t) + |B(\tilde{\eta}_t) - B(\eta_t)|) \\ &= (1+\alpha)|\eta_t - \tilde{\eta}_t|(Y_t + X_t(\theta) + |\widetilde{X}_t(\theta) - X_t(\theta)|). \end{aligned}$$

Hence,

$$\sup_{\theta\in\Theta} II_t(\theta) \le (1+\alpha)\sup_{\theta\in\Theta}|\eta_t - \tilde{\eta}_t| \left(Y_t + \sup_{\theta\in\Theta} X_t(\theta) + \sup_{\theta\in\Theta} |\widetilde{X}_t(\theta) - X_t(\theta)| \right).$$

By using Lemma 2.1 of Straumann and Mikosch (2006) again with Lemma 1 and **(A4)**, it holds that $Y_t \sup_{\theta\in\Theta} |\eta_t - \tilde{\eta}_t| \to 0$ a.s. as $t \to \infty$. By applying the same method to the remaining terms, we have $\sup_{\theta\in\Theta} II_t(\theta)$ converges to 0 a.s. Therefore, the lemma is validated. □

Lemma 3 *Suppose that conditions* **(A0)–(A5)** *hold. Then, we have*

$$E\left(\sup_{\theta\in\Theta} |l_{\alpha,t}(\theta)|\right) < \infty \ \text{ and if } \theta \neq \theta_0, \text{ then } El_{\alpha,t}(\theta) > El_{\alpha,t}(\theta_0).$$

Proof From (E1) and (E2), it can be seen that

$$|l_{\alpha,t}(\theta)| \le \sum_{y=0}^{\infty} p(y|\eta_t)^{1+\alpha} + \left(1 + \frac{1}{\alpha}\right) p(Y_t|\eta_t)^\alpha \le 2 + \frac{1}{\alpha},$$

and thus the first part of the lemma is established. Note that

$$\begin{aligned} &El_{\alpha,t}(\theta) - El_{\alpha,t}(\theta_0) \\ &= E\left[E(l_{\alpha,t}(\theta) - l_{\alpha,t}(\theta_0)|\mathcal{F}_{t-1})\right] \\ &= E\left[\sum_{y=0}^{\infty}\left(p(y|\eta_t)^{1+\alpha} - \left(1+\frac{1}{\alpha}\right)p(y|\eta_t)^\alpha p(y|\eta_t^0) + \frac{1}{\alpha}p(y|\eta_t^0)^{1+\alpha}\right)\right] \\ &\ge 0, \end{aligned}$$

where equality holds if and only if $\eta_t = \eta_t^0$ a.s. Therefore, by **(A5)** and the fact that B is strictly increasing, the lemma is asserted. □

Proof of Theorem 1 We can express

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\tilde{l}_{\alpha,t}(\theta)-El_{\alpha,t}(\theta)\right| \le \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\tilde{l}_{\alpha,t}(\theta)-\frac{1}{n}\sum_{t=1}^{n}l_{\alpha,t}(\theta)\right| + \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}l_{\alpha,t}(\theta)-El_{\alpha,t}(\theta)\right|.$$

By Lemma 2, the first term on the RHS of the above inequality converges to 0 a.s. Since $l_{\alpha,t}(\theta)$ is stationary and ergodic with $E(\sup_{\theta\in\Theta}|l_{\alpha,t}(\theta)|)<\infty$ by Lemma 3, the second term on the RHS also converges to 0 a.s. (cf. Theorem 2.7 of Straumann and Mikosch 2006). Moreover, since $El_{\alpha,t}(\theta)$ has a unique minimum at θ_0 from Lemma 3, the theorem is established. □

In order to derive the first and second derivatives of $l_{\alpha,t}(\theta)$, we define two functions $h_\alpha(\eta)$ and $m_\alpha(\eta)$ as

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

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

Then, we have

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

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

The following four lemmas are useful for proving Theorem 2.

