Article

Robust Estimation for Bivariate Poisson INGARCH Models

Byungsoo Kim [1,*]**, Sangyeol Lee** [2] **and Dongwon Kim** [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 (S.L.);
    dongwon.k@snu.ac.kr (D.K.)
*   Correspondence: bkim@yu.ac.kr

Abstract: In the integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) models, parameter estimation is conventionally based on the conditional maximum likelihood estimator (CMLE). However, because the CMLE is sensitive to outliers, we consider a robust estimation method for bivariate Poisson INGARCH models while using the minimum density power divergence estimator. We demonstrate the proposed estimator is consistent and asymptotically normal under certain regularity conditions. Monte Carlo simulations are conducted to evaluate the performance of the estimator in the presence of outliers. Finally, a real data analysis using monthly count series of crimes in New South Wales and an artificial data example are provided as an illustration.

Keywords: integer-valued time series; bivariate Poisson INGARCH model; outliers; robust estimation; minimum density power divergence estimator

1. Introduction

Integer-valued time series models have received widespread attention from researchers and practitioners, due to their versatile applications in many scientific areas, including finance, insurance, marketing, and quality control. Numerous studies focus on integer-valued autoregressive (INAR) models to analyze the time series of counts, see Weiß [1] and Scotto et al. [2] for general reviews. Taking a different approach, Ferland et al. [3] proposed using Poisson integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) models and Fokianos et al. [4] developed Poisson AR models to generalize the linear assumption on INGARCH models. The Poisson assumption on INGARCH models has been extended to negative binomial INGARCH models (Davis and Wu [5] and Christou and Fokianos [6]), zero-inflated generalized Poisson INGARCH models (Zhu [7,8] and Lee et al. [9]), and one-parameter exponential family AR models (Davis and Liu [10]). We refer to the review papers by Fokianos [11,12] and Tjøstheim [13,14] for more details.

Researchers invested considerable efforts to extend the univariate integer-valued time series models to bivariate (multivariate) models. For INAR type models, Quoreshi [15] proposed bivariate integer-valued moving average models and Pedeli and Karlis [16] introduced bivariate INAR models with Poisson and negative binomial innovations. Liu [17] proposed bivariate Poisson INGARCH models with a bivariate Poisson distribution that was constructed via the trivariate reduction method and established the stationarity and ergodicity of the model. Andreassen [18] later verified the consistency of the conditional maximum likelihood estimator (CMLE) and Lee et al. [19] studied the asymptotic normality of the CMLE and developed the CMLE- and residual-based change point tests. However, this model has the drawback that it can only accommodate positive correlation between two time series of counts. To cope with this issue, Cui and Zhu [20] recently introduced a new bivariate Poisson INGARCH model based on Lakshminarayana et al.'s [21] bivariate Poisson distribution. Their model can deal with positive or negative correlation, depending on the multiplicative factor parameter. They employed the CMLE for parameter estimation. However, because the CMLE is unduly influenced by outliers, the robust estimation in bivariate Poisson INGARCH models is crucial and deserves thorough investigation.

As such, here we develop a robust estimator for Cui and Zhu's [20] bivariate Poisson INGARCH models. Among the robust estimation methods, we employ the minimum density power divergence estimator (MDPDE) approach that was originally proposed by Basu et al. [22], because it is well known to consistently provide robust estimators in various situations. For previous works in the context of time series of counts, see Kang and Lee [23], Kim and Lee [24,25], Diop and Kengne [26], Kim and Lee [27], and Lee and Kim [28], who studied the MDPDE for Poisson AR models, zero-inflated Poisson AR models, one-parameter exponential family AR models, and change point tests. For another robust estimation approach in INGARCH models, see Xiong and Zhu [29] and Li et al. [30], who studied Mallows' quasi-likelihood method. To the best of our knowledge, the robust estimation method for bivariate Poisson INGARCH models has not been previously studied. In earlier studies, the MDPDE was proven to possess strong robust properties against outliers with little loss in asymptotic efficiency relative to the CMLE. This study confirms the same conclusion for bivariate Poisson INGARCH models.

The rest of this paper is organized, as follows. Section 2 constructs the MDPDE for bivariate Poisson INGARCH models. Section 3 shows the asymptotic properties of the MDPDE. Section 4 conducts empirical studies to evaluate the performance of the MDPDE. Section 5 provides concluding remarks. Appendix A provides the proof.

2. MDPDE for Bivariate Poisson Ingarch Models

Basu et al. [22] defined the density power divergence d_α between two densities f and g, with a tuning parameter α, as

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

For a parametric family $\{F_\theta; \ \theta \in \Theta\}$ having densities $\{f_\theta\}$ and a distribution G with density g, they defined the minimum density power divergence functional $T_\alpha(G)$ by $d_\alpha(g, f_{T_\alpha(G)}) = \min_{\theta \in \Theta} d_\alpha(g, f_\theta)$. If G belongs to $\{F_\theta\}$, which is, $G = F_{\theta^0}$ for some $\theta^0 \in \Theta$, then $T_\alpha(F_{\theta^0}) = \theta^0$. Let g be the density function of a random sample $Y_1, \ldots, Y_n$. Using the empirical distribution G_n to approximate G, Basu et al. [22] defined the MDPDE by

$$\hat{\theta}_{\alpha,n} = \underset{\theta \in \Theta}{\arg\min} \ H_{\alpha,n}(\theta),$$

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

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

The tuning parameter α controls the trade-off between the robustness and asymptotic efficiency of the MDPDE. Namely, relatively large α values improve the robustness but the estimator's efficiency decreases. The MDPDE with $\alpha = 0$ and 1 leads to the MLE and L_2-distance estimator, respectively. Basu et al. [22] showed the consistency and asymptotic normality of the MDPDE and demonstrated that the estimator is robust against outliers, but it still retains high efficiency when the true distribution belongs to a parametric family $\{F_\theta\}$ and α is close to zero.

We need to define the conditional version of the MDPDE in order to apply the above procedure to bivariate Poisson INGARCH models. Let $\{f_\theta(\cdot|\mathcal{F}_{t-1})\}$ denote the parametric family of autoregressive models, being indexed by the parameter θ, and let $f_{\theta^0}(\cdot|\mathcal{F}_{t-1})$ 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 $Y_{t-1}, Y_{t-2}, \ldots$. Subsequently, the MDPDE of θ^0 is given by

$$\hat{\theta}_{\alpha,n} = \underset{\theta \in \Theta}{\arg\min} \ H_{\alpha,n}(\theta),$$

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

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

(cf. Section 2 of Kang and Lee [23]).

Let $\boldsymbol{Y}_t = (Y_{t,1}, Y_{t,2})^T$ be a two-dimensional vector of counts at time t, namely, $\{Y_{t,1}, t \geq 1\}$ and $\{Y_{t,2}, t \geq 1\}$ are the two time series of counts under consideration. Liu [17] proposed the bivariate Poisson INGARCH model, as follows

$$
\boldsymbol{Y}_t|\mathcal{F}_{t-1} \sim BP^*(\lambda_{t,1}, \lambda_{t,2}, \phi), \quad \boldsymbol{\lambda}_t = (\lambda_{t,1}, \lambda_{t,2})^T = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1},
$$

where $\mathcal{F}_t$ is the σ-field generated by $\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \dots$, $\phi \geq 0$, $\boldsymbol{\omega} = (\omega_1, \omega_2)^T \in \mathbb{R}_+^2$, $\boldsymbol{A} = \{a_{ij}\}_{i,j=1,2}$ and $\boldsymbol{B} = \{b_{ij}\}_{i,j=1,2}$ are 2×2 matrices with non-negative entries. $BP^*(\lambda_{t,1}, \lambda_{t,2}, \phi)$ denotes the bivariate Poisson distribution constructed via the trivariate reduction method, whose probability mass function (PMF) is

$$
\begin{aligned}
&P(Y_{t,1} = y_1, Y_{t,2} = y_2|\mathcal{F}_{t-1}) \\
=\ & e^{-(\lambda_{t,1}+\lambda_{t,2}-\phi)}\frac{(\lambda_{t,1}-\phi)^{y_1}}{y_1!}\frac{(\lambda_{t,2}-\phi)^{y_2}}{y_2!}\sum_{s=0}^{\min(y_1,y_2)}\binom{y_1}{s}\binom{y_2}{s}s!\left\{\frac{\phi}{(\lambda_{t,1}-\phi)(\lambda_{t,2}-\phi)}\right\}^s.
\end{aligned}
$$

In this model, $Cov(Y_{t,1}, Y_{t,2}|\mathcal{F}_{t-1}) = \phi \in [0, \min(\lambda_{t,1}, \lambda_{t,2}))$, so that the model has a drawback that it can only deal with positive correlation between two components.

To overcome this defect, Cui and Zhu [20] proposed a new bivariate Poisson IN-GARCH model using the distribution that was proposed by Lakshminarayana et al. [21]. They considered the model:

$$
\boldsymbol{Y}_t|\mathcal{F}_{t-1} \sim BP(\lambda_{t,1}, \lambda_{t,2}, \delta), \quad \boldsymbol{\lambda}_t = (\lambda_{t,1}, \lambda_{t,2})^T = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1} \tag{2}
$$

and $BP(\lambda_{t,1}, \lambda_{t,2}, \delta)$ is the bivariate Poisson distribution constructed as a product of Poisson marginals with a multiplicative factor, whose PMF is given by

$$
\begin{aligned}
&P(Y_{t,1} = y_1, Y_{t,2} = y_2|\mathcal{F}_{t-1}) \\
=\ & \frac{\lambda_{t,1}^{y_1}\lambda_{t,2}^{y_2}}{y_1!y_2!}e^{-(\lambda_{t,1}+\lambda_{t,2})}\left\{1 + \delta(e^{-y_1} - e^{-c\lambda_{t,1}})(e^{-y_2} - e^{-c\lambda_{t,2}})\right\},
\end{aligned} \tag{3}
$$

where $c = 1 - e^{-1}$. The marginal conditional distribution of $Y_{t,1}$ and $Y_{t,2}$ are Poisson with parameters $\lambda_{t,1}$ and $\lambda_{t,2}$, respectively, and $Cov(Y_{t,1}, Y_{t,2}|\mathcal{F}_{t-1}) = \delta c^2 \lambda_{t,1}\lambda_{t,2}e^{-c(\lambda_{t,1}+\lambda_{t,2})}$. Hence, this model supports positive or negative correlation, depending on the multiplicative factor parameter δ. Cui and Zhu [20] established the stationarity and ergodicity of the model under certain conditions and showed the consistency and asymptotic normality of the CMLE.

In this study, we apply the MDPDE to the model (2). We focus on the case that $\boldsymbol{A}$ is a diagonal matrix, because this simplification can reduce the number of model parameters and makes it easy to use in practice, as Heinen and Rengifo [31] suggested. Further, the diagonal setup of $\boldsymbol{A}$ eases the verification of the asymptotic properties of the MDPDE. Similar approaches can be found in Liu [17], Lee et al. [19], and Cui et al. [32]. Let $\boldsymbol{A} = diag(a_1, a_2)$. Subsequently, we set $\theta = (\theta_1^T, \theta_2^T, \delta)^T$, where $\theta_1 = (\omega_1, a_1, b_{11}, b_{12})^T$ and $\theta_2 = (\omega_2, a_2, b_{21}, b_{22})^T$, and write the true parameter as $\theta^0 = (\theta_1^{0T}, \theta_2^{0T}, \delta^0)^T$, where $\theta_1^0 = (\omega_1^0, a_1^0, b_{11}^0, b_{12}^0)^T$ and $\theta_2^0 = (\omega_2^0, a_2^0, b_{21}^0, b_{22}^0)^T$.

Given $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ that is generated from (2), from (1), we obtain the MDPDE of θ^0 by

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

where

$$
\tilde{h}_{\alpha,t}(\theta) = \begin{cases} \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta^{1+\alpha}(\boldsymbol{y}|\tilde{\boldsymbol{\lambda}}_t) - \left(1 + \frac{1}{\alpha}\right) f_\theta^\alpha(\boldsymbol{Y}_t|\tilde{\boldsymbol{\lambda}}_t), & \alpha > 0, \\ -\log f_\theta(\boldsymbol{Y}_t|\tilde{\boldsymbol{\lambda}}_t), & \alpha = 0, \end{cases} \tag{4}
$$

$f_\theta(\boldsymbol{y}|\boldsymbol{\lambda}_t)$ for $\boldsymbol{y} = (y_1, y_2)^T$ is the conditional PMF in (3), and $\tilde{\boldsymbol{\lambda}}_t$ is recursively defined by

$$
\tilde{\boldsymbol{\lambda}}_t = (\tilde{\lambda}_{t,1}, \tilde{\lambda}_{t,2})^T = \boldsymbol{\omega} + \boldsymbol{A}\tilde{\boldsymbol{\lambda}}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1}, \ \ t \geq 2
$$

with an arbitrarily chosen initial value $\tilde{\boldsymbol{\lambda}}_1$. We also use notations $\boldsymbol{\lambda}_t(\theta)$ and $\tilde{\boldsymbol{\lambda}}_t(\theta)$ to denote $\boldsymbol{\lambda}_t$ and $\tilde{\boldsymbol{\lambda}}_t$, respectively, in order to emphasize the role of θ.

3. Asymptotic Properties of the MDPDE

In this section, we establish the consistency and asymptotic normality of the MDPDE. Throughout this study, $\|\boldsymbol{A}\|_p$ denotes the p-induced norm of matrix $\boldsymbol{A}$ for $1 \leq p \leq \infty$ and $\|\boldsymbol{x}\|_p$ is the p-norm of vector $\boldsymbol{x}$. When $p = 1$ and ∞, $\|\boldsymbol{A}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$ and $\|\boldsymbol{A}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$ for $\boldsymbol{A} = \{a_{ij}\}_{1 \leq i \leq m, 1 \leq j \leq n}$, respectively. $E(\cdot)$ is taken under θ^0. We assume that the following conditions hold in order to verify the asymptotic properties of the MDPDE.

(A1) θ_1^0, θ_2^0, and δ^0 are interior points in the compact parameter spaces Θ_1, Θ_2, and Θ_3, respectively, and $\Theta = \Theta_1 \times \Theta_2 \times \Theta_3$. In addition, there exist positive constants ω_L, ω_U, a_L, a_U, b_L, b_U, and δ_U, such that for $i, j = 1, 2$,

$$
0 < \omega_L \leq \omega_i \leq \omega_U, \ 0 < a_L \leq a_i \leq a_U, \ 0 < b_L \leq b_{ij} \leq b_U, \ \text{and} \ |\delta| \leq \delta_U.
$$

(A2) There exist positive constants φ_L and φ_U such that for $\boldsymbol{y} = (y_1, y_2)^T \in \mathbb{N}_0^2$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T \in (0, \infty)^2$, and $\delta \in \Theta_3$,

$$
0 < \varphi_L \leq \varphi(\boldsymbol{y}, \boldsymbol{\lambda}, \delta) \leq \varphi_U, \ \text{where} \ \varphi(\boldsymbol{y}, \boldsymbol{\lambda}, \delta) = 1 + \delta(e^{-y_1} - e^{-c\lambda_1})(e^{-y_2} - e^{-c\lambda_2}).
$$

(A3) There exists a $p \in [1, \infty]$ such that $\|\boldsymbol{A}\|_p + 2^{(1-1/p)}\|\boldsymbol{B}\|_p < 1$.

Remark 1. *These conditions can be found in Cui and Zhu [20]. According to Theorem 1 in their study, $\{(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t)\}$ is stationary and ergodic under (A1) and (A3).*

Subsequently, we obtain the following results; the proofs are provided in the Appendix A.

Theorem 1. *Under the conditions (A1)–(A3),*

$$
\hat{\theta}_{\alpha,n} \xrightarrow{a.s.} \theta^0 \ \text{as} \ n \to \infty.
$$

Theorem 2. *Under the conditions (A1)–(A3),*

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

where

$$
J_\alpha = -E\left(\frac{\partial^2 h_{\alpha,t}(\theta^0)}{\partial\theta\partial\theta^T}\right), \ K_\alpha = E\left(\frac{\partial h_{\alpha,t}(\theta^0)}{\partial\theta}\frac{\partial h_{\alpha,t}(\theta^0)}{\partial\theta^T}\right),
$$

and $h_{\alpha,t}(\theta)$ is defined by replacing $\tilde{\boldsymbol{\lambda}}_t(\theta)$ with $\boldsymbol{\lambda}_t(\theta)$ in (4).

Remark 2. *Because the tuning parameter α controls the trade-off between the robustness and asymptotic efficiency, choosing the optimal α is an important issue in practice. Several researchers investigated the selection criterion of optimal α; see Fujisawa and Eguchi [33], Durio and Isaia [34],*

and Toma and Broniatowski [35]. Among them, we adopt the method of Warwick [36] to choose α that minimizes the trace of the estimated asymptotic mean squared error $(\widehat{AMSE})$ defined by

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

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

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

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

4. Empirical Studies

4.1. Simulation

In this section, we report the simulation results to evaluate the performance of the MDPDE. The simulation settings are described, as follows. Using the inverse transformation sampling method (cf. Section 2.3 of Verges [37]), we generate $Y_1, \ldots, Y_n$ from (2) with the initial value $\lambda_1 = (0,0)^T$. For the estimation, $\tilde{\lambda}_1$ is set to be the sample mean of the data. We first consider $\theta = (\omega_1, a_1, b_{11}, b_{12}, \omega_2, a_2, b_{21}, b_{22}, \delta)^T = (1, 0.2, 0.1, 0.2, 0.5, 0.3, 0.4, 0.2, 0.5)^T$, which satisfies (A3) with $p = 1$. In this simulation, we compare the performance of the MDPDE with $\alpha > 0$ with that of the CMLE ($\alpha = 0$). We examine the sample mean, variance, and mean squared error (MSE) of the estimators. The sample size under consideration is $n = 1000$ and the number of repetitions for each simulation is 1000. In Tables 1–16, the symbol * represents the minimal MSEs for each parameter.

Table 1 indicates that, when the data are not contaminated by outliers, the CMLE exhibits minimal MSEs for all parameters, and the MSEs of the MDPDE with small α are close to those of the CMLE. The MSE of the MDPDE shows an increasing tendency as α increases. Hence, we can conclude that the CMLE outperforms the MDPDE in the absence of outliers.

Table 1. Sample mean, variance, and mean squared error (MSE) of estimators when $\theta = (1, 0.2, 0.1, 0.2, 0.5, 0.3, 0.4, 0.2, 0.5)^T$, $n = 1000$, and no outliers exist.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	1.010	0.198	0.099	0.199	0.510	0.298	0.401	0.198	0.583
	Var $\times 10^2$	3.421	1.062	0.105	0.083	1.344	0.366	0.119	0.100	15.54
	MSE $\times 10^2$	3.429 *	1.061 *	0.105 *	0.083 *	1.352 *	0.366 *	0.119 *	0.100 *	16.22 *
0.1	Mean	1.012	0.198	0.099	0.199	0.510	0.297	0.401	0.199	0.577
	Var $\times 10^2$	3.527	1.091	0.108	0.083	1.379	0.372	0.121	0.103	15.83
	MSE $\times 10^2$	3.537	1.091	0.108	0.084	1.387	0.372	0.121	0.103	16.41
0.2	Mean	1.013	0.197	0.099	0.199	0.510	0.297	0.401	0.199	0.572
	Var $\times 10^2$	3.671	1.134	0.113	0.086	1.453	0.387	0.126	0.108	16.42
	MSE $\times 10^2$	3.684	1.134	0.113	0.086	1.463	0.388	0.126	0.108	16.92
0.3	Mean	1.013	0.197	0.099	0.199	0.511	0.296	0.401	0.199	0.568
	Var $\times 10^2$	3.870	1.195	0.120	0.090	1.555	0.410	0.133	0.114	17.22
	MSE $\times 10^2$	3.883	1.195	0.120	0.090	1.565	0.411	0.133	0.114	17.66
0.5	Mean	1.012	0.197	0.100	0.199	0.511	0.294	0.402	0.200	0.559
	Var $\times 10^2$	4.336	1.340	0.137	0.101	1.817	0.469	0.151	0.130	19.51
	MSE $\times 10^2$	4.347	1.340	0.137	0.101	1.828	0.472	0.152	0.130	19.84
1	Mean	1.007	0.198	0.101	0.200	0.513	0.289	0.405	0.203	0.544
	Var $\times 10^2$	6.094	1.864	0.198	0.148	2.805	0.690	0.222	0.189	29.18
	MSE $\times 10^2$	6.094	1.863	0.198	0.148	2.818	0.701	0.224	0.190	29.35

Now, we consider the situation that the data are contaminated by outliers. To this end, we generate contaminated data $Y_{c,t} = (Y_{c,t,1}, Y_{c,t,2})^T$ when considering

$$Y_{c,t,i} = Y_{t,i} + P_{t,i} Y_{o,t,i}, \quad i = 1, 2,$$

where $Y_{t,i}$ are generated from (2), $P_{t,i}$ are i.i.d. Bernoulli random variables with success probability p, and $Y_{o,t,i}$ are i.i.d. Poisson random variables with mean γ. We consider three

cases: $(p, \gamma) = (0.03, 5)$, $(0.03, 10)$, and $(0.05, 10)$. Tables 2–4 report the results. In the tables, the MDPDE appears to have smaller MSEs than the CMLE for all cases, except for the case of $\alpha = 1$ when $(p, \gamma) = (0.03, 5)$. As p or γ increases, the MSEs of the CMLE increase faster than those of the MDPDE, which indicates that the MDPDE outperforms the CMLE, as the data are more contaminated by outliers. Moreover, as p or γ increases, the symbol * tends to move downward. This indicates that, when the data are severely contaminated by outliers, the MDPDE with large α performs better.

Table 2. Sample mean, variance, and MSE of estimators when $\theta = (1, 0.2, 0.1, 0.2, 0.5, 0.3, 0.4, 0.2, 0.5)^T$, $n = 1000$, and $(p, \gamma) = (0.03, 5)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	1.073	0.266	0.077	0.167	0.650	0.339	0.325	0.176	0.728
	Var $\times 10^2$	6.363	1.707	0.109	0.105	2.333	0.553	0.168	0.115	17.16
	MSE $\times 10^2$	6.897	2.140	0.160	0.213	4.577	0.704	0.736	0.170	22.36
0.1	Mean	1.028	0.264	0.080	0.170	0.607	0.335	0.331	0.179	0.697
	Var $\times 10^2$	5.299	1.510	0.098	0.097	2.040	0.512	0.160	0.108	17.23
	MSE $\times 10^2$	5.375	1.915	0.139	0.188	3.185	0.635	0.636	0.151	21.09
0.2	Mean	1.008	0.261	0.081	0.171	0.587	0.331	0.335	0.181	0.679
	Var $\times 10^2$	5.114	1.491	0.098	0.097	2.031	0.526	0.165	0.110	17.70
	MSE $\times 10^2$	5.116 *	1.855	0.133	0.179	2.789	0.621 *	0.583	0.147 *	20.87 *
0.3	Mean	1.000	0.257	0.083	0.172	0.578	0.327	0.339	0.182	0.662
	Var $\times 10^2$	5.182	1.526	0.101	0.100	2.099	0.558	0.177	0.115	18.34
	MSE $\times 10^2$	5.177	1.846 *	0.131 *	0.177 *	2.701 *	0.628	0.548	0.148	20.95
0.5	Mean	0.997	0.248	0.086	0.174	0.572	0.317	0.346	0.184	0.633
	Var $\times 10^2$	5.729	1.682	0.114	0.116	2.381	0.658	0.220	0.136	20.02
	MSE $\times 10^2$	5.724	1.910	0.134	0.183	2.899	0.686	0.516	0.162	21.77
1	Mean	1.007	0.230	0.094	0.179	0.578	0.296	0.363	0.191	0.587
	Var $\times 10^2$	7.297	2.213	0.166	0.168	3.435	0.965	0.315	0.205	29.90
	MSE $\times 10^2$	7.294	2.301	0.170	0.210	4.039	0.966	0.449 *	0.214	30.62

Table 3. Sample mean, variance, and MSE of estimators when $\theta = (1, 0.2, 0.1, 0.2, 0.5, 0.3, 0.4, 0.2, 0.5)^T$, $n = 1000$, and $(p, \gamma) = (0.03, 10)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	1.141	0.349	0.052	0.123	0.846	0.398	0.230	0.141	1.113
	Var $\times 10^2$	16.43	3.478	0.101	0.140	5.886	1.087	0.265	0.138	21.91
	MSE $\times 10^2$	18.39	5.702	0.335	0.736	17.88	2.051	3.138	0.487	59.51
0.1	Mean	1.015	0.329	0.057	0.131	0.706	0.382	0.248	0.150	0.865
	Var $\times 10^2$	7.844	2.031	0.069	0.095	3.087	0.672	0.224	0.100	19.42
	MSE $\times 10^2$	7.860	3.703	0.250	0.566	7.329	1.348	2.523	0.355	32.72
0.2	Mean	0.995	0.314	0.060	0.134	0.680	0.365	0.259	0.153	0.802
	Var $\times 10^2$	7.073	1.948	0.068	0.095	2.912	0.677	0.244	0.104	19.42
	MSE $\times 10^2$	7.068	3.252	0.225	0.529	6.156 *	1.105	2.245	0.321	28.54
0.3	Mean	1.002	0.298	0.064	0.137	0.681	0.349	0.269	0.157	0.765
	Var $\times 10^2$	6.995	1.972	0.075	0.102	3.030	0.742	0.280	0.114	19.94
	MSE $\times 10^2$	6.989 *	2.936	0.207	0.499	6.287	0.977	2.005	0.301	26.92
0.5	Mean	1.034	0.264	0.072	0.145	0.695	0.314	0.293	0.165	0.706
	Var $\times 10^2$	7.365	2.137	0.097	0.125	3.415	0.913	0.382	0.146	21.81
	MSE $\times 10^2$	7.475	2.545	0.176	0.430	7.223	0.932 *	1.536	0.266 *	26.01 *
1	Mean	1.088	0.198	0.095	0.167	0.719	0.242	0.353	0.191	0.604
	Var $\times 10^2$	7.825	2.377	0.171	0.203	4.553	1.273	0.601	0.258	30.55
	MSE $\times 10^2$	8.592	2.375 *	0.173 *	0.309 *	9.328	1.611	0.818*	0.267	31.61

Table 4. Sample mean, variance, and MSE of estimators when $\theta = (1, 0.2, 0.1, 0.2, 0.5, 0.3, 0.4, 0.2, 0.5)^T$, $n = 1000$, and $(p, \gamma) = (0.05, 10)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	1.223	0.404	0.040	0.093	0.990	0.449	0.167	0.114	1.635
	Var $\times 10^2$	28.47	4.763	0.086	0.128	11.74	1.691	0.229	0.131	29.21
	MSE $\times 10^2$	33.40	8.909	0.442	1.281	35.70	3.897	5.645	0.867	158.1
0.1	Mean	1.012	0.390	0.046	0.103	0.772	0.437	0.185	0.125	1.057
	Var $\times 10^2$	11.78	2.695	0.056	0.083	4.883	0.952	0.188	0.095	21.44
	MSE $\times 10^2$	11.78	6.291	0.349	1.031	12.27	2.820	4.823	0.661	52.48
0.2	Mean	0.967	0.377	0.048	0.105	0.724	0.421	0.192	0.128	0.935
	Var $\times 10^2$	9.531	2.414	0.052	0.080	4.163	0.896	0.203	0.093	20.74
	MSE $\times 10^2$	9.633	5.529	0.324	0.986	9.168	2.359	4.525	0.608	39.63
0.3	Mean	0.971	0.361	0.050	0.107	0.720	0.405	0.199	0.131	0.879
	Var $\times 10^2$	9.450	2.465	0.055	0.086	4.189	0.962	0.236	0.101	20.90
	MSE $\times 10^2$	9.526 *	5.040	0.308	0.953	9.029 *	2.068	4.296	0.578	35.21
0.5	Mean	1.004	0.327	0.056	0.113	0.741	0.369	0.217	0.138	0.801
	Var $\times 10^2$	9.878	2.724	0.071	0.112	4.689	1.209	0.363	0.132	22.32
	MSE $\times 10^2$	9.870	4.336	0.269	0.861	10.51	1.687 *	3.700	0.511	31.33 *
1	Mean	1.102	0.229	0.084	0.142	0.807	0.257	0.300	0.170	0.651
	Var $\times 10^2$	10.28	3.134	0.183	0.238	5.959	1.804	0.946	0.304	30.79
	MSE $\times 10^2$	11.32	3.214 *	0.208 *	0.574 *	15.35	1.990	1.936 *	0.392 *	33.03

We also consider smaller sample size $n = 200$. The results are presented in Tables 5–8 and they show results similar to those in Tables 1–4. The variances and MSEs of both the CMLE and MDPDE are larger than those in Tables 1–4.

Table 5. Sample mean, variance, and MSE of estimators when $\theta = (1, 0.2, 0.1, 0.2, 0.5, 0.3, 0.4, 0.2, 0.5)^T$, $n = 200$, and no outliers exist.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	1.005	0.208	0.089	0.199	0.541	0.281	0.411	0.195	0.893
	Var $\times 10^2$	12.41	3.866	0.426	0.394	7.816	2.078	0.651	0.553	71.23
	MSE $\times 10^2$	12.40 *	3.869 *	0.437 *	0.394 *	7.973 *	2.112 *	0.663 *	0.555 *	86.57
0.1	Mean	0.975	0.203	0.087	0.193	0.529	0.271	0.400	0.191	0.786
	Var $\times 10^2$	14.98	3.919	0.439	0.498	8.317	2.212	1.097	0.649	59.99
	MSE $\times 10^2$	15.03	3.916	0.455	0.502	8.392	2.296	1.096	0.658	68.12
0.2	Mean	0.970	0.203	0.087	0.192	0.527	0.267	0.400	0.191	0.756
	Var $\times 10^2$	15.48	3.965	0.458	0.520	8.672	2.292	1.176	0.687	60.78
	MSE $\times 10^2$	15.55	3.962	0.473	0.526	8.734	2.396	1.174	0.695	67.27 *
0.3	Mean	0.962	0.204	0.088	0.191	0.525	0.263	0.400	0.191	0.730
	Var $\times 10^2$	16.41	4.166	0.477	0.555	9.040	2.366	1.274	0.734	63.50
	MSE $\times 10^2$	16.54	4.163	0.492	0.563	9.096	2.497	1.273	0.741	68.71
0.5	Mean	0.945	0.202	0.088	0.188	0.521	0.254	0.398	0.192	0.685
	Var $\times 10^2$	18.64	4.513	0.527	0.653	10.34	2.653	1.561	0.873	70.39
	MSE $\times 10^2$	18.93	4.509	0.540	0.666	10.38	2.863	1.560	0.879	73.75
1	Mean	0.968	0.209	0.102	0.204	0.537	0.249	0.433	0.213	0.684
	Var $\times 10^2$	18.37	5.307	0.757	0.817	11.99	3.117	1.327	1.159	135.3
	MSE $\times 10^2$	18.45	5.310	0.757	0.817	12.12	3.374	1.437	1.175	138.5

Table 6. Sample mean, variance and MSE of estimators when $\theta = (1, 0.2, 0.1, 0.2, 0.5, 0.3, 0.4, 0.2, 0.5)^T$, $n = 200$, and $(p, \gamma) = (0.03, 5)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	1.056	0.276	0.077	0.164	0.662	0.324	0.339	0.173	1.054
	Var $\times 10^2$	20.83	5.501	0.419	0.489	12.34	2.785	0.918	0.619	80.49
	MSE $\times 10^2$	21.13	6.078	0.471	0.616 *	14.94	2.839	1.292 *	0.690 *	111.2
0.1	Mean	0.992	0.262	0.077	0.163	0.605	0.311	0.334	0.171	0.925
	Var $\times 10^2$	20.38	5.118	0.411	0.510	11.65	2.801	1.153	0.641	67.19
	MSE $\times 10^2$	20.37	5.496	0.463 *	0.648	12.75	2.810 *	1.581	0.724	85.15
0.2	Mean	0.973	0.253	0.079	0.165	0.585	0.305	0.338	0.172	0.882
	Var $\times 10^2$	19.71	4.993	0.422	0.525	11.55	2.817	1.207	0.652	68.88
	MSE $\times 10^2$	19.76	5.265	0.465	0.645	12.26 *	2.816	1.594	0.730	83.37
0.3	Mean	0.958	0.247	0.081	0.165	0.577	0.296	0.340	0.172	0.840
	Var $\times 10^2$	19.93	5.028	0.445	0.563	12.33	2.962	1.321	0.690	70.67
	MSE $\times 10^2$	20.09	5.244	0.483	0.682	12.90	2.961	1.681	0.766	82.17 *
0.5	Mean	0.944	0.234	0.084	0.167	0.572	0.281	0.344	0.174	0.774
	Var $\times 10^2$	20.94	5.080	0.503	0.647	13.53	3.241	1.574	0.806	78.15
	MSE $\times 10^2$	21.23	5.193 *	0.528	0.756	14.04	3.273	1.885	0.873	85.55
1	Mean	0.960	0.236	0.101	0.187	0.592	0.266	0.388	0.198	0.770
	Var $\times 10^2$	19.00	5.571	0.755	0.859	15.57	3.851	1.689	1.119	147.0
	MSE $\times 10^2$	19.14 *	5.696	0.754	0.876	16.40	3.962	1.702	1.119	154.2