Lemma 4 *Suppose that conditions* **(A3)** *and* **(A6)** *hold. Then, we have*

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

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

$$|m_\alpha(\eta_t)|\le\frac{\alpha}{\underline{c}^2}Y_t^2+KY_t+\frac{\alpha}{\underline{c}^2}X_t(\theta)^2+3KX_t(\theta)+\frac{3+\alpha}{\underline{c}},$$

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

Proof Due to (E1)–(E4), **(A3)**, and **(A6)**, we have

$$\begin{aligned}|h_\alpha(\eta_t)| &\leq \frac{1}{\underline{c}}\left(\sum_{y=0}^{\infty} y p(y|\eta_t) + B(\eta_t)\sum_{y=0}^{\infty} p(y|\eta_t) + Y_t + B(\eta_t)\right)\\ &= \frac{1}{\underline{c}}\left(Y_t + 3B(\eta_t)\right)\end{aligned}$$

and

$$\begin{aligned}|m_\alpha(\eta_t)| \leq\ & \frac{1+\alpha}{B'(\eta_t)^2}\sum_{y=0}^{\infty} p(y|\eta_t)\left(y - B(\eta_t)\right)^2 + \left|\frac{B''(\eta_t)}{B'(\eta_t)^3}\right| \sum_{y=0}^{\infty} p(y|\eta_t)\left(y + B(\eta_t)\right)\\ & + \frac{1}{B'(\eta_t)}\sum_{y=0}^{\infty} p(y|\eta_t) + \frac{\alpha}{B'(\eta_t)^2}\left(Y_t - B(\eta_t)\right)^2\\ & + \left|\frac{B''(\eta_t)}{B'(\eta_t)^3}\right|\left(Y_t + B(\eta_t)\right) + \frac{1}{B'(\eta_t)}\\ \leq\ & \frac{1+\alpha}{\underline{c}} + 2K B(\eta_t) + \frac{1}{\underline{c}} + \frac{\alpha}{\underline{c}^2}\left(Y_t^2 + B(\eta_t)^2\right) + K\left(Y_t + B(\eta_t)\right) + \frac{1}{\underline{c}}\\ =\ & \frac{\alpha}{\underline{c}^2} Y_t^2 + K Y_t + \frac{\alpha}{\underline{c}^2} B(\eta_t)^2 + 3K B(\eta_t) + \frac{3+\alpha}{\underline{c}}.\end{aligned}$$

Hence, the first and third parts of the lemma are verified. The second part of the lemma can be obtained using the fact that $B(\tilde{\eta}_t) \leq |B(\eta_t) - B(\tilde{\eta}_t)| + B(\eta_t)$.

Note that since $m_\alpha(\eta_t) = \partial h_\alpha(\eta_t)/\partial X_t(\theta)$, it holds that

$$\begin{aligned}&|h_\alpha(\eta_t) - h_\alpha(\tilde{\eta}_t)|\\ &= |m_\alpha(\eta_t^*)||X_t(\theta) - \widetilde{X}_t(\theta)|\\ &\leq \left(\frac{\alpha}{\underline{c}^2} Y_t^2 + K Y_t + \frac{\alpha}{\underline{c}^2} B(\eta_t^*)^2 + 3K B(\eta_t^*) + \frac{3+\alpha}{\underline{c}}\right)|X_t(\theta) - \widetilde{X}_t(\theta)|\\ &\leq \left(\frac{\alpha}{\underline{c}^2} Y_t^2 + K Y_t + \frac{2\alpha}{\underline{c}^2}\left(B(\eta_t)^2 + |B(\tilde{\eta}_t) - B(\eta_t)|^2\right)\right.\\ &\quad \left. +3K\left(B(\eta_t) + |B(\tilde{\eta}_t) - B(\eta_t)|\right) + \frac{3+\alpha}{\underline{c}}\right)|X_t(\theta) - \widetilde{X}_t(\theta)|\end{aligned}$$

by the MVT, the third part of the lemma, and the fact that $B(\eta_t^*) \leq B(\eta_t) + |B(\tilde{\eta}_t) - B(\eta_t)|$. Therefore, the fourth part is established. This completes the proof. □