Table 7. Sample mean, variance, and MSE of estimators when $\theta = (1, 0.2, 0.1, 0.2, 0.5, 0.3, 0.4, 0.2, 0.5)^T$, $n = 200$, and $(p, \gamma) = (0.03, 10)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	1.128	0.349	0.052	0.126	0.860	0.388	0.241	0.135	1.365
	Var $\times 10^2$	38.79	8.126	0.345	0.618	26.05	4.690	1.145	0.746	96.45
	MSE $\times 10^2$	40.38	10.33	0.574	1.161	38.95	5.467	3.659	1.174	171.2
0.1	Mean	1.003	0.314	0.054	0.128	0.715	0.355	0.250	0.141	1.050
	Var $\times 10^2$	28.42	6.643	0.269	0.507	16.99	3.644	1.158	0.616	69.06
	MSE $\times 10^2$	28.39	7.925	0.480	1.021	21.59	3.938	3.403	0.961	99.19
0.2	Mean	0.980	0.296	0.057	0.130	0.679	0.337	0.258	0.146	0.953
	Var $\times 10^2$	26.04	6.348	0.270	0.505	15.82	3.612	1.262	0.628	67.71
	MSE $\times 10^2$	26.05	7.268	0.455 *	0.991	19.02	3.749	3.268	0.914	88.19
0.3	Mean	0.972	0.289	0.060	0.133	0.678	0.320	0.270	0.151	0.893
	Var $\times 10^2$	25.20	6.357	0.299	0.535	15.69	3.649	1.407	0.660	69.05
	MSE $\times 10^2$	25.26	7.142	0.457	0.987 *	18.84 *	3.683 *	3.096	0.894 *	84.43
0.5	Mean	0.974	0.264	0.070	0.139	0.673	0.287	0.294	0.160	0.783
	Var $\times 10^2$	24.72	6.143	0.399	0.643	16.00	3.836	1.847	0.794	75.64
	MSE $\times 10^2$	24.76	6.548	0.490	1.019	18.96	3.848	2.963	0.953	83.56 *
1	Mean	1.007	0.232	0.100	0.171	0.677	0.235	0.374	0.200	0.657
	Var $\times 10^2$	21.91	6.221	0.778	1.007	16.89	3.717	2.460	1.238	130.0
	MSE $\times 10^2$	21.89 *	6.319 *	0.777	1.088	20.01	4.133	2.526 *	1.237	132.3

Table 8. Sample mean, variance, and MSE of estimators when $\theta = (1, 0.2, 0.1, 0.2, 0.5, 0.3, 0.4, 0.2, 0.5)^T$, $n = 200$, and $(p, \gamma) = (0.05, 10)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	1.171	0.406	0.046	0.097	1.041	0.420	0.183	0.108	1.814
	Var $\times 10^2$	53.26	9.255	0.326	0.521	45.18	6.131	1.054	0.654	133.0
	MSE $\times 10^2$	56.14	13.48	0.619	1.572	74.38	7.569	5.761	1.504	305.4
0.1	Mean	1.037	0.347	0.047	0.102	0.821	0.389	0.192	0.117	1.203
	Var $\times 10^2$	36.17	7.578	0.227	0.430	26.89	4.810	1.034	0.549	80.49
	MSE $\times 10^2$	36.28	9.719	0.509	1.388	37.17	5.600	5.372	1.244	129.8
0.2	Mean	0.989	0.334	0.049	0.104	0.772	0.370	0.199	0.122	1.064
	Var $\times 10^2$	31.43	7.373	0.218	0.421	23.29	4.607	1.106	0.554	77.26
	MSE $\times 10^2$	31.41	9.171	0.477	1.344	30.69	5.097	5.144	1.156	108.9
0.3	Mean	0.989	0.320	0.051	0.106	0.762	0.355	0.207	0.126	0.984
	Var $\times 10^2$	30.35	7.338	0.234	0.443	22.64	4.685	1.247	0.602	76.99
	MSE $\times 10^2$	30.33	8.773	0.472 *	1.327	29.47	4.985	4.985	1.149 *	100.4
0.5	Mean	0.984	0.293	0.058	0.112	0.764	0.314	0.229	0.135	0.855
	Var $\times 10^2$	30.12	7.263	0.332	0.558	22.81	4.884	1.791	0.781	80.40
	MSE $\times 10^2$	30.12	8.122	0.505	1.331	29.73	4.897	4.726	1.206	92.95 *
1	Mean	1.046	0.239	0.097	0.151	0.774	0.243	0.333	0.178	0.696
	Var $\times 10^2$	23.99	6.497	0.805	1.059	21.95	4.517	3.261	1.366	136.2
	MSE $\times 10^2$	24.17 *	6.645 *	0.805	1.302 *	29.46 *	4.839 *	3.708 *	1.413	139.9

In order to evaluate the performance of the MDPDE for negatively cross-correlated data, we consider $\theta = (\omega_1, a_1, b_{11}, b_{12}, \omega_2, a_2, b_{21}, b_{22}, \delta)^T = (0.5, 0.1, 0.2, 0.4, 0.3, 0.3, 0.2, 0.1, -0.4)^T$ with the same p and γ, as above. The results are reported in Tables 9–16 for $n = 1000$ and 200, respectively. These tables exhibit results that are similar to those in Tables 1–8. Overall, our findings strongly support the assertion that the MDPDE is a functional tool for yielding a robust estimator for bivariate Poisson INGARCH models in the presence of outliers.

Table 9. Sample mean, variance, and MSE of estimators when $\theta = (0.5, 0.1, 0.2, 0.4, 0.3, 0.3, 0.2, 0.1, -0.4)^T$, $n = 1000$, and no outliers exist.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	0.501	0.103	0.199	0.397	0.306	0.296	0.200	0.098	−0.385
	Var $\times 10^2$	0.570	0.508	0.105	0.159	0.518	1.029	0.080	0.108	6.231
	MSE $\times 10^2$	0.569 *	0.508 *	0.105 *	0.160 *	0.522 *	1.030 *	0.080 *	0.109 *	6.247 *
0.1	Mean	0.501	0.103	0.199	0.397	0.306	0.295	0.200	0.098	−0.384
	Var $\times 10^2$	0.578	0.515	0.107	0.160	0.530	1.040	0.082	0.111	6.347
	MSE $\times 10^2$	0.578	0.515	0.108	0.161	0.534	1.041	0.082	0.112	6.367
0.2	Mean	0.501	0.103	0.199	0.397	0.307	0.295	0.200	0.098	−0.383
	Var $\times 10^2$	0.600	0.532	0.113	0.166	0.556	1.082	0.086	0.117	6.564
	MSE $\times 10^2$	0.600	0.533	0.113	0.167	0.560	1.083	0.086	0.117	6.588
0.3	Mean	0.501	0.104	0.199	0.397	0.307	0.294	0.200	0.098	−0.381
	Var $\times 10^2$	0.627	0.554	0.119	0.175	0.591	1.145	0.092	0.124	6.848
	MSE $\times 10^2$	0.627	0.555	0.119	0.176	0.595	1.147	0.092	0.125	6.876
0.5	Mean	0.500	0.105	0.198	0.398	0.308	0.292	0.201	0.099	−0.380
	Var $\times 10^2$	0.702	0.615	0.137	0.199	0.685	1.320	0.106	0.142	7.577
	MSE $\times 10^2$	0.701	0.617	0.137	0.200	0.690	1.325	0.106	0.142	7.610
1	Mean	0.495	0.110	0.198	0.399	0.310	0.287	0.203	0.100	−0.382
	Var $\times 10^2$	0.972	0.839	0.201	0.290	0.942	1.864	0.155	0.195	10.09
	MSE $\times 10^2$	0.974	0.848	0.201	0.290	0.951	1.878	0.156	0.194	10.12

Table 10. Sample mean, variance, and MSE of estimators when $\theta = (0.5, 0.1, 0.2, 0.4, 0.3, 0.3, 0.2, 0.1, -0.4)^T$, $n = 1000$, and $(p, \gamma) = (0.03, 5)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	0.633	0.194	0.143	0.269	0.399	0.368	0.147	0.064	−0.097
	Var $\times 10^2$	1.794	1.343	0.152	0.263	1.741	2.315	0.125	0.123	5.603
	MSE $\times 10^2$	3.560	2.219	0.474	1.974	2.728	2.769	0.409	0.255	14.79
0.1	Mean	0.572	0.186	0.149	0.280	0.350	0.360	0.153	0.067	−0.143
	Var $\times 10^2$	1.191	1.013	0.126	0.235	1.047	1.659	0.100	0.094	5.787
	MSE $\times 10^2$	1.711	1.743	0.390	1.676	1.297	2.016	0.325	0.205	12.38
0.2	Mean	0.550	0.177	0.151	0.286	0.335	0.350	0.155	0.068	−0.169
	Var $\times 10^2$	1.082	0.958	0.124	0.240	0.950	1.608	0.100	0.090	6.076
	MSE $\times 10^2$	1.335	1.543	0.361	1.536	1.074 *	1.861 *	0.305	0.191	11.43
0.3	Mean	0.543	0.167	0.154	0.292	0.333	0.340	0.156	0.070	−0.187
	Var $\times 10^2$	1.055	0.950	0.129	0.254	0.976	1.706	0.107	0.095	6.375
	MSE $\times 10^2$	1.241	1.401	0.344	1.427	1.083	1.868	0.297	0.184	10.89
0.5	Mean	0.542	0.148	0.159	0.304	0.339	0.318	0.161	0.075	−0.214
	Var $\times 10^2$	1.050	0.951	0.147	0.290	1.118	2.017	0.125	0.113	7.038
	MSE $\times 10^2$	1.229 *	1.185	0.315	1.203	1.270	2.049	0.279	0.175 *	10.49 *
1	Mean	0.548	0.112	0.176	0.340	0.360	0.268	0.176	0.090	−0.247
	Var $\times 10^2$	1.136	0.953	0.214	0.399	1.425	2.636	0.188	0.184	9.324
	MSE $\times 10^2$	1.363	0.966 *	0.271 *	0.756 *	1.783	2.733	0.244 *	0.194	11.64

Table 11. Sample mean, variance, and MSE of estimators when $\theta = (0.5, 0.1, 0.2, 0.4, 0.3, 0.3, 0.2, 0.1, -0.4)^T$, $n = 1000$, and $(p, \gamma) = (0.03, 10)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	0.774	0.295	0.087	0.149	0.525	0.415	0.094	0.037	0.336
	Var $\times 10^2$	7.976	4.279	0.176	0.346	7.305	6.070	0.187	0.103	7.338
	MSE $\times 10^2$	15.50	8.069	1.442	6.644	12.37	7.392	1.303	0.501	61.46
0.1	Mean	0.612	0.254	0.100	0.173	0.373	0.399	0.106	0.040	0.019
	Var $\times 10^2$	2.368	2.008	0.104	0.287	1.694	2.548	0.105	0.049	6.426
	MSE $\times 10^2$	3.628	4.369	1.109	5.441	2.231	3.521	0.982	0.406	23.94
0.2	Mean	0.596	0.227	0.103	0.182	0.364	0.378	0.108	0.042	−0.035
	Var $\times 10^2$	2.029	1.866	0.106	0.336	1.567	2.548	0.107	0.050	6.677
	MSE $\times 10^2$	2.944	3.490	1.048	5.081	1.971 *	3.160	0.949	0.384	20.01
0.3	Mean	0.600	0.203	0.108	0.195	0.372	0.355	0.112	0.046	−0.069
	Var $\times 10^2$	1.894	1.823	0.122	0.425	1.601	2.697	0.123	0.060	6.909
	MSE $\times 10^2$	2.884	2.873	0.973	4.637	2.111	2.997	0.889	0.353	17.86
0.5	Mean	0.611	0.145	0.125	0.240	0.401	0.287	0.130	0.061	−0.135
	Var $\times 10^2$	1.518	1.521	0.177	0.691	1.608	2.883	0.181	0.107	7.489
	MSE $\times 10^2$	2.744	1.725	0.742	3.259	2.619	2.898 *	0.674	0.259	14.50
1	Mean	0.594	0.059	0.178	0.360	0.440	0.155	0.185	0.107	−0.249
	Var $\times 10^2$	0.941	0.570	0.268	0.634	1.291	2.214	0.278	0.216	9.962
	MSE $\times 10^2$	1.828 *	0.737 *	0.316 *	0.794 *	3.262	4.327	0.301 *	0.220 *	12.22 *

Table 12. Sample mean, variance, and MSE of estimators when $\theta = (0.5, 0.1, 0.2, 0.4, 0.3, 0.3, 0.2, 0.1, -0.4)^T$, $n = 1000$, and $(p, \gamma) = (0.05, 10)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	0.870	0.382	0.059	0.086	0.645	0.451	0.062	0.027	0.829
	Var $\times 10^2$	17.32	6.408	0.129	0.206	14.69	8.150	0.135	0.073	8.777
	MSE $\times 10^2$	30.96	14.37	2.130	10.09	26.58	10.41	2.026	0.604	159.9
0.1	Mean	0.621	0.346	0.070	0.103	0.396	0.446	0.074	0.029	0.170
	Var $\times 10^2$	4.575	3.255	0.073	0.158	2.948	3.708	0.076	0.036	6.635
	MSE $\times 10^2$	6.034	9.327	1.762	8.971	3.862	5.837	1.659	0.545	39.12
0.2	Mean	0.585	0.327	0.070	0.104	0.370	0.431	0.073	0.029	0.089
	Var $\times 10^2$	3.641	2.988	0.065	0.164	2.294	3.311	0.068	0.031	6.911
	MSE $\times 10^2$	4.360	8.156	1.749	8.898	2.788 *	5.022	1.670	0.540	30.79
0.3	Mean	0.586	0.311	0.071	0.107	0.374	0.417	0.074	0.029	0.058
	Var $\times 10^2$	3.517	3.054	0.072	0.193	2.353	3.531	0.074	0.033	7.089
	MSE $\times 10^2$	4.249	7.500	1.727	8.805	2.893	4.895	1.661	0.532	28.06
0.5	Mean	0.608	0.265	0.080	0.124	0.399	0.371	0.083	0.035	0.016
	Var $\times 10^2$	3.465	3.335	0.114	0.398	2.559	4.119	0.120	0.055	7.515
	MSE $\times 10^2$	4.628	6.044	1.555	8.030	3.537	4.613 *	1.492	0.482	24.83
1	Mean	0.637	0.087	0.148	0.296	0.481	0.161	0.153	0.096	−0.144
	Var $\times 10^2$	1.536	1.591	0.410	1.613	1.732	3.105	0.408	0.306	9.724
	MSE $\times 10^2$	3.424 *	1.606 *	0.682 *	2.695 *	4.999	5.042	0.626 *	0.308 *	16.28 *

Table 13. Sample mean, variance, and MSE of estimators when $\theta = (0.5, 0.1, 0.2, 0.4, 0.3, 0.3, 0.2, 0.1, -0.4)^T$, $n = 200$, and no outliers exist.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	0.487	0.131	0.182	0.394	0.316	0.287	0.203	0.092	−0.313
	Var $\times 10^2$	2.173	2.095	0.526	0.806	2.213	4.245	0.411	0.475	33.35
	MSE $\times 10^2$	2.187 *	2.186 *	0.558 *	0.809 *	2.237 *	4.257 *	0.412 *	0.481 *	34.07
0.1	Mean	0.483	0.129	0.181	0.390	0.314	0.284	0.202	0.091	−0.294
	Var $\times 10^2$	2.391	2.104	0.571	0.958	2.337	4.358	0.452	0.487	31.43
	MSE $\times 10^2$	2.416	2.188	0.609	0.967	2.353	4.379	0.452	0.495	32.53 *
0.2	Mean	0.481	0.131	0.180	0.388	0.312	0.283	0.202	0.090	−0.285
	Var $\times 10^2$	2.542	2.158	0.608	1.040	2.427	4.490	0.484	0.512	31.25
	MSE $\times 10^2$	2.577	2.250	0.649	1.054	2.439	4.513	0.483	0.520	32.55
0.3	Mean	0.479	0.134	0.180	0.388	0.311	0.284	0.202	0.091	−0.284
	Var $\times 10^2$	2.612	2.225	0.636	1.069	2.531	4.657	0.503	0.537	32.24
	MSE $\times 10^2$	2.653	2.337	0.677	1.082	2.541	4.679	0.503	0.545	33.55
0.5	Mean	0.477	0.137	0.180	0.389	0.313	0.280	0.204	0.092	−0.285
	Var $\times 10^2$	2.860	2.423	0.726	1.183	2.681	4.733	0.554	0.601	35.16
	MSE $\times 10^2$	2.909	2.555	0.766	1.194	2.695	4.769	0.555	0.606	36.43
1	Mean	0.473	0.145	0.185	0.399	0.314	0.276	0.212	0.100	−0.324
	Var $\times 10^2$	3.364	3.057	1.016	1.549	3.003	5.321	0.746	0.858	48.77
	MSE $\times 10^2$	3.434	3.252	1.038	1.548	3.021	5.375	0.760	0.857	49.31

Table 14. Sample mean, variance, and MSE of estimators when $\theta = (0.5, 0.1, 0.2, 0.4, 0.3, 0.3, 0.2, 0.1, -0.4)^T$, $n = 200$, and $(p, \gamma) = (0.03, 5)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	0.611	0.215	0.138	0.270	0.396	0.349	0.157	0.068	−0.056
	Var $\times 10^2$	6.324	4.595	0.661	1.242	5.415	6.896	0.600	0.472	28.97
	MSE $\times 10^2$	7.555	5.916	1.050	2.927	6.330	7.124	0.786	0.574	40.76
0.1	Mean	0.561	0.202	0.141	0.278	0.348	0.345	0.160	0.068	−0.086
	Var $\times 10^2$	4.635	3.815	0.593	1.110	3.799	5.860	0.497	0.380	28.62
	MSE $\times 10^2$	5.004	4.853	0.942	2.597	4.023	6.053	0.653	0.483	38.44
0.2	Mean	0.537	0.192	0.142	0.282	0.329	0.338	0.161	0.068	−0.099
	Var $\times 10^2$	4.374	3.640	0.598	1.175	3.512	5.797	0.497	0.367	28.89
	MSE $\times 10^2$	4.504	4.487	0.930 *	2.562	3.592	5.933 *	0.649 *	0.468 *	37.90 *
0.3	Mean	0.526	0.187	0.144	0.287	0.325	0.330	0.162	0.070	−0.115
	Var $\times 10^2$	4.313	3.636	0.619	1.264	3.494	5.913	0.517	0.383	29.94
	MSE $\times 10^2$	4.377	4.383	0.932	2.529 *	3.553 *	5.998	0.660	0.472	38.02
0.5	Mean	0.516	0.177	0.149	0.300	0.329	0.312	0.166	0.075	−0.141
	Var $\times 10^2$	4.305	3.602	0.689	1.529	3.657	6.121	0.572	0.454	33.05
	MSE $\times 10^2$	4.327 *	4.188	0.950	2.532	3.739	6.128	0.689	0.514	39.75
1	Mean	0.503	0.162	0.167	0.340	0.346	0.272	0.183	0.092	−0.194
	Var $\times 10^2$	4.432	3.750	0.989	2.266	3.911	6.436	0.772	0.726	47.97
	MSE $\times 10^2$	4.428	4.130 *	1.100	2.619	4.119	6.507	0.799	0.731	52.14

Table 15. Sample mean, variance, and MSE of estimators when $\theta = (0.5, 0.1, 0.2, 0.4, 0.3, 0.3, 0.2, 0.1, -0.4)^T$, $n = 200$, and $(p, \gamma) = (0.03, 10)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	0.736	0.312	0.084	0.165	0.497	0.416	0.103	0.047	0.291
	Var $\times 10^2$	16.70	8.406	0.635	1.549	12.16	10.20	0.674	0.401	33.37
	MSE $\times 10^2$	22.23	12.91	1.971	7.078	16.04	11.54	1.610	0.680	81.11
0.1	Mean	0.600	0.264	0.092	0.181	0.381	0.368	0.110	0.047	0.032
	Var $\times 10^2$	8.133	5.973	0.445	1.314	5.185	7.196	0.442	0.233	27.95
	MSE $\times 10^2$	9.120	8.669	1.613	6.104	5.843	7.657	1.255	0.515	46.63
0.2	Mean	0.570	0.252	0.095	0.188	0.367	0.353	0.110	0.050	−0.012
	Var $\times 10^2$	7.112	5.817	0.451	1.441	4.576	6.859	0.446	0.244	28.91
	MSE $\times 10^2$	7.592	8.112	1.544	5.920	5.016	7.136	1.250	0.495 *	43.96
0.3	Mean	0.563	0.235	0.100	0.200	0.366	0.338	0.113	0.054	−0.046
	Var $\times 10^2$	6.572	5.489	0.513	1.736	4.477	6.853	0.502	0.294	29.07
	MSE $\times 10^2$	6.965	7.299	1.513	5.748	4.910	6.988	1.263	0.503	41.58
0.5	Mean	0.553	0.193	0.113	0.235	0.369	0.294	0.124	0.066	−0.101
	Var $\times 10^2$	6.273	4.985	0.736	2.594	4.586	7.030	0.689	0.441	29.86
	MSE $\times 10^2$	6.548	5.840	1.484	5.318	5.061	7.027	1.263	0.555	38.79 *
1	Mean	0.552	0.126	0.166	0.340	0.383	0.227	0.176	0.104	−0.262
	Var $\times 10^2$	4.097	3.239	1.193	3.141	3.787	6.138	1.084	0.816	46.47
	MSE $\times 10^2$	4.360 *	3.304 *	1.307 *	3.502 *	4.479 *	6.658 *	1.141 *	0.817	48.32

Table 16. Sample mean, variance, and MSE of estimators when $\theta = (0.5, 0.1, 0.2, 0.4, 0.3, 0.3, 0.2, 0.1, -0.4)^T$, $n = 200$, and $(p, \gamma) = (0.05, 10)$.

α		$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$
0(CMLE)	Mean	0.829	0.378	0.064	0.102	0.613	0.435	0.081	0.037	0.749
	Var $\times 10^2$	27.18	10.13	0.521	0.889	18.60	10.72	0.611	0.307	39.05
	MSE $\times 10^2$	37.97	17.82	2.375	9.795	28.35	12.52	2.018	0.702	171.0
0.1	Mean	0.652	0.313	0.070	0.114	0.441	0.374	0.082	0.034	0.172
	Var $\times 10^2$	12.51	7.707	0.348	0.815	7.550	8.532	0.365	0.162	29.09
	MSE $\times 10^2$	14.81	12.21	2.040	9.010	9.521	9.078	1.748	0.592	61.80
0.2	Mean	0.612	0.294	0.071	0.117	0.417	0.354	0.080	0.035	0.097
	Var $\times 10^2$	9.876	7.005	0.317	0.871	6.395	8.221	0.332	0.147	30.13
	MSE $\times 10^2$	11.13	10.74	1.979	8.885	7.751	8.510	1.768	0.571	54.75
0.3	Mean	0.604	0.283	0.073	0.121	0.414	0.343	0.081	0.037	0.063
	Var $\times 10^2$	9.469	7.048	0.348	1.031	6.125	8.167	0.358	0.167	30.74
	MSE $\times 10^2$	10.54	10.38	1.970	8.819	7.414	8.347	1.771	0.567 *	52.11
0.5	Mean	0.607	0.241	0.085	0.151	0.420	0.310	0.091	0.048	−0.006
	Var $\times 10^2$	8.559	6.604	0.565	1.957	5.915	8.036	0.536	0.337	32.13
	MSE $\times 10^2$	9.688	8.590	1.881	8.142	7.350	8.038	1.713	0.608	47.63 *
1	Mean	0.600	0.135	0.146	0.292	0.425	0.220	0.152	0.097	−0.195
	Var $\times 10^2$	5.457	4.147	1.395	4.172	4.697	6.676	1.279	0.911	46.82
	MSE $\times 10^2$	6.453 *	4.268 *	1.690 *	5.343 *	6.263 *	7.302 *	1.508 *	0.910	50.98

4.2. Illustrative Examples

First, we illustrate the proposed method by examining the monthly count series of crimes provided by the New South Wales Police Force in Australia. The data set is classified by local government area and offence type. This data set has been studied in many literatures, including Lee et al. [9], Chen and Lee [38,39], Kim and Lee [24], and Lee et al. [40]. To investigate the behavior of the MDPDE in the presence of outliers, we consider the data series of liquor offences (LO) and transport regulatory offences (TRO)

in Botany Bay from January 1995 to December 2012, which has 216 observations in each series. Figure 1 plots the monthly count series of LO and TRO and it shows the presence of some deviant observations in each series. The sample mean and variance are 1.912 and 13.14 for LO, and 2.463 and 20.41 for TRO. A large value of the variance of each series is expected to be influenced by outliers. The autocorrelation function (ACF) and partial autocorrelation function (PACF) of LO and TRO, as well as cross-correlation function (CCF) between two series, are given in Figure 2, indicating that the data are both serially and crossly correlated. The cross-correlation coefficient between two series is 0.141.
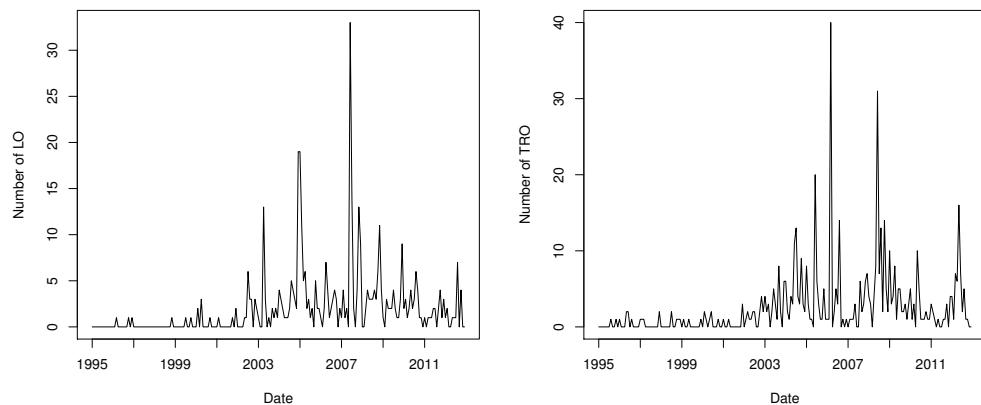
Figure 1. Monthly count series of liquor offences (LO) (**left**) and transport regulatory offences (TRO) (**right**) in Botany Bay.
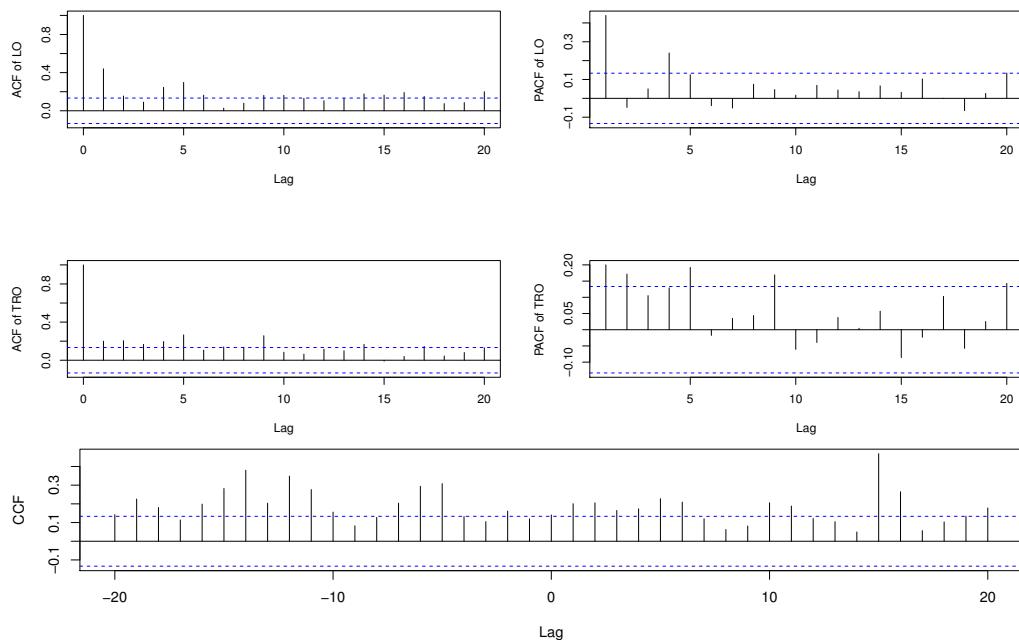
Figure 2. Autocorrelation function (ACF) and partial autocorrelation function (PACF) of LO (**top**) and TRO (**middle**), and cross-correlation function (CCF) (**bottom**) between two series.

We fit the model (2) to the data using both the CMLE and the MDPDE. $\tilde{\lambda}_1$ is set to be the sample mean of the data. Table 17 reports the estimated parameters with various α. The standard errors are given in parentheses and the symbol $\bullet$ represents the minimal $\widehat{\text{AMSE}}$ provided in Remark 2. In the table, we can observe that the MDPDE has smaller $\widehat{\text{AMSE}}$ than the CMLE and the optimal α is chosen to be 0.1. The MDPDE with optimal α is quite different from the CMLE, in particular, $\hat{\delta}$ is about half of the CMLE. This result indicates that outliers can seriously affect the parameter estimation and, thus, the robust estimation method is required when the data are contaminated by outliers.

Table 17. Parameter estimates for bivariate Poisson integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) model for crime data; the symbol $^\bullet$ represents the minimal $\widehat{\text{AMSE}}$.

α	$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$	$\widehat{\text{AMSE}}$
0(CMLE)	0.019	0.779	0.125	0.073	0.032	0.865	0.090	0.057	1.312	1.578
	(0.054)	(0.290)	(0.166)	(0.075)	(0.028)	(0.090)	(0.032)	(0.086)	(1.096)	
0.1	0.034	0.609	0.172	0.094	0.097	0.654	0.095	0.156	0.685	0.699 $^\bullet$
	(0.034)	(0.149)	(0.104)	(0.026)	(0.047)	(0.091)	(0.043)	(0.069)	(0.678)	
0.2	0.026	0.643	0.134	0.087	0.117	0.575	0.124	0.159	0.509	0.858
	(0.032)	(0.163)	(0.109)	(0.026)	(0.060)	(0.121)	(0.052)	(0.069)	(0.692)	
0.3	0.021	0.666	0.113	0.085	0.129	0.523	0.154	0.155	0.401	0.991
	(0.029)	(0.149)	(0.096)	(0.027)	(0.067)	(0.133)	(0.053)	(0.068)	(0.710)	
0.4	0.019	0.673	0.107	0.085	0.130	0.508	0.176	0.145	0.356	1.081
	(0.029)	(0.143)	(0.093)	(0.029)	(0.067)	(0.135)	(0.055)	(0.069)	(0.736)	
0.5	0.018	0.675	0.105	0.086	0.125	0.514	0.196	0.131	0.365	1.108
	(0.029)	(0.138)	(0.093)	(0.032)	(0.065)	(0.136)	(0.059)	(0.071)	(0.768)	
0.6	0.017	0.676	0.104	0.088	0.119	0.527	0.216	0.115	0.418	1.094
	(0.029)	(0.133)	(0.091)	(0.036)	(0.062)	(0.135)	(0.065)	(0.073)	(0.807)	
0.7	0.017	0.675	0.104	0.089	0.114	0.540	0.238	0.100	0.509	1.073
	(0.029)	(0.130)	(0.092)	(0.041)	(0.059)	(0.133)	(0.075)	(0.075)	(0.859)	
0.8	0.018	0.674	0.104	0.090	0.111	0.551	0.261	0.087	0.638	1.079
	(0.031)	(0.130)	(0.094)	(0.045)	(0.057)	(0.133)	(0.089)	(0.076)	(0.929)	
0.9	0.018	0.672	0.104	0.091	0.109	0.560	0.285	0.076	0.808	1.158
	(0.033)	(0.133)	(0.098)	(0.050)	(0.056)	(0.134)	(0.105)	(0.077)	(1.021)	
1	0.019	0.668	0.104	0.092	0.108	0.568	0.312	0.066	1.025	1.383
	(0.035)	(0.138)	(0.103)	(0.054)	(0.057)	(0.136)	(0.122)	(0.079)	(1.143)	

We clean the data by using the approach that was introduced by Fokianos and Fried [41] and apply the CMLE and the MDPDE to this data in order to illustrate the behavior of the estimators in the absence of outliers. Table 18 reports the results. The standard errors and $\widehat{\text{AMSE}}$ tend to decrease compared to Table 17. The CMLE has minimal $\widehat{\text{AMSE}}$ and the MDPDE with small α appears to be similar to the CMLE.