Lemma 5 *Suppose that conditions* **(A0)–(A7)** *hold. Then, we have*

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

Proof Note that

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

Using Lemma 4 and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&\frac{1}{1+\alpha}E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T}\right\|\right)\\
&\le\sqrt{E\left(\sup_{\theta\in\Theta}|h_\alpha(\eta_t)|\right)^2}\sqrt{E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 X_t(\theta)}{\partial\theta\partial\theta^T}\right\|\right)^2}\\
&\quad+\sqrt{E\left(\sup_{\theta\in\Theta}|m_\alpha(\eta_t)|\right)^2}\sqrt{E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\frac{\partial X_t(\theta)}{\partial\theta^T}\right\|\right)^2}\\
&\le\sqrt{E\left[\frac{1}{\underline{c}}\left(Y_t+3\sup_{\theta\in\Theta}X_t(\theta)\right)\right]^2}\sqrt{E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 X_t(\theta)}{\partial\theta\partial\theta^T}\right\|\right)^2}\\
&\quad+\sqrt{E\left(\frac{\alpha}{\underline{c}^2}Y_t^2+KY_t+\frac{\alpha}{\underline{c}^2}\sup_{\theta\in\Theta}X_t(\theta)^2+3K\sup_{\theta\in\Theta}X_t(\theta)+\frac{3+\alpha}{\underline{c}}\right)^2}\\
&\quad\times\sqrt{E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\frac{\partial X_t(\theta)}{\partial\theta^T}\right\|\right)^2}.
\end{aligned}$$

Owing to **(A2), (A4)**, and **(A7)**, the RHS of the last inequality is finite.

In a similar manner, we can show that

$$\begin{aligned}
&\frac{1}{(1+\alpha)^2}E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial l_{\alpha,t}(\theta)}{\partial\theta}\frac{\partial l_{\alpha,t}(\theta)}{\partial\theta^T}\right\|\right)\\
&\le\sqrt{E\left(\sup_{\theta\in\Theta}|h_\alpha(\eta_t)|^2\right)^2}\sqrt{E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\frac{\partial X_t(\theta)}{\partial\theta^T}\right\|\right)^2}\\
&\le\sqrt{E\left[\sup_{\theta\in\Theta}\left(\frac{1}{\underline{c}}(Y_t+3X_t(\theta))\right)^2\right]^2}\sqrt{E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\frac{\partial X_t(\theta)}{\partial\theta^T}\right\|\right)^2}\\
&\le\sqrt{\frac{4}{\underline{c}^4}E\left(Y_t^2+9\sup_{\theta\in\Theta}X_t(\theta)^2\right)^2}\sqrt{E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\frac{\partial X_t(\theta)}{\partial\theta^T}\right\|\right)^2}\\
&<\infty.
\end{aligned}$$

Therefore, the lemma is verified. □

Lemma 6 *Suppose that conditions* **(A0)–(A8)** *hold. Then, we have*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left\|\frac{\partial l_{\alpha,t}(\theta)}{\partial\theta}-\frac{\partial\tilde{l}_{\alpha,t}(\theta)}{\partial\theta}\right\|\xrightarrow{a.s.}0 \ \ as \ \ n\to\infty.$$