Table 18. Parameter estimates for bivariate Poisson INGARCH model for cleaned data; the symbol $^\bullet$ represents the minimal $\widehat{\text{AMSE}}$.

α	$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$	$\widehat{\text{AMSE}}$
0(CMLE)	0.0018	0.943	0.025	0.021	0.092	0.682	0.067	0.184	0.118	0.430 $^\bullet$
	(0.007)	(0.118)	(0.084)	(0.017)	(0.050)	(0.115)	(0.069)	(0.069)	(0.609)	
0.1	0.0002	0.942	0.026	0.022	0.074	0.680	0.066	0.183	0.159	0.445
	(0.006)	(0.097)	(0.076)	(0.013)	(0.035)	(0.084)	(0.063)	(0.053)	(0.626)	
0.2	0.0001	0.940	0.025	0.023	0.066	0.679	0.066	0.182	0.199	0.497
	(0.009)	(0.100)	(0.079)	(0.011)	(0.032)	(0.075)	(0.063)	(0.049)	(0.657)	
0.3	0.0001	0.939	0.024	0.023	0.060	0.678	0.066	0.182	0.220	0.549
	(0.010)	(0.102)	(0.082)	(0.011)	(0.032)	(0.073)	(0.064)	(0.049)	(0.688)	
0.4	0.0001	0.939	0.024	0.023	0.057	0.678	0.066	0.182	0.228	0.591
	(0.010)	(0.104)	(0.086)	(0.011)	(0.032)	(0.074)	(0.066)	(0.049)	(0.715)	
0.5	0.0001	0.938	0.025	0.023	0.056	0.676	0.066	0.182	0.293	0.665
	(0.010)	(0.104)	(0.087)	(0.011)	(0.034)	(0.076)	(0.068)	(0.051)	(0.742)	
0.6	0.0001	0.938	0.024	0.023	0.054	0.677	0.066	0.182	0.263	0.688
	(0.009)	(0.102)	(0.088)	(0.011)	(0.035)	(0.077)	(0.071)	(0.053)	(0.769)	
0.7	0.0002	0.939	0.024	0.023	0.051	0.678	0.066	0.182	0.237	0.719
	(0.008)	(0.106)	(0.093)	(0.012)	(0.035)	(0.078)	(0.074)	(0.055)	(0.795)	
0.8	0.0002	0.940	0.015	0.027	0.053	0.678	0.067	0.179	0.100	0.812
	(0.011)	(0.126)	(0.093)	(0.013)	(0.038)	(0.081)	(0.076)	(0.057)	(0.873)	
0.9	0.0001	0.944	0.011	0.028	0.050	0.679	0.068	0.176	-0.029	0.924
	(0.012)	(0.150)	(0.108)	(0.015)	(0.039)	(0.083)	(0.080)	(0.058)	(0.933)	
1	0.0002	0.944	0.010	0.028	0.054	0.677	0.070	0.173	0.010	1.079
	(0.012)	(0.151)	(0.107)	(0.015)	(0.044)	(0.087)	(0.084)	(0.061)	(1.012)	

Now, we consider an artificial example that has negative cross-correlation coefficient. Following Cui and Zhu [20], we generate 1000 samples from univariate Poisson INGARCH model, i.e.,

$$X_t | \mathcal{F}_{t-1} \sim P(\lambda_t), \ \ \lambda_t = 1 + 0.35\lambda_{t-1} + 0.45X_{t-1},$$

where $P(\lambda_t)$ denotes the Poisson distribution with mean λ_t. Further, we observe the contaminated data $X_{c,t}$ as follows

$$X_{c,t} = X_t + P_t X_{o,t},$$

where P_t are i.i.d. Bernoulli random variables with a success probability of 0.03 and $X_{o,t}$ are i.i.d. Poisson random variables with mean 5. Let $\boldsymbol{Y}_t = (Y_{t,1}, Y_{t,2})^T$, where $Y_{t,1} = X_{c,t}$ and $Y_{t,2} = X_{c,t+500}$ for $t = 1, \ldots, 500$. The sample mean and variance are 5.196 and 7.380 for $Y_{t,1}$, and 4.538 and 8.129 for $Y_{t,2}$. The cross-correlation coefficient between $Y_{t,1}$ and $Y_{t,2}$ is -0.161. We fit the model (2) to $\boldsymbol{Y}_t$ and the results are presented in Table 19. Similar to Table 17, the MDPDE has smaller $\widehat{\text{AMSE}}$ than the CMLE. The optimal α is chosen to be 0.3 and the corresponding $\hat{\delta}$ is -0.329, whereas the CMLE is 0.772.

Table 19. Parameter estimates for bivariate Poisson INGARCH model for artificial data; the symbol • represents the minimal $\widehat{\text{AMSE}}$.

α	$\hat{\omega}_1$	$\hat{a}_1$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{\omega}_2$	$\hat{a}_2$	$\hat{b}_{21}$	$\hat{b}_{22}$	$\hat{\delta}$	$\widehat{\text{AMSE}}$
0(CMLE)	1.507	0.274	0.438	0.000	0.976	0.410	0.000	0.375	0.772	9.468
	(0.442)	(0.102)	(0.053)	(0.031)	(0.241)	(0.069)	(0.029)	(0.048)	(2.939)	
0.1	1.442	0.274	0.449	0.000	0.952	0.412	0.000	0.372	0.308	7.647
	(0.432)	(0.100)	(0.053)	(0.031)	(0.236)	(0.066)	(0.030)	(0.048)	(2.688)	
0.2	1.402	0.273	0.457	0.000	0.918	0.417	0.000	0.371	-0.064	6.611
	(0.443)	(0.102)	(0.054)	(0.033)	(0.237)	(0.065)	(0.031)	(0.049)	(2.487)	
0.3	1.373	0.271	0.464	0.000	0.883	0.422	0.000	0.372	-0.329	6.216 •
	(0.465)	(0.105)	(0.056)	(0.034)	(0.242)	(0.065)	(0.033)	(0.050)	(2.367)	
0.4	1.349	0.269	0.471	0.000	0.849	0.425	0.000	0.375	-0.485	6.276
	(0.494)	(0.111)	(0.058)	(0.036)	(0.250)	(0.064)	(0.034)	(0.052)	(2.339)	
0.5	1.326	0.268	0.476	0.000	0.817	0.427	0.000	0.380	-0.540	6.607
	(0.528)	(0.118)	(0.060)	(0.039)	(0.259)	(0.064)	(0.037)	(0.054)	(2.388)	
0.6	1.302	0.267	0.482	0.000	0.786	0.428	0.000	0.386	-0.509	7.031
	(0.567)	(0.126)	(0.063)	(0.041)	(0.271)	(0.064)	(0.039)	(0.056)	(2.476)	
0.7	1.277	0.267	0.487	0.000	0.758	0.428	0.000	0.394	-0.407	7.412
	(0.610)	(0.135)	(0.066)	(0.044)	(0.285)	(0.064)	(0.042)	(0.058)	(2.566)	
0.8	1.250	0.267	0.491	0.000	0.732	0.427	0.000	0.401	-0.250	7.698
	(0.657)	(0.145)	(0.069)	(0.047)	(0.299)	(0.064)	(0.045)	(0.060)	(2.639)	
0.9	1.223	0.267	0.496	0.000	0.708	0.425	0.000	0.410	-0.055	7.916
	(0.707)	(0.156)	(0.072)	(0.050)	(0.314)	(0.065)	(0.048)	(0.062)	(2.688)	
1	1.196	0.268	0.500	0.000	0.686	0.423	0.000	0.418	0.165	8.131
	(0.761)	(0.168)	(0.076)	(0.053)	(0.330)	(0.065)	(0.051)	(0.064)	(2.719)	

5. Concluding Remarks

In this study, we developed the robust estimator for bivariate Poisson INGARCH models based on the MDPDE. In practice, this is important, because outliers can strongly affect the CMLE, which is commonly employed for parameter estimation in INGARCH models. We proved that the MDPDE is consistent and asymptotically normal under regularity conditions. Our simulation study compared the performances of the MDPDE and the CMLE, and confirmed the superiority of the proposed estimator in the presence of outliers. The real data analysis also confirmed the validity of our method as a robust estimator in practice. Although we focused on Cui and Zhu's [20] bivariate Poisson INGARCH models here, the MDPDE method can be extended to other bivariate INGARCH models. For example, one can consider the copula-based bivariate INGARCH models that were studied by Heinen and Rengifo [42], Andreassen [18], and Fokianos et al. [43]. We leave this issue of extension as our future research.

Author Contributions: Conceptualization, B.K. and S.L.; methodology, B.K., S.L. and D.K.; software, B.K. and D.K.; formal analysis, B.K., S.L. and D.K.; data curation, B.K.; writing—original draft preparation, B.K. and S.L.; writing—review and editing, B.K. and S.L.; funding acquisition, B.K. and S.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: data.gov.au (accessed on 19 March 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this Appendix, we provide the proofs for Theorems 1 and 2 in Section 3 when $\alpha > 0$. We refer to Cui and Zhu [20] for the case of $\alpha = 0$. In what follows, we denote

V and $\rho \in (0, 1)$ as a generic positive integrable random variable and a generic constant, respectively, and $H_{\alpha,n}(\theta) = n^{-1} \sum_{t=1}^{n} h_{\alpha,t}(\theta)$. Furthermore, we employ the notation $\lambda_t = \lambda_t(\theta)$, $\tilde{\lambda}_t = \tilde{\lambda}_t(\theta)$, $\lambda_t^0 = \lambda_t(\theta^0)$, $\lambda_{t,i} = \lambda_{t,i}(\theta_i)$, $\tilde{\lambda}_{t,i} = \tilde{\lambda}_{t,i}(\theta_i)$, $\lambda_{t,i}^0 = \lambda_{t,i}(\theta_i^0)$ for $i = 1, 2$, and $u(y, \lambda) = e^{-y} - e^{-c\lambda}$ for brevity.

Lemma A1. *Under (A1)–(A3), we have for $i = 1, 2$,*

(i) $E(\sup_{\theta_i \in \Theta_i} \lambda_{t,i}) < \infty$.

(ii) $\lambda_{t,i} = \lambda_{t,i}^0$ *a.s. implies* $\theta_i = \theta_i^0$.

(iii) $\lambda_{t,i}$ *is twice continuously differentiable with respect to θ_i and satisfies*

$$E\left(\sup_{\theta_i \in \Theta_i} \left\| \frac{\partial \lambda_{t,i}}{\partial \theta_i} \right\|_1 \right)^4 < \infty \text{ and } E\left(\sup_{\theta_i \in \Theta_i} \left\| \frac{\partial^2 \lambda_{t,i}}{\partial \theta_i \partial \theta_i^T} \right\|_1 \right)^2 < \infty.$$

(iv) $\sup_{\theta_i \in \Theta_i} \left\| \frac{\partial \lambda_{t,i}}{\partial \theta_i} - \frac{\partial \tilde{\lambda}_{t,i}}{\partial \theta_i} \right\|_1 \leq V\rho^t$ *a.s.*

(v) $\nu^T \frac{\partial \lambda_{t,i}^0}{\partial \theta_i} = 0$ *a.s. implies $\nu = 0$.*

(vi) $\sup_{\theta_i \in \Theta_i} |\lambda_{t,i} - \tilde{\lambda}_{t,i}| \leq V\rho^t$ *a.s.*

Proof. By recursion of (2), we have

$$\lambda_t = (I_2 - A)^{-1}\omega + \sum_{k=0}^{\infty} A^k B Y_{t-k-1},$$

$$\tilde{\lambda}_t = (I_2 + A + \cdots + A^{t-2})\omega + A^{t-1}\tilde{\lambda}_1 + \sum_{k=0}^{t-2} A^k B Y_{t-k-1}$$

and thus, for $i = 1, 2$,

$$\lambda_{t,i} = \frac{\omega_i}{1 - a_i} + \sum_{k=0}^{\infty} a_i^k (b_{i1} Y_{t-k-1,1} + b_{i2} Y_{t-k-1,2}),$$

$$\tilde{\lambda}_{t,i} = \frac{\omega_i}{1 - a_i} + \sum_{k=0}^{t-2} a_i^k (b_{i1} Y_{t-k-1,1} + b_{i2} Y_{t-k-1,2}),$$

where I_2 denotes 2×2 identity matrix and the initial value $\tilde{\lambda}_{1,i}$ is taken as $\omega_i/(1 - a_i)$ for simplicity. Subsequently, $(i) - (v)$ can be shown following the arguments in the proof of Theorem 3 in Kang and Lee [44]. For (vi), let $\rho = \sup_{\theta_i \in \Theta_i} a_i < 1$. Afterwards, from (2), it holds that

$$\sup_{\theta_i \in \Theta_i} |\lambda_{t,i} - \tilde{\lambda}_{t,i}| = \sup_{\theta_i \in \Theta_i} |a_i(\lambda_{t-1,i} - \tilde{\lambda}_{t-1,i})| = \cdots = \sup_{\theta_i \in \Theta_i} |a_i^{t-1}(\lambda_{1,i} - \tilde{\lambda}_{1,i})| \leq V\rho^t$$

with $V = \sup_{\theta_i \in \Theta_i} |\lambda_{1,i} - \tilde{\lambda}_{1,i}|/\rho$. Therefore, the lemma is established. □

Lemma A2. *Under (A1)–(A3), we have*

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

Proof. It is sufficient to show that

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

We write

$$|h_{\alpha,t}(\theta) - \tilde{h}_{\alpha,t}(\theta)|$$

$$\leq \left| \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} \left\{ f_{\theta}^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t) - f_{\theta}^{1+\alpha}(\boldsymbol{y}|\tilde{\boldsymbol{\lambda}}_t) \right\} \right| + \left(1 + \frac{1}{\alpha}\right) |f_{\theta}^{\alpha}(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t) - f_{\theta}^{\alpha}(\boldsymbol{Y}_t|\tilde{\boldsymbol{\lambda}}_t)|$$

$$:= I_t(\theta) + II_t(\theta).$$

From **(A1)**, **(A2)**, and the mean value theorem (MVT), we have

$$
\begin{aligned}
I_t(\theta) &= (1+\alpha) \left| \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_{\theta}^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t^*) \left\{ \frac{y_1}{\lambda_{t,1}^*} - 1 + \frac{c\delta e^{-c\lambda_{t,1}^*} u(y_2, \lambda_{t,2}^*)}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t^*, \delta)} \right\} (\lambda_{t,1} - \tilde{\lambda}_{t,1}) \right. \\
&\quad \left. + \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_{\theta}^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t^*) \left\{ \frac{y_2}{\lambda_{t,2}^*} - 1 + \frac{c\delta e^{-c\lambda_{t,2}^*} u(y_1, \lambda_{t,1}^*)}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t^*, \delta)} \right\} (\lambda_{t,2} - \tilde{\lambda}_{t,2}) \right| \\
&\leq (1+\alpha) \left[\sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_{\theta}(\boldsymbol{y}|\boldsymbol{\lambda}_t^*) \left\{ \frac{y_1}{\lambda_{t,1}^*} + 1 + \frac{c|\delta|e^{-c\lambda_{t,1}^*}|u(y_2, \lambda_{t,2}^*)|}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t^*, \delta)} \right\} |\lambda_{t,1} - \tilde{\lambda}_{t,1}| \right. \\
&\quad \left. + \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_{\theta}(\boldsymbol{y}|\boldsymbol{\lambda}_t^*) \left\{ \frac{y_2}{\lambda_{t,2}^*} + 1 + \frac{c|\delta|e^{-c\lambda_{t,2}^*}|u(y_1, \lambda_{t,1}^*)|}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t^*, \delta)} \right\} |\lambda_{t,2} - \tilde{\lambda}_{t,2}| \right] \\
&\leq (1+\alpha) \left\{ \left(1 + 1 + \frac{2c\delta_U}{\varphi_L}\right) |\lambda_{t,1} - \tilde{\lambda}_{t,1}| + \left(1 + 1 + \frac{2c\delta_U}{\varphi_L}\right) |\lambda_{t,2} - \tilde{\lambda}_{t,2}| \right\} \\
&= 2(1+\alpha) \left(1 + \frac{c\delta_U}{\varphi_L}\right) (|\lambda_{t,1} - \tilde{\lambda}_{t,1}| + |\lambda_{t,2} - \tilde{\lambda}_{t,2}|),
\end{aligned}
$$

where $\boldsymbol{\lambda}_t^* = (\lambda_{t,1}^*, \lambda_{t,2}^*)^T$ and $\lambda_{t,i}^*$ is an intermediate point between $\lambda_{t,i}$ and $\tilde{\lambda}_{t,i}$ for $i = 1, 2$. Hence, $\sup_{\theta \in \Theta} I_t(\theta)$ converges to 0 a.s. as $t \to \infty$ by (vi) of Lemma A1.