Proof From Lemma 1, 4, and **(A8)**, we can write

$$\begin{aligned}
&\frac{1}{1+\alpha}\sup_{\theta\in\Theta}\left\|\frac{\partial l_{\alpha,t}(\theta)}{\partial\theta}-\frac{\partial\tilde{l}_{\alpha,t}(\theta)}{\partial\theta}\right\|\\
&\le\sup_{\theta\in\Theta}|h_\alpha(\tilde{\eta}_t)|\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right\|\\
&\quad+\sup_{\theta\in\Theta}|h_\alpha(\eta_t)-h_\alpha(\tilde{\eta}_t)|\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\right\|\\
&\le\frac{1}{\underline{c}}\left(Y_t+3\sup_{\theta\in\Theta}X_t(\theta)+3\sup_{\theta\in\Theta}|X_t(\theta)-\widetilde{X}_t(\theta)|\right)\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right\|\\
&\quad+\left[\frac{\alpha}{\underline{c}^2}Y_t^2+KY_t+\frac{2\alpha}{\underline{c}^2}\left(\sup_{\theta\in\Theta}X_t(\theta)^2+\sup_{\theta\in\Theta}|X_t(\theta)-\widetilde{X}_t(\theta)|^2\right)\right.\\
&\quad\left.+3K\left(\sup_{\theta\in\Theta}X_t(\theta)+\sup_{\theta\in\Theta}|X_t(\theta)-\widetilde{X}_t(\theta)|\right)+\frac{3+\alpha}{\underline{c}}\right]\\
&\sup_{\theta\in\Theta}|X_t(\theta)-\widetilde{X}_t(\theta)|\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\right\|\\
&\le\frac{1}{\underline{c}}\left(Y_t+3\sup_{\theta\in\Theta}X_t(\theta)+3V\rho^t\right)V\rho^t+\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\right\|\\
&\quad\times\left[\frac{\alpha}{\underline{c}^2}Y_t^2+KY_t+\frac{2\alpha}{\underline{c}^2}\left(\sup_{\theta\in\Theta}X_t(\theta)^2+V^2\rho^{2t}\right)\right.\\
&\quad\left.+3K\left(\sup_{\theta\in\Theta}X_t(\theta)+V\rho^t\right)+\frac{3+\alpha}{\underline{c}}\right]V\rho^t.
\end{aligned}$$

Hence, due to Lemma 2.1 of Straumann and Mikosch (2006) together with **(A2), (A4)**, and **(A7)**, the RHS of the last inequality converges to 0 exponentially fast a.s., and the lemma is established. For details of the concept and properties of exponentially fast a.s. convergence, we refer the reader to Straumann and Mikosch (2006) and Cui and Zheng (2017). □

Lemma 7 *Suppose that*

$$\hat{\theta}_{\alpha,n}^H=\underset{\theta\in\Theta}{argmin}\,H_{\alpha,n}(\theta),$$

and conditions **(A0)–(A9)** *hold. Then, we have*

$$\hat{\theta}_{\alpha,n}^H\xrightarrow{a.s.}\theta_0$$

and

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

Proof As in the proof of Theorem 1, we can see that $\sup_{\theta\in\Theta} |n^{-1}\sum_{t=1}^{n} l_{\alpha,t}(\theta) - El_{\alpha,t}(\theta)|$ converges to 0 a.s., and because $El_{\alpha,t}(\theta)$ has a unique minimum at θ_0 by Lemma 3, the first part of the lemma is validated.

Now, we verify the second part of the lemma. By using the MVT, we obtain

$$0 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta} + \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_{\alpha,t}(\theta_{\alpha,n}^*)}{\partial\theta\partial\theta^T}\right)\sqrt{n}(\hat{\theta}_{\alpha,n}^{H} - \theta_0),$$

where $\theta_{\alpha,n}^*$ is an intermediate point between θ_0 and $\hat{\theta}_{\alpha,n}^{H}$. First, we show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, K_\alpha). \tag{6}$$

For $\nu \in \mathbb{R}^d$, we have

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

and

$$E\left(\nu^{\mathrm{T}}\frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta}\right)^2 = \nu^{\mathrm{T}}E\left(\frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta}\frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta^T}\right)\nu < \infty$$

owing to the second part of Lemma 5. Thus, using the central limit theorem in Billingsley (1961), we obtain

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\nu^{\mathrm{T}}\frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, \nu^{\mathrm{T}}K_\alpha\nu),$$

which asserts (6).

Next, we show that

$$-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_{\alpha,t}(\theta_{\alpha,n}^*)}{\partial\theta\partial\theta^T} \xrightarrow{a.s.} J_\alpha. \tag{7}$$

In view of the first part of Lemma 5, J_α is finite. Moreover, since

$$E\left(m_\alpha(\eta_t^0)|\mathcal{F}_{t-1}\right) = \sum_{y=0}^{\infty}p(y|\eta_t^0)^{1+\alpha}\left(\frac{y - B(\eta_t^0)}{B'(\eta_t^0)}\right)^2 > 0,$$

it holds that

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

by **(A9)**, which implies that J_α is non-singular. Note that

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_{\alpha,t}(\theta^*_{\alpha,n})}{\partial\theta\partial\theta^T} - E\left(\frac{\partial^2 l_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T}\right)\right\|$$
$$\leq \sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - E\left(\frac{\partial^2 l_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T}\right)\right\|$$
$$+\left\|E\left(\frac{\partial^2 l_{\alpha,t}(\theta^*_{\alpha,n})}{\partial\theta\partial\theta^T}\right) - E\left(\frac{\partial^2 l_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T}\right)\right\|.$$

The stationarity and ergodicity of $\partial^2 l_{\alpha,t}(\theta)/\partial\theta\partial\theta^{\mathrm{T}}$ and the first part of Lemma 5 imply that the first term on the RHS of the above inequality converges to 0 a.s. Furthermore, the second term goes to 0 by the dominated convergence theorem, so that (7) is verified. Therefore, from (6) and (7), the second part of the lemma is established. □

Proof of Theorem 2 Owing to the MVT, we have

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_{\alpha,t}(\hat\theta^H_{\alpha,n})}{\partial\theta} - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_{\alpha,t}(\hat\theta_{\alpha,n})}{\partial\theta} = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_{\alpha,t}(\zeta_{\alpha,n})}{\partial\theta\partial\theta^T}\right)(\hat\theta^H_{\alpha,n} - \hat\theta_{\alpha,n}),$$

where $\zeta_{\alpha,n}$ is an intermediate point between $\hat\theta^H_{\alpha,n}$ and $\hat\theta_{\alpha,n}$. Furthermore, from the facts that $n^{-1}\sum_{t=1}^{n}\partial l_{\alpha,t}(\hat\theta^H_{\alpha,n})/\partial\theta = 0$ and $n^{-1}\sum_{t=1}^{n}\partial\tilde l_{\alpha,t}(\hat\theta_{\alpha,n})/\partial\theta = 0$, we obtain

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde l_{\alpha,t}(\hat\theta_{\alpha,n})}{\partial\theta} - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_{\alpha,t}(\hat\theta_{\alpha,n})}{\partial\theta} = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_{\alpha,t}(\zeta_{\alpha,n})}{\partial\theta\partial\theta^T}\right)\sqrt{n}(\hat\theta^H_{\alpha,n} - \hat\theta_{\alpha,n}).$$

The LHS of the above equation converges to 0 a.s. by Lemma 6, and we can show that $n^{-1}\sum_{t=1}^{n}\partial^2 l_{\alpha,t}(\zeta_{\alpha,n})/\partial\theta\partial\theta^{\mathrm{T}}$ converges to $E(\partial^2 l_{\alpha,t}(\theta_0)/\partial\theta\partial\theta^T)$ a.s. in a similar manner as in the proof of Lemma 7. Therefore, the theorem is established by Lemma 7.
□

References

Ahmad, A., Francq, C. (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis*, *37*, 291–314.

Al-Osh, M. A., Alzaid, A. A. (1987). First order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, *8*, 261–275.

Basu, A., Harris, I. R., Hjort, N. L., Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, *85*, 549–559.

Billingsley, P. (1961). The Lindeberg-Lévy theorem for martingales. *Proceedings of the American Mathematical Society*, *12*, 788–792.

Chang, L. (2010). Conditional modeling and conditional inference. Ph.D. thesis, Brown University, Providence, Rhode Island.