Because $\lambda_{t,i}^* = m\lambda_{t,i} + (1-m)\tilde{\lambda}_{t,i}$ for some $m \in (0,1)$, it holds that $(\lambda_{t,i}^*)^{-1} \leq (m\lambda_{t,i})^{-1} \leq (m\omega_L)^{-1}$. Thus, we obtain

$$
\begin{aligned}
II_t(\theta) &= (1+\alpha) \left| f_{\theta}^{\alpha}(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t^*) \left\{ \frac{Y_{t,1}}{\lambda_{t,1}^*} - 1 + \frac{c\delta e^{-c\lambda_{t,1}^*} u(Y_{t,2}, \lambda_{t,2}^*)}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t^*, \delta)} \right\} (\lambda_{t,1} - \tilde{\lambda}_{t,1}) \right. \\
&\quad \left. + f_{\theta}^{\alpha}(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t^*) \left\{ \frac{Y_{t,2}}{\lambda_{t,2}^*} - 1 + \frac{c\delta e^{-c\lambda_{t,2}^*} u(Y_{t,1}, \lambda_{t,1}^*)}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t^*, \delta)} \right\} (\lambda_{t,2} - \tilde{\lambda}_{t,2}) \right| \\
&\leq (1+\alpha) \left\{ \left(\frac{Y_{t,1}}{m\omega_L} + 1 + \frac{2c\delta_U}{\varphi_L} \right) |\lambda_{t,1} - \tilde{\lambda}_{t,1}| + \left(\frac{Y_{t,2}}{m\omega_L} + 1 + \frac{2c\delta_U}{\varphi_L} \right) |\lambda_{t,2} - \tilde{\lambda}_{t,2}| \right\}.
\end{aligned}
$$

According to Lemma 2.1 in Straumann and Mikosch [45], together with (vi) of Lemma A1, $\sup_{\theta \in \Theta} II_t(\theta)$ converges to 0 a.s. as $t \to \infty$. Therefore, the lemma is verified. □

Lemma A3. *Under (A1)–(A3), we have*

$$E\left(\sup_{\theta \in \Theta} |h_{\alpha,t}(\theta)| \right) < \infty \text{ and if } \theta \neq \theta^0, \text{ then } E(h_{\alpha,t}(\theta)) > E(h_{\alpha,t}(\theta^0)).$$

Proof. Because

$$|h_{\alpha,t}(\theta)| \leq \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_{\theta}^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t) + \left(1 + \frac{1}{\alpha}\right) f_{\theta}^{\alpha}(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t) \leq 2 + \frac{1}{\alpha},$$

the first part of the lemma is validated. Note that

$$
\begin{aligned}
& E(h_{\alpha,t}(\theta)) - E(h_{\alpha,t}(\theta^0)) \\
= \; & E\Big[E\big\{h_{\alpha,t}(\theta) - h_{\alpha,t}(\theta^0)|\mathcal{F}_{t-1}\big\}\Big] \\
= \; & E\left[\sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\Big\{f_\theta^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t) - \Big(1+\frac{1}{\alpha}\Big)f_\theta^\alpha(\boldsymbol{y}|\boldsymbol{\lambda}_t)f_{\theta^0}(\boldsymbol{y}|\boldsymbol{\lambda}_t) + \frac{1}{\alpha}f_{\theta^0}^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t)\Big\}\right] \\
\geq \; & 0,
\end{aligned}
$$

where the equality holds if and only if $\delta = \delta^0$ and $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_t^0$ a.s. Therefore, the second part of the lemma is established by (ii) of Lemma A1. $\square$

Proof of Theorem 1. We can write

$$
\begin{aligned}
& \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\tilde{h}_{\alpha,t}(\theta) - E(h_{\alpha,t}(\theta))\right| \\
\leq \; & \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\tilde{h}_{\alpha,t}(\theta) - \frac{1}{n}\sum_{t=1}^{n}h_{\alpha,t}(\theta)\right| + \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}h_{\alpha,t}(\theta) - E(h_{\alpha,t}(\theta))\right|.
\end{aligned}
$$

The first term on the RHS of the inequality converges to 0 a.s. from Lemma A2. Moreover, because $h_{\alpha,t}(\theta)$ is stationary and ergodic with $E(\sup_{\theta\in\Theta}|h_{\alpha,t}(\theta)|) < \infty$ by Lemma A3, the second term also converges to 0 a.s. Finally, as $E(h_{\alpha,t}(\theta))$ has a unique minimum at θ^0 from Lemma A3, the theorem is established. $\square$

Now, we derive the first and second derivatives of $h_{\alpha,t}(\theta)$. The first derivatives are obtained as

$$
\begin{aligned}
\frac{\partial h_{\alpha,t}(\theta)}{\partial\theta} &= (1+\alpha)\Big(D_{t,1}(\theta)s_{t,1}(\theta_1)^T, D_{t,2}(\theta)s_{t,2}(\theta_2)^T, D_{t,3}(\theta)\Big)^T \\
&= (1+\alpha)\begin{pmatrix} D_{t,1}(\theta)\boldsymbol{I}_4 & \boldsymbol{0}_{4\times 4} & \boldsymbol{0}_{4\times 1} \\ \boldsymbol{0}_{4\times 4} & D_{t,2}(\theta)\boldsymbol{I}_4 & \boldsymbol{0}_{4\times 1} \\ \boldsymbol{0}_{1\times 4} & \boldsymbol{0}_{1\times 4} & D_{t,3}(\theta) \end{pmatrix}\begin{pmatrix} s_{t,1}(\theta_1) \\ s_{t,2}(\theta_2) \\ 1 \end{pmatrix} \\
&:= (1+\alpha)\boldsymbol{D}_t(\theta)\Lambda_t(\theta),
\end{aligned}
$$

where $\boldsymbol{I}_4$ denotes the 4×4 identity matrix, $\boldsymbol{0}_{m\times n}$ means the $m\times n$ matrix with zero elements, and

$$
\begin{aligned}
s_{t,i}(\theta_i) &= \frac{\partial\lambda_{t,i}}{\partial\theta_i}\ \text{ for }\ i=1,2, \\
D_{t,i}(\theta) &= \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}f_\theta^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t)\left\{\frac{y_i}{\lambda_{t,i}}-1+\frac{c\delta e^{-c\lambda_{t,i}}u(y_j,\lambda_{t,j})}{\varphi(\boldsymbol{y},\boldsymbol{\lambda}_t,\delta)}\right\} \\
& \quad -f_\theta^\alpha(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t)\left\{\frac{Y_{t,i}}{\lambda_{t,i}}-1+\frac{c\delta e^{-c\lambda_{t,i}}u(Y_{t,j},\lambda_{t,j})}{\varphi(\boldsymbol{Y}_t,\boldsymbol{\lambda}_t,\delta)}\right\}\ \text{ for }\ (i,j)=(1,2),(2,1), \\
D_{t,3}(\theta) &= \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}f_\theta^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t)\frac{u(y_1,\lambda_{t,1})u(y_2,\lambda_{t,2})}{\varphi(\boldsymbol{y},\boldsymbol{\lambda}_t,\delta)}-f_\theta^\alpha(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t)\frac{u(Y_{t,1},\lambda_{t,1})u(Y_{t,2},\lambda_{t,2})}{\varphi(\boldsymbol{Y}_t,\boldsymbol{\lambda}_t,\delta)}.
\end{aligned}
$$

The second derivatives are expressed as

$$
\begin{aligned}
\frac{\partial^2 h_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} &= (1+\alpha)\begin{pmatrix} F_{t,11}(\theta)s_{t,1}(\theta_1)s_{t,1}(\theta_1)^T & F_{t,12}(\theta)s_{t,1}(\theta_1)s_{t,2}(\theta_2)^T & F_{t,13}(\theta)s_{t,1}(\theta_1) \\ F_{t,21}(\theta)s_{t,2}(\theta_2)s_{t,1}(\theta_1)^T & F_{t,22}(\theta)s_{t,2}(\theta_2)s_{t,2}(\theta_2)^T & F_{t,23}(\theta)s_{t,2}(\theta_2) \\ F_{t,31}(\theta)s_{t,1}(\theta_1)^T & F_{t,32}(\theta)s_{t,2}(\theta_2)^T & F_{t,33}(\theta) \end{pmatrix} \\
& \quad +(1+\alpha)\begin{pmatrix} D_{t,1}(\theta)s_{t,11}(\theta_1) & \boldsymbol{0}_{4\times 4} & \boldsymbol{0}_{4\times 1} \\ \boldsymbol{0}_{4\times 4} & D_{t,2}(\theta)s_{t,22}(\theta_2) & \boldsymbol{0}_{4\times 1} \\ \boldsymbol{0}_{1\times 4} & \boldsymbol{0}_{1\times 4} & 0 \end{pmatrix} \\
& := (1+\alpha)\left\{\boldsymbol{F}_t(\theta) + \boldsymbol{D}_t(\theta)\frac{\partial\Lambda_t(\theta)}{\partial\theta^T}\right\},
\end{aligned}
$$

where

$$s_{t,ii}(\theta_i) = \frac{\partial^2 \lambda_{t,i}}{\partial \theta_i \partial \theta_i^T} \text{ for } i = 1, 2,$$

$$
\begin{aligned}
F_{t,ii}(\theta) &= \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t) \Bigg[(1+\alpha) \left\{ \frac{y_i}{\lambda_{t,i}} - 1 + \frac{c\delta e^{-c\lambda_{t,i}} u(y_j, \lambda_{t,j})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\}^2 \\
&\quad - \frac{y_i}{\lambda_{t,i}^2} - \frac{c^2 \delta e^{-c\lambda_{t,i}} u(y_j, \lambda_{t,j}) \left\{ 1 + \delta e^{-y_i} u(y_j, \lambda_{t,j}) \right\}}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)^2} \Bigg] \\
&\quad - f_\theta^\alpha(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t) \Bigg[\alpha \left\{ \frac{Y_{t,i}}{\lambda_{t,i}} - 1 + \frac{c\delta e^{-c\lambda_{t,i}} u(Y_{t,j}, \lambda_{t,j})}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\}^2 \\
&\quad - \frac{Y_{t,i}}{\lambda_{t,i}^2} - \frac{c^2 \delta e^{-c\lambda_{t,i}} u(Y_{t,j}, \lambda_{t,j}) \left\{ 1 + \delta e^{-Y_{t,i}} u(Y_{t,j}, \lambda_{t,j}) \right\}}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)^2} \Bigg] \\
&\quad \text{for } (i, j) = (1, 2), (2, 1),
\end{aligned}
$$

$$
\begin{aligned}
F_{t,33}(\theta) &= \alpha \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t) \left\{ \frac{u(y_1, \lambda_{t,1}) u(y_2, \lambda_{t,2})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\}^2 \\
&\quad - (\alpha - 1) f_\theta^\alpha(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t) \left\{ \frac{u(Y_{t,1}, \lambda_{t,1}) u(Y_{t,2}, \lambda_{t,2})}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\}^2,
\end{aligned}
$$

$$
\begin{aligned}
F_{t,12}(\theta) &= \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t) \Bigg[(1+\alpha) \left\{ \frac{y_1}{\lambda_{t,1}} - 1 + \frac{c\delta e^{-c\lambda_{t,1}} u(y_2, \lambda_{t,2})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\} \\
&\quad \times \left\{ \frac{y_2}{\lambda_{t,2}} - 1 + \frac{c\delta e^{-c\lambda_{t,2}} u(y_1, \lambda_{t,1})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\} + \frac{c^2 \delta e^{-c(\lambda_{t,1}+\lambda_{t,2})}}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)^2} \Bigg] \\
&\quad - f_\theta^\alpha(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t) \Bigg[\alpha \left\{ \frac{Y_{t,1}}{\lambda_{t,1}} - 1 + \frac{c\delta e^{-c\lambda_{t,1}} u(Y_{t,2}, \lambda_{t,2})}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\} \\
&\quad \times \left\{ \frac{Y_{t,2}}{\lambda_{t,2}} - 1 + \frac{c\delta e^{-c\lambda_{t,2}} u(Y_{t,1}, \lambda_{t,1})}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\} + \frac{c^2 \delta e^{-c(\lambda_{t,1}+\lambda_{t,2})}}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)^2} \Bigg],
\end{aligned}
$$

$$
\begin{aligned}
F_{t,i3}(\theta) &= \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta^{1+\alpha}(\boldsymbol{y}|\boldsymbol{\lambda}_t) \Bigg[(1+\alpha) \left\{ \frac{y_i}{\lambda_{t,i}} - 1 + \frac{c\delta e^{-c\lambda_{t,i}} u(y_j, \lambda_{t,j})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\} \\
&\quad \times \frac{u(y_1, \lambda_{t,1}) u(y_2, \lambda_{t,2})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} + \frac{ce^{-c\lambda_{t,i}} u(y_j, \lambda_{t,j})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)^2} \Bigg] \\
&\quad - f_\theta^\alpha(\boldsymbol{Y}_t|\boldsymbol{\lambda}_t) \Bigg[\alpha \left\{ \frac{Y_{t,i}}{\lambda_{t,i}} - 1 + \frac{c\delta e^{-c\lambda_{t,i}} u(Y_{t,j}, \lambda_{t,j})}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\} \frac{u(Y_{t,1}, \lambda_{t,1}) u(Y_{t,2}, \lambda_{t,2})}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \\
&\quad + \frac{ce^{-c\lambda_{t,i}} u(Y_{t,j}, \lambda_{t,j})}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)^2} \Bigg] \text{ for } (i, j) = (1, 2), (2, 1).
\end{aligned}
$$

The following four lemmas are helpful for proving Theorem 2.

Lemma A4. *Let $\widetilde{D}_{t,i}(\theta)$ denote the counterpart of $D_{t,i}(\theta)$ by substituting λ_t with $\tilde{\lambda}_t$ for $i = 1, 2, 3$. Subsequently, under **(A1)–(A3)**, we have that for $i = 1, 2$,*

$$
\begin{aligned}
&|D_{t,i}(\theta)| \le C(Y_{t,i} + 1), \; |\widetilde{D}_{t,i}(\theta)| \le C(Y_{t,i} + 1), \; |D_{t,3}(\theta)| \le C, \; |\widetilde{D}_{t,3}(\theta)| \le C, \\
&|F_{t,ii}(\theta)| \le C(Y_{t,i}^2 + Y_{t,i} + 1), \; |F_{t,33}(\theta)| \le C, \; |F_{t,12}(\theta)| \le C(Y_{t,1} Y_{t,2} + Y_{t,1} + Y_{t,2} + 1), \\
&|F_{t,i3}(\theta)| \le C(Y_{t,i} + 1),
\end{aligned}
$$

and for $(i, j) = (1, 2), (2, 1)$,

$$|D_{t,i}(\theta) - \widetilde{D}_{t,i}(\theta)| \leq C(Y_{t,i}^2 + Y_{t,i} + 1)|\lambda_{t,i} - \tilde{\lambda}_{t,i}|$$
$$+ C(Y_{t,1}Y_{t,2} + Y_{t,1} + Y_{t,2} + 1)|\lambda_{t,j} - \tilde{\lambda}_{t,j}|,$$
$$|D_{t,3}(\theta) - \widetilde{D}_{t,3}(\theta)| \leq C(Y_{t,1} + 1)|\lambda_{t,1} - \tilde{\lambda}_{t,1}| + C(Y_{t,2} + 1)|\lambda_{t,2} - \tilde{\lambda}_{t,2}|,$$

where C is some positive constant.

Proof. From **(A1)–(A3)** and the fact that $\lambda_{t,i}^{-1} \leq \omega_L^{-1}$, we obtain

$$
\begin{aligned}
|D_{t,i}(\theta)| &\leq \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta(\boldsymbol{y}|\boldsymbol{\lambda}_t) \left\{ \frac{y_i}{\lambda_{t,i}} + 1 + \frac{c|\delta|e^{-c\lambda_{t,i}}|u(y_j, \lambda_{t,j})|}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\} \\
&\quad + \left\{ \frac{Y_{t,i}}{\lambda_{t,i}} + 1 + \frac{c|\delta|e^{-c\lambda_{t,i}}|u(Y_{t,j}, \lambda_{t,j})|}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\} \\
&\leq 1 + 1 + \frac{2c\delta_U}{\varphi_L} + \frac{Y_{t,i}}{\omega_L} + 1 + \frac{2c\delta_U}{\varphi_L} \\
&= \frac{Y_{t,i}}{\omega_L} + 3 + \frac{4c\delta_U}{\varphi_L}
\end{aligned}
$$

for $(i, j) = (1, 2), (2, 1)$ and

$$
\begin{aligned}
|D_{t,3}(\theta)| &\leq \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta(\boldsymbol{y}|\boldsymbol{\lambda}_t) \frac{|u(y_1, \lambda_{t,1})||u(y_2, \lambda_{t,2})|}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} + \frac{|u(Y_{t,1}, \lambda_{t,1})||u(Y_{t,2}, \lambda_{t,2})|}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \\
&= \frac{8}{\varphi_L}.
\end{aligned}
$$

The second and fourth parts of the lemma also hold in the same manner. Furthermore, we can show that

$$
\begin{aligned}
|F_{t,ii}(\theta)| &\leq \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta(\boldsymbol{y}|\boldsymbol{\lambda}_t) \Bigg[2(1+\alpha) \left\{ \left(\frac{y_i - \lambda_{t,i}}{\lambda_{t,i}} \right)^2 + \left(\frac{c\delta e^{-c\lambda_{t,i}} u(y_j, \lambda_{t,j})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right)^2 \right\} \\
&\quad + \frac{y_i}{\lambda_{t,i}^2} + \frac{c^2|\delta|e^{-c\lambda_{t,i}}|u(y_j, \lambda_{t,j})| \left\{ 1 + |\delta|e^{-y_i}|u(y_j, \lambda_{t,j})| \right\}}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)^2} \Bigg] \\
&\quad + 2\alpha \left\{ \left(\frac{Y_{t,i}}{\lambda_{t,i}} - 1 \right)^2 + \left(\frac{c\delta e^{-c\lambda_{t,i}} u(Y_{t,j}, \lambda_{t,j})}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right)^2 \right\} \\
&\quad - \frac{Y_{t,i}}{\lambda_{t,i}^2} - \frac{c^2|\delta|e^{-c\lambda_{t,i}}|u(Y_{t,j}, \lambda_{t,j})| \left\{ 1 + |\delta|e^{-Y_{t,i}}|u(Y_{t,j}, \lambda_{t,j})| \right\}}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)^2} \\
&\leq 2(1+\alpha) \frac{1}{\omega_L} + 2(1+\alpha) \frac{4c^2\delta_U^2}{\varphi_L^2} + \frac{1}{\omega_L} + \frac{2c^2\delta_U(1 + 2\delta_U)}{\varphi_L^2} \\
&\quad + 4\alpha \frac{Y_{t,i}^2}{\omega_L^2} + 4\alpha + 2\alpha \frac{4c^2\delta_U^2}{\varphi_L^2} + \frac{Y_{t,i}}{\omega_L^2} + \frac{2c^2\delta_U(1 + 2\delta_U)}{\varphi_L^2} \\
&= \frac{4\alpha}{\omega_L^2} Y_{t,i}^2 + \frac{1}{\omega_L^2} Y_{t,i} + \frac{3 + 2\alpha}{\omega_L} + \frac{4c^2\delta_U\{4(1+\alpha)\delta_U + 1\}}{\varphi_L^2} + 4\alpha
\end{aligned}
$$

for $(i,j) = (1,2), (2,1)$,

$$
\begin{aligned}
|F_{t,33}(\theta)| &\leq \alpha \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta(\boldsymbol{y}|\boldsymbol{\lambda}_t) \left\{ \frac{u(y_1, \lambda_{t,1}) u(y_2, \lambda_{t,2})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\}^2 \\
&\quad + (1+\alpha) \left\{ \frac{u(Y_{t,1}, \lambda_{t,1}) u(Y_{t,2}, \lambda_{t,2})}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\}^2 \\
&\leq \frac{16\alpha}{\varphi_L^2} + \frac{16(1+\alpha)}{\varphi_L^2} \\
&= \frac{16(1+2\alpha)}{\varphi_L^2},
\end{aligned}
$$

$$
\begin{aligned}
|F_{t,12}(\theta)| &\leq \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta(\boldsymbol{y}|\boldsymbol{\lambda}_t) \left[(1+\alpha) \left\{ \frac{y_1}{\lambda_{t,1}} + 1 + \frac{c|\delta|e^{-c\lambda_{t,1}}|u(y_2, \lambda_{t,2})|}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\} \right. \\
&\quad \left. \times \left\{ \frac{y_2}{\lambda_{t,2}} + 1 + \frac{c|\delta|e^{-c\lambda_{t,2}}|u(y_1, \lambda_{t,1})|}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\} + \frac{c^2|\delta|e^{-c(\lambda_{t,1}+\lambda_{t,2})}}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)^2} \right] \\
&\quad + \alpha \left\{ \frac{Y_{t,1}}{\lambda_{t,1}} + 1 + \frac{c|\delta|e^{-c\lambda_{t,1}}|u(Y_{t,2}, \lambda_{t,2})|}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\} \\
&\quad \times \left\{ \frac{Y_{t,2}}{\lambda_{t,2}} + 1 + \frac{c|\delta|e^{-c\lambda_{t,2}}|u(Y_{t,1}, \lambda_{t,1})|}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\} + \frac{c^2|\delta|e^{-c(\lambda_{t,1}+\lambda_{t,2})}}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)^2} \\
&\leq (1+\alpha) \left\{ c^2\delta_U + 1 + \left(1 + \frac{2c\delta_U}{\varphi_L}\right) + \left(1 + \frac{2c\delta_U}{\varphi_L}\right) + \frac{4c^2\delta_U^2}{\varphi_L^2} + \frac{4c\delta_U}{\varphi_L} + 1 \right\} \\
&\quad + \frac{c^2\delta_U}{\varphi_L^2} + \alpha \left\{ \frac{Y_{t,1}Y_{t,2}}{\omega_L^2} + \frac{1}{\omega_L}\left(1 + \frac{2c\delta_U}{\varphi_L}\right)(Y_{t,1} + Y_{t,2}) + \frac{4c^2\delta_U^2}{\varphi_L^2} + \frac{4c\delta_U}{\varphi_L} + 1 \right\} \\
&\quad + \frac{c^2\delta_U}{\varphi_L^2} \\
&= \frac{\alpha}{\omega_L^2} Y_{t,1}Y_{t,2} + \frac{\alpha}{\omega_L}\left(1 + \frac{2c\delta_U}{\varphi_L}\right)(Y_{t,1} + Y_{t,2}) + 4(1+2\alpha)\frac{c^2\delta_U^2}{\varphi_L^2} + \frac{2c^2\delta_U}{\varphi_L^2} \\
&\quad + 4(2+3\alpha)\frac{c\delta_U}{\varphi_L} + (1+\alpha)c^2\delta_U + 4 + 5\alpha,
\end{aligned}
$$

and

$$
\begin{aligned}
|F_{t,i3}(\theta)| &\leq \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} f_\theta(\boldsymbol{y}|\boldsymbol{\lambda}_t) \left[(1+\alpha) \left\{ \frac{y_i}{\lambda_{t,i}} + 1 + \frac{c|\delta|e^{-c\lambda_{t,i}}|u(y_j, \lambda_{t,j})|}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} \right\} \right. \\
&\quad \left. \times \frac{|u(y_1, \lambda_{t,1})||u(y_2, \lambda_{t,2})|}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)} + \frac{ce^{-c\lambda_{t,i}}|u(y_j, \lambda_{t,j})|}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}_t, \delta)^2} \right] \\
&\quad + \alpha \left\{ \frac{Y_{t,i}}{\lambda_{t,i}} + 1 + \frac{c|\delta|e^{-c\lambda_{t,i}}|u(Y_{t,j}, \lambda_{t,j})|}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \right\} \frac{|u(Y_{t,1}, \lambda_{t,1})||u(Y_{t,2}, \lambda_{t,2})|}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)} \\
&\quad + \frac{ce^{-c\lambda_{t,i}}|u(Y_{t,j}, \lambda_{t,j})|}{\varphi(\boldsymbol{Y}_t, \boldsymbol{\lambda}_t, \delta)^2} \\
&\leq \frac{4(1+\alpha)}{\varphi_L}\left(1 + 1 + \frac{2c\delta_U}{\varphi_L}\right) + \frac{2c}{\varphi_L^2} + \frac{4\alpha}{\varphi_L}\left(\frac{Y_{t,i}}{\omega_L} + 1 + \frac{2c\delta_U}{\varphi_L}\right) + \frac{2c}{\varphi_L^2} \\
&= \frac{4\alpha}{\varphi_L\omega_L} Y_{t,i} + \frac{4c\{1 + 2(1+2\alpha)\delta_U\}}{\varphi_L^2} + \frac{4(2+3\alpha)}{\varphi_L}
\end{aligned}
$$

for $(i, j) = (1, 2), (2, 1)$.

Now, we prove the last two parts of the lemma. Because $F_{t,ij}(\theta) = \partial D_{t,i}(\theta)/\partial \lambda_{t,j}$ for $i = 1, 2, 3$, $j = 1, 2$, owing to MVT, it holds that for $i = 1, 2, 3$,

$$
\begin{aligned}
|D_{t,i}(\theta) - \widetilde{D}_{t,i}(\theta)| &\leq \left| \frac{\partial D_{t,i}(\theta)}{\partial \lambda_{t,1}} \right|_{\lambda_t = \lambda_t^*} |\lambda_{t,1} - \tilde{\lambda}_{t,1}| + \left| \frac{\partial D_{t,i}(\theta)}{\partial \lambda_{t,2}} \right|_{\lambda_t = \lambda_t^*} |\lambda_{t,2} - \tilde{\lambda}_{t,2}| \\
&= \left| F_{t,i1}(\theta) \right|_{\lambda_t = \lambda_t^*} |\lambda_{t,1} - \tilde{\lambda}_{t,1}| + \left| F_{t,i2}(\theta) \right|_{\lambda_t = \lambda_t^*} |\lambda_{t,2} - \tilde{\lambda}_{t,2}|,
\end{aligned}
$$

where $F_{t,ij}(\theta)\big|_{\lambda_t = \lambda_t^*}$ is the same as $F_{t,ij}(\theta)$ with λ_t replaced by λ_t^* for $j = 1, 2$. Because $(\lambda_{t,i}^*)^{-1} \leq \omega_L^{-1}$, it can be easily shown that $\left| F_{t,ij}(\theta)\big|_{\lambda_t = \lambda_t^*} \right|$ has the same upper bound as $\left| F_{t,ij}(\theta) \right|$ by following the aforementioned arguments. Therefore, the lemma is established. $\quad\square$

Lemma A5. *Under (A1)–(A3), we have*

$$
E \left(\sup_{\theta \in \Theta} \left\| \frac{\partial^2 h_{\alpha,t}(\theta)}{\partial \theta \partial \theta^T} \right\|_1 \right) < \infty \ \text{and} \ E \left(\sup_{\theta \in \Theta} \left\| \frac{\partial h_{\alpha,t}(\theta)}{\partial \theta} \frac{\partial h_{\alpha,t}(\theta)}{\partial \theta^T} \right\|_1 \right) < \infty.
$$

Proof. We can write

$$
E \left(\sup_{\theta \in \Theta} \left\| \frac{\partial^2 h_{\alpha,t}(\theta)}{\partial \theta \partial \theta^T} \right\|_1 \right) \leq (1 + \alpha) \left\{ E \left(\sup_{\theta \in \Theta} \| F_t(\theta) \|_1 \right) + E \left(\sup_{\theta \in \Theta} \left\| D_t(\theta) \frac{\partial \Lambda_t(\theta)}{\partial \theta^T} \right\|_1 \right) \right\}.
$$

Hence, to show the first part of the lemma, it is sufficient to show that, for $i, j = 1, 2$,

$$
E \left(\sup_{\theta \in \Theta} \left\| F_{t,ij}(\theta) \frac{\partial \lambda_{t,i}}{\partial \theta_i} \frac{\partial \lambda_{t,j}}{\partial \theta_j^T} \right\|_1 \right) < \infty, \ E \left(\sup_{\theta \in \Theta} \left\| F_{t,i3}(\theta) \frac{\partial \lambda_{t,i}}{\partial \theta_i} \right\|_1 \right) < \infty,
$$

$$
E \left(\sup_{\theta \in \Theta} |F_{t,33}(\theta)| \right) < \infty, \ \text{and} \ E \left(\sup_{\theta \in \Theta} \left\| D_{t,i}(\theta) \frac{\partial^2 \lambda_{t,i}}{\partial \theta_i \partial \theta_i^T} \right\|_1 \right) < \infty,
$$

which can be directly obtained from (iii) of Lemma A1, Lemma A4, and Cauchy–Schwarz inequality. For example,

$$
\begin{aligned}
&E \left(\sup_{\theta \in \Theta} \left\| F_{t,12}(\theta) \frac{\partial \lambda_{t,1}}{\partial \theta_1} \frac{\partial \lambda_{t,2}}{\partial \theta_2^T} \right\|_1 \right) \\
&\leq \left\{ E \left(\sup_{\theta \in \Theta} |F_{t,12}(\theta)| \right)^2 \right\}^{1/2} \left\{ E \left(\sup_{\theta \in \Theta} \left\| \frac{\partial \lambda_{t,1}}{\partial \theta_1} \frac{\partial \lambda_{t,2}}{\partial \theta_2^T} \right\|_1 \right)^2 \right\}^{1/2} \\
&\leq \left[E \{ C(Y_{t,1} Y_{t,2} + Y_{t,1} + Y_{t,2} + 1) \}^2 \right]^{1/2} \\
&\quad \times \left\{ E \left(\sup_{\theta_1 \in \Theta_1} \left\| \frac{\partial \lambda_{t,1}}{\partial \theta_1} \right\|_1 \right)^4 \right\}^{1/4} \left\{ E \left(\sup_{\theta_2 \in \Theta_2} \left\| \frac{\partial \lambda_{t,2}}{\partial \theta_2} \right\|_1 \right)^4 \right\}^{1/4} \\
&< \infty
\end{aligned}
$$

and

$$E\left(\sup_{\theta\in\Theta}\left\|D_{t,1}(\theta)\frac{\partial^2\lambda_{t,1}}{\partial\theta_1\partial\theta_1^T}\right\|_1\right)$$

$$\leq \quad \left\{E\left(\sup_{\theta\in\Theta}|D_{t,1}(\theta)|\right)^2\right\}^{1/2}\left\{E\left(\sup_{\theta_1\in\Theta_1}\left\|\frac{\partial^2\lambda_{t,1}}{\partial\theta_1\partial\theta_1^T}\right\|_1\right)^2\right\}^{1/2}$$

$$\leq \quad \left[E\{C(Y_{t,1}+1)\}^2\right]^{1/2}\left\{E\left(\sup_{\theta_1\in\Theta_1}\left\|\frac{\partial^2\lambda_{t,1}}{\partial\theta_1\partial\theta_1^T}\right\|_1\right)^2\right\}^{1/2}$$

$$< \quad \infty.$$

The second part of the lemma can be shown in the same manner. □

Lemma A6. *Under (A1)–(A3), we have*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left\|\frac{\partial h_{\alpha,t}(\theta)}{\partial\theta}-\frac{\partial\tilde{h}_{\alpha,t}(\theta)}{\partial\theta}\right\|_1\xrightarrow{a.s.}0 \ as \ n\to\infty.$$

Proof. Owing to (iv) and (vi) of Lemma A1 and Lemma A4, we obtain a.s.,

$$\frac{1}{1+\alpha}\sup_{\theta\in\Theta}\left\|\frac{\partial h_{\alpha,t}(\theta)}{\partial\theta}-\frac{\partial\tilde{h}_{\alpha,t}(\theta)}{\partial\theta}\right\|_1$$

$$\leq \quad \sup_{\theta\in\Theta}\left\|\tilde{D}_t(\theta)\right\|_1\sup_{\theta\in\Theta}\left\|\Lambda_t(\theta)-\tilde{\Lambda}_t(\theta)\right\|_1+\sup_{\theta\in\Theta}\|\Lambda_t(\theta)\|_1\sup_{\theta\in\Theta}\left\|D_t(\theta)-\tilde{D}_t(\theta)\right\|_1$$

$$\leq \quad \left(\sum_{i=1}^{3}\sup_{\theta\in\Theta}|\tilde{D}_{t,i}(\theta)|\right)\left(\sum_{i=1}^{2}\sup_{\theta_i\in\Theta_i}\left\|\frac{\partial\lambda_{t,i}}{\partial\theta_i}-\frac{\partial\tilde{\lambda}_{t,i}}{\partial\theta_i}\right\|_1\right)$$

$$\quad +\left(\sum_{i=1}^{2}\sup_{\theta_i\in\Theta_i}\left\|\frac{\partial\lambda_{t,i}}{\partial\theta_i}\right\|_1+1\right)\left(\sum_{i=1}^{3}\sup_{\theta\in\Theta}|D_{t,i}(\theta)-\tilde{D}_{t,i}(\theta)|\right)$$

$$\leq \quad 2C(Y_{t,1}+Y_{t,2}+3)V\rho^t$$

$$\quad +\left(\sum_{i=1}^{2}\sup_{\theta_i\in\Theta_i}\left\|\frac{\partial\lambda_{t,i}}{\partial\theta_i}\right\|_1+1\right)\times C\{Y_{t,1}^2+Y_{t,2}^2+2Y_{t,1}Y_{t,2}+4(Y_{t,1}+Y_{t,2})+6\}V\rho^t,$$

where $\tilde{D}_t(\theta)$ and $\tilde{\Lambda}_t(\theta)$ are the same as $D_t(\theta)$ and $\Lambda_t(\theta)$ with λ_t replaced by $\tilde{\lambda}_t$. Therefore, from Lemma 2.1 in Straumann and Mikosch [45], together with (iii) of Lemma A1, the RHS of the last inequality converges to 0 exponentially fast a.s. and, thus, the lemma is validated. We refer the reader to Straumann and Mikosch [45] and Cui and Zheng [46] for more details on exponentially fast a.s. convergence. □

Lemma A7. *Let $\hat{\theta}_{\alpha,n}^H=\operatorname{argmin}_{\theta\in\Theta}H_{\alpha,n}(\theta)$. Subsequently, under (A1)–(A3), 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 seen in the proof of Theorem 1, $\sup_{\theta\in\Theta}\left|n^{-1}\sum_{t=1}^{n}h_{\alpha,t}(\theta)-E(h_{\alpha,t}(\theta))\right|$ converges to 0 a.s. and $E(h_{\alpha,t}(\theta))$ has a unique minimum at θ^0. Hence, the first part of the lemma is verified.