Christou, V., Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, *35*, 55–78.

Cui, Y., Zheng, Q. (2017). Conditional maximum likelihood estimation for a class of observation-driven time series models for count data. *Statistics and Probability Letters*, *123*, 193–201.

Davis, R. A., Liu, H. (2016). Theory and inference for a class of observation-driven models with application to time series of counts. *Statistica Sinica*, *26*, 1673–1707.

Davis, R. A., Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika*, *96*, 735–749.

Diop, M. L., Kengne, W. (2017). Testing parameter change in general integer-valued time series. *Journal of Time Series Analysis*, *38*, 880–894.

Doukhan, P., Kengne, W. (2015). Inference and testing for structural change in general poisson autoregressive models. *Electronic Journal of Statistics*, *9*, 1267–1314.

Durio, A., Isaia, E. D. (2011). The minimum density power divergence approach in building robust regression models. *Informatica*, *22*, 43–56.

Ferland, R., Latour, A., Oraichi, D. (2006). Integer-valued GARCH processes. *Journal of Time Series Analysis*, *27*, 923–942.

Fokianos, K., Rahbek, A., Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, *104*, 1430–1439.

Fried, R., Agueusop, I., Bornkamp, B., Fokianos, K., Fruth, J., Ickstadt, K. (2015). Retrospective Bayesian outlier detection in INGARCH series. *Statistics and Computing*, *25*, 365–374.

Fujisawa, H., Eguchi, S. (2006). Robust estimation in the normal mixture model. *Journal of Statistical Planning and Inference*, *136*, 3989–4011.

Kang, J., Lee, S. (2014a). Minimum density power divergence estimator for Poisson autoregressive models. *Computational Statistics and Data Analysis*, *80*, 44–56.

Kang, J., Lee, S. (2014b). Parameter change test for Poisson autoregressive models. *Scandinavian Journal of Statistics*, *41*, 1136–1152.

Kim, B., Lee, S. (2013). Robust estimation for the covariance matrix of multivariate time series based on normal mixtures. *Computational Statistics and Data Analysis*, *57*, 125–140.

Kim, B., Lee, S. (2017). Robust estimation for zero-inflated Poisson autoregressive models based on density power divergence. *Journal of Statistical Computation and Simulation*, *87*, 2981–2996.

Lee, S., Lee, Y., Chen, C. W. S. (2016). Parameter change test for zero-inflated generalized Poisson autoregressive models. *Statistics*, *50*, 540–557.

Lee, S., Song, J. (2009). Minimum density power divergence estimator for GARCH models. *Test*, *18*, 316–341.

Lee, Y., Lee, S. (2018). CUSUM test for general nonlinear integer-valued GARCH models: Comparison study. Annals of the Institute of Statistical Mathematics. https://doi.org/10.1007/s10463-018-0676-7.

Lehmann, E., Casella, G. (1998). *Theory of point estimation*2nd ed. New York: Springer.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, *21*, 645–650.

Mihoko, M., Eguchi, S. (2002). Robust blind source separation by beta divergence. *Neural Computation*, *14*, 1859–1886.

Straumann, D., Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics*, *34*, 2449–2495.

Toma, A., Broniatowski, M. (2011). Dual divergence estimators and tests: Robustness results. *Journal of Multivariate Analysis*, *102*, 20–36.

Warwick, J. (2005). A data-based method for selecting tuning parameters in minimum distance estimators. *Computational Statistics and Data Analysis*, *48*, 571–585.

Warwick, J., Jones, M. C. (2005). Choosing a robustness tuning parameter. *Journal of Statistical Computation and Simulation*, *75*, 581–588.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts-a survey. *AStA-Advances in Statistical Analysis*, *92*, 319–341.
Zhu, F. (2012a). Modeling overdispersed or underdispersed count data with generalized poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, *389*, 58–71.
Zhu, F. (2012b). Zero-inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference*, *142*, 826–839.