Next, we handle the second part. Let $\theta(i), i=1,\ldots,9$ be the i-th element of θ. Using MVT, we have

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

for some vector $\theta^*_{\alpha,n,i}$ between θ_0 and $\hat{\theta}^H_{\alpha,n}$, so that, eventually, we can write

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

where the term $\partial^2 h_{\alpha,t}(\theta^*_{\alpha,n})/\partial\theta\partial\theta^T$ actually represents a 9×9 matrix whose (i,j)-th entry is $\partial^2 h_{\alpha,t}(\theta^*_{\alpha,n,ij})/\partial\theta(i)\partial\theta(j)$ for some vector $\theta^*_{\alpha,n,ij}$ between θ_0 and $\hat{\theta}^H_{\alpha,n}$. We first show that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial h_{\alpha,t}(\theta^0)}{\partial\theta} \xrightarrow{d} N(0, K_\alpha). \tag{A1}$$

For $\nu = (\nu_1^T, \nu_2^T, \nu_3)^T \in \mathbb{R}^4 \times \mathbb{R}^4 \times \mathbb{R}$, we obtain

$$
\begin{aligned}
E\left(\nu^T \frac{\partial h_{\alpha,t}(\theta^0)}{\partial\theta} \Big| \mathcal{F}_{t-1} \right) &= (1+\alpha)\Bigg\{ \nu_1^T \frac{\partial\lambda^0_{t,1}}{\partial\theta_1} E\left(D_{t,1}(\theta^0) | \mathcal{F}_{t-1} \right) \\
&\qquad + \nu_2^T \frac{\partial\lambda^0_{t,2}}{\partial\theta_2} E\left(D_{t,2}(\theta^0) | \mathcal{F}_{t-1} \right) + \nu_3 E\left(D_{t,3}(\theta^0) | \mathcal{F}_{t-1} \right) \Bigg\} \\
&= 0
\end{aligned}
$$

and

$$E\left(\nu^T \frac{\partial h_{\alpha,t}(\theta^0)}{\partial\theta} \right)^2 = \nu^T E\left(\frac{\partial h_{\alpha,t}(\theta^0)}{\partial\theta} \frac{\partial h_{\alpha,t}(\theta^0)}{\partial\theta^T} \right) \nu < \infty$$

by Lemma A5. Hence, it follows from the central limit theorem in Billingsley [47] that

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

which implies (A1).

Now, we claim that

$$-\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 h_{\alpha,t}(\theta^*_{\alpha,n,ij})}{\partial\theta(i)\partial\theta(j)} \xrightarrow{a.s.} J^{ij}_\alpha, \tag{A2}$$

where J^{ij}_α denotes the (i,j)-th entry of J_α. From Lemma A5, J_α is finite. Further, after some algebras, we have

$$
\begin{aligned}
&\nu^T(-J_\alpha)\nu \\
&= (1+\alpha)E\Bigg[\sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty} f^{1+\alpha}_{\theta^0}(\boldsymbol{y}|\boldsymbol{\lambda}_t) \Bigg\{ \left(\nu_1^T \frac{\partial\lambda^0_{t,1}}{\partial\theta_1} \right) \left(\frac{y_1}{\lambda^0_{t,1}} - 1 + \frac{c\delta^0 e^{-c\lambda^0_{t,1}} u(y_2, \lambda^0_{t,2})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}^0_t, \delta^0)} \right) \\
&\qquad + \left(\nu_2^T \frac{\partial\lambda^0_{t,2}}{\partial\theta_2} \right) \left(\frac{y_2}{\lambda^0_{t,2}} - 1 + \frac{c\delta^0 e^{-c\lambda^0_{t,2}} u(y_1, \lambda^0_{t,1})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}^0_t, \delta^0)} \right) \\
&\qquad + \nu_3 \frac{u(y_1, \lambda^0_{t,1}) u(y_2, \lambda^0_{t,2})}{\varphi(\boldsymbol{y}, \boldsymbol{\lambda}^0_t, \delta^0)} \Bigg\}^2 \Bigg] > 0
\end{aligned}
$$

by (v) of Lemma A1, which implies that J_α is non-singular. Note that we can write

$$\left| \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 h_{\alpha,t}(\theta^*_{\alpha,n,ij})}{\partial \theta(i) \partial \theta(j)} - E\left(\frac{\partial^2 h_{\alpha,t}(\theta^0)}{\partial \theta(i) \partial \theta(j)} \right) \right|$$
$$\leq \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 h_{\alpha,t}(\theta)}{\partial \theta(i) \partial \theta(j)} - E\left(\frac{\partial^2 h_{\alpha,t}(\theta)}{\partial \theta(i) \partial \theta(j)} \right) \right| + \left| E\left(\frac{\partial^2 h_{\alpha,t}(\theta^*_{\alpha,n,ij})}{\partial \theta(i) \partial \theta(j)} \right) - E\left(\frac{\partial^2 h_{\alpha,t}(\theta^0)}{\partial \theta(i) \partial \theta(j)} \right) \right|.$$

Because $\partial^2 h_{\alpha,t}(\theta)/\partial \theta(i) \partial \theta(j)$ is stationary and ergodic, from Lemma A5, the first term on the RHS of the inequality converges to 0 a.s. Moreover, the second term converges to 0 by the dominated convergence theorem. Hence, (A2) is asserted. Therefore, from (A1) and (A2), the second part of the lemma is established. $\square$

Proof of Theorem 2. From MVT, we get

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

for some vector $\zeta_{\alpha,n,i}$ between $\hat{\theta}^H_{\alpha,n}$ and $\hat{\theta}_{\alpha,n}$. Thus, we can write

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

where the (i,j)-th entry of $\partial^2 h_{\alpha,t}(\zeta_{\alpha,n})/\partial \theta \partial \theta^T$ is $\partial^2 h_{\alpha,t}(\zeta_{\alpha,n,ij})/\partial \theta(i) \partial \theta(j)$ for some vector $\zeta_{\alpha,n,ij}$ between $\hat{\theta}^H_{\alpha,n}$ and $\hat{\theta}_{\alpha,n}$. Since $n^{-1} \sum_{t=1}^n \partial h_{\alpha,t}(\hat{\theta}^H_{\alpha,n})/\partial \theta = 0$ and $n^{-1} \sum_{t=1}^n \partial \tilde{h}_{\alpha,t}(\hat{\theta}_{\alpha,n})/\partial \theta = 0$, we have

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

The LHS of the above equation converges to 0 a.s. by Lemma A6, and $n^{-1} \sum_{t=1}^n \partial^2 h_{\alpha,t}(\zeta_{\alpha,n})/\partial \theta \partial \theta^T$ converges to $-J_\alpha$ a.s. in a similar way as in the proof of Lemma A7. Therefore, the theorem is established due to Lemma A7. $\square$

References

1. Weiß, C.H. Thinning operations for modeling time series of counts-a survey. *AStA Adv. Stat. Anal.* **2008**, *92*, 319–341. [CrossRef]
2. Scotto, M.G.; Weiß, C.H.; Gouveia, S. Thinning-based models in the analysis of integer-valued time series: A review. *Stat. Model.* **2015**, *15*, 590–618. [CrossRef]
3. Ferl, ; R.; Latour, A.; Oraichi, D. Integer-valued GARCH processes. *J. Time Ser. Anal.* **2006**, *27*, 923–942. [CrossRef]
4. Fokianos, K.; Rahbek, A.; Tjøstheim, D. Poisson autoregression. *J. Am. Stat. Assoc.* **2009**, *104*, 1430–1439. [CrossRef]
5. Davis, R.A.; Wu, R. A negative binomial model for time series of counts. *Biometrika* **2009**, *96*, 735–749. [CrossRef]
6. Christou, V.; Fokianos, K. Quasi-likelihood inference for negative binomial time series models. *J. Time Ser. Anal.* **2014**, *35*, 55–78. [CrossRef]
7. Zhu, F. Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *J. Math. Anal. Appl.* **2012**, *389*, 58–71. [CrossRef]
8. Zhu, F. Zero-inflated Poisson and negative binomial integer-valued GARCH models. *J. Stat. Plan. Infer.* **2012**, *142*, 826–839. [CrossRef]
9. Lee, S.; Lee, Y.; Chen, C.W.S. Parameter change test for zero-inflated generalized Poisson autoregressive models. *Statistics* **2016**, *50*, 540–557. [CrossRef]
10. 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.
11. Fokianos, K. Count time series models. In *Handbook of Statistics: Time Series Analysis-Methods and Applications*; Rao, T.S., Rao, S.S., Rao, C.R., Eds.; Elsevier: Amsterdam, The Netherlands, 2012; Volume 30, pp. 315–347.
12. Fokianos, K. Statistical analysis of count time series models: A GLM perspective. In *Handbook of Discrete-Valued Time Series*; Davis, R.A., Holan, S.H., Lund,R., Ravishanker, N., Eds.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2016; pp. 3–27.

13. Tjøstheim, D. Some recent theory for autoregressive count time series (with discussions). *Test* **2012**, *21*, 413–476. [CrossRef]
14. Tjøstheim, D. Count time series with observation-driven autoregressive parameter dynamics. In *Handbook of Discrete-Valued Time Series*; Davis, R.A., Holan, S.H., Lund, R., Ravishanker, N., Eds.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2016; pp. 77–100.
15. Quoreshi, A.S. Bivariate time series modeling of financial count data. *Commun. Stat. Theory Methods* **2006**, *35*, 1343–1358. [CrossRef]
16. Pedeli, X.; Karlis, D. A bivariate INAR(1) process with application. *Stat. Model.* **2011**, *11*, 325–349. [CrossRef]
17. Liu, H. Some Models for Time Series of Counts. Ph.D. Thesis, Columbia University, New York, NY, USA, 2012.
18. Andreassen, C.M. Models and Inference for Correlated Count Data. Ph.D. Thesis, Aarhus University, Aarhus, Denmark, 2013.
19. Lee, Y.; Lee, S.; Tjøstheim, D. Asymptotic normality and parameter change test for bivariate Poisson INGARCH models. *Test* **2018**, *27*, 52–69. [CrossRef]
20. Cui, Y.; Zhu, F. A new bivariate integer-valued GARCH model allowing for negative cross-correlation. *Test* **2018**, *27*, 428–452. [CrossRef]
21. Lakshminarayana, J.; Pandit, S.N.N.; Rao, K.S. On a bivariate Poisson distribution. *Commun. Stat. Theory Methods* **1999**, *28*, 267–276. [CrossRef]
22. 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]
23. Kang, J.; Lee, S. Minimum density power divergence estimator for Poisson autoregressive models. *Comput. Stat. Data Anal.* **2014**, *80*, 44–56. [CrossRef]
24. 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]
25. Kim, B.; Lee, S. Robust estimation for general integer-valued time series models. *Ann. Inst. Stat. Math.* **2020**, *72*, 1371–1396. [CrossRef]
26. Diop, M.L.; Kengne, W. Density power divergence for general integer-valued time series with multivariate exogenous covariate. *arXiv* **2020**, arXiv:2006.11948.
27. Kim, B.; Lee, S. Robust change point test for general integer-valued time series models based on density power divergence. *Entropy* **2020**, *22*, 493. [CrossRef]
28. Lee, S.; Kim, D. Monitoring parameter change for time series models of counts based on minimum density power divergence estimator. *Entropy* **2020**, *22*, 1304. [CrossRef]
29. Xiong, L.; Zhu, F. Robust quasi-likelihood estimation for the negative binomial integer-valued GARCH(1,1) model with an application to transaction counts. *J. Stat. Plan. Infer.* **2019**, *203*, 178–198. [CrossRef]
30. Li, Q.; Chen, H.; Zhu, F. Robust estimation for Poisson integer-valued GARCH models using a new hybrid loss. *J. Syst. Sci. Complex.* **2021**, in press. [CrossRef]
31. Heinen, A.; Rengifo, E. Multivariate modeling of time series count data: An AR conditional Poisson model. In *CORE Discussion Paper*; Université Catholique de Louvain, Center for Operations Research and Econometrics (CORE): Ottignies-Louvain-la-Neuve, Belgium, 2003.
32. Cui, Y.; Li, Q.; Zhu, F. Flexible bivariate Poisson integer-valued GARCH model. *Ann. Inst. Stat. Math.* **2020**, *72*, 1449–1477. [CrossRef]
33. Fujisawa, H.; Eguchi, S. Robust estimation in the normal mixture model. *J. Stat. Plan. Infer.* **2006**, *136*, 3989–4011. [CrossRef]
34. Durio, A.; Isaia, E.D. The minimum density power divergence approach in building robust regression models. *Informatica* **2011**, *22*, 43–56. [CrossRef]
35. Toma, A.; Broniatowski, M. Dual divergenceestimators and tests: Robustness results. *J. Multivar. Anal.* **2011**, *102*, 20–36. [CrossRef]
36. Warwick, J. A data-based method for selecting tuning parameters in minimum distance estimators. *Comput. Stat. Data Anal.* **2005**, *48*, 571–585. [CrossRef]
37. Verges, Y. The Bivariate Integer-Valued GARCH Model: A Bayesian Estimation Framework. Master's Thesis, University of São Paulo, São Paulo, Brazil, 2019.
38. Chen, C.W.S.; Lee, S. Generalized Poisson autoregressive models for time series of counts. *Comput. Stat. Data Anal.* **2016**, *99*, 51–67. [CrossRef]
39. Chen, C.W.S.; Lee, S. Bayesian causality test for integer-valued time series models with applications to climate and crime data. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2017**, *66*, 797–814. [CrossRef]
40. Lee, S.; Park, S.; Chen, C.W.S. On Fisher's dispersion test for integer-valued autoregressive Poisson models with applications. *Commun. Stat. Theory Methods* **2017**, *46*, 9985–9994. [CrossRef]
41. Fokianos, K.; Fried, R. Interventions in INGARCH processes. *J. Time Ser. Anal.* **2010**, *31*, 210–225. [CrossRef]
42. Heinen, A.; Rengifo, E. Multivariate autoregressive modeling of time series count data using copulas. *J. Empir. Financ.* **2007**, *14*, 564–583. [CrossRef]
43. Fokianos, K.; Støve, B.; Tjøstheim, D.; Doukhan, P. Multivariate count autoregression. *Bernoulli* **2020**, *26*, 471–499. [CrossRef]
44. Kang, J.; Lee, S. Parameter change test for Poisson autoregressive models. *Scand. J. Stat.* **2014**, *41*, 1136–1152. [CrossRef]
45. 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]

46. Cui, Y.; Zheng, Q. Conditional maximum likelihood estimation for a class of observation-driven time series models for count data. *Stat. Probab. Lett.* **2017**, *123*, 193–201. [CrossRef]
47. Billingsley, P. The Lindeberg-Lévy theorem for martingales. *Proc. Am. Math. Soc.* **1961**, *12*, 788–792.