MDPI

Article

Consistent Model Selection Procedure for Random Coefficient INAR Models

Kaizhi Yu and Tielai Tao *

School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China; yukz@swufe.edu.cn
* Correspondence: 118020209005@smail.swufe.edu.cn

Abstract: In the realm of time series data analysis, information criteria constructed on the basis of likelihood functions serve as crucial instruments for determining the appropriate lag order. However, the intricate structure of random coefficient integer-valued time series models, which are founded on thinning operators, complicates the establishment of likelihood functions. Consequently, employing information criteria such as AIC and BIC for model selection becomes problematic. This study introduces an innovative methodology that formulates a penalized criterion by utilizing the estimation equation within conditional least squares estimation, effectively addressing the aforementioned challenge. Initially, the asymptotic properties of the penalized criterion are derived, followed by a numerical simulation study and a comparative analysis. The findings from both theoretical examinations and simulation investigations reveal that this novel approach consistently selects variables under relatively relaxed conditions. Lastly, the applications of this method to infectious disease data and seismic frequency data produce satisfactory outcomes.

Keywords: integer-valued time series; model selection; thinning operator; conditional least squares; information criteria

1. Introduction

Integer-valued time series are ubiquitous in scientific research and everyday life, encompassing examples such as the daily count of hospitalized patients admitted to hospitals and the frequency of crimes committed daily or monthly. Consequently, integer-valued time series have increasingly garnered attention from scholars. However, traditional continuous-valued time series models fail to capture the integer-valued characteristics, only approximating integer-valued data through continuous-valued time series models. This approximation may result in model misspecification issues, complicating statistical inference. As a result, the modeling and analysis of integer-valued time series data have become a growing area of focus in academia. Among the variety of integer-valued time series modeling methods, thinning operator models have gained favor due to their resemblance to autoregressive moving average (ARMA) models found in traditional continuous-valued time series theory. Thinning operator models substitute the multiplication in ARMA models with the binomial thinning operator introduced by Steutel and Van Harn [1]:

$$\phi \circ Y_i = \sum_{i=1}^{Y_i} B_i \tag{1}$$

In this equation, Y_i represents a count sequence, while $\{B_i\}$ denotes a series of Bernoulli random variables independent of $\{Y_i\}$. The probability mass function satisfies $P(B_i = 1) = 1 - P(B_i = 0) = \phi$ with $\phi \in [0, 1)$. Building on this foundation, Al-Osh and Alzaid [2] developed the first-order integer-valued autoregressive (INAR (1)) model for $t \in \mathbb{N}^+$:

$$Y_t = \phi \circ Y_{t-1} + Z_t \tag{2}$$

where Z_t is regarded as the innovation term entering the model at period t, with its marginal distribution being a Poisson distribution with an expected value of λ. Consequently, model (2) is called the Poisson INAR(1) model. Later, Du and Li [3] introduced the INAR(p) model and provided conditions for ensuring the stationarity and ergodicity of the INAR(p) process. The incorporation of additional lag terms increased the model's flexibility. Subsequently, Joe [4] and Zheng, Basawa, and Datta [5] developed the random coefficient thinning operator model (RCINAR(1)) by allowing the parameter ϕ in the INAR(1) model to follow a specific random distribution. Zheng, Basawa, and Datta [6] extended the RCINAR(1) model to the p-th order integer-valued autoregressive model, known as the RCINAR(p) model. Zhang, Wang, and Zhu [7] established a data-driven empirical likelihood interval estimation for the RCINAR(p) model using the empirical likelihood (EL) estimation method. By employing the geometric thinning operator (also referred to as the negative binomial thinning operator) proposed by Ristić, Bakouch, and Nastić [8], Tian, Wang, and Cui [9]) constructed an INAR(1) model capable of describing seasonal effects. Lu [10] investigated the prediction problem of the thinning operator model using the Taylor expansion. For further discussions on thinning operator models, readers can consult the textbook by Weiß [11].

In general, researchers engaged in statistical analysis, particularly during the initial stages of time series data investigation, frequently encounter the challenge of model selection. Current model selection techniques can be broadly categorized into three groups: The first group relies on sample autocorrelation (ACF) and partial autocorrelation (PACF) functions for model selection, as exemplified by Latour [12]; the second group, which is the most prevalent method for variable selection, comprises a series of information criteria founded on maximum likelihood estimation. Akaike [13] introduced the Akaike Information Criterion (AIC) by performing an unbiased estimation of the expected log-likelihood function, while Schwarz [14] established the Bayesian Information Criterion (BIC) by employing a Laplace expansion for the posterior estimation of the expected log-likelihood function. Ding, Tarokh, and Yang [15] devised a novel information criterion for autoregressive time series models by connecting AIC and BIC. Furthermore, given that empirical likelihood estimation can substantially circumvent issues stemming from model misspecification and maintain certain maximum likelihood estimation features, researchers have started to investigate data-driven information criteria based on empirical likelihood estimation. Variyath, Chen, and Abraham [16] formulated the Empirical Akaike Information Criterion (EAIC) and the Empirical Bayesian Information Criterion (EBIC) by drawing on the principles of AIC and BIC with empirical likelihood estimation. They demonstrated that EBIC possesses consistency in variable selection. Chen, Wang, Wu, and Li [17] addressed potential computational convergence problems in empirical likelihood estimation by incorporating external estimators (typically moment estimators) into the empirical likelihood function, thereby developing a robust and consistent information criterion. For additional discussions on information criteria, readers may consult the textbook by Konishi and Kitagawa [18] and the review article by Ding, Tarokh, and Hong [19].

In the specific domain of integer-valued time series analysis, our objective is to determine which lagged variables of Y_t ought to be incorporated into the model. Extensive research has been conducted on model selection for integer-valued autoregressive conditional heteroskedasticity (INARCH) models, which allow for relatively straightforward likelihood function establishment. Notable examples include Weiß and Feld [20], who provided comprehensive numerical simulations for integer-valued time series model selection using information criteria, and Diop and Kengne [21], who introduced consistent model selection methods for INARCH models based on quasi-maximum likelihood estimation. However, the process becomes more challenging when dealing with higher-order and random coefficient INAR(p) models constructed using thinning operators. The complexity of the likelihood functions and the substantial computational requirements make it difficult to establish and utilize information criteria. Consequently, Zheng, Basawa, and Datta [6] proposed estimating the model based on its conditional moments rather than

relying on likelihood functions. While this approach facilitates the estimation of unknown parameters for researchers, it creates complications for variable selection. To overcome this hurdle, Wang, Wang, and Yang [22] implemented penalty functions and pseudo-quasi-maximum likelihood estimation (PQML) for variable selection, demonstrating the robustness of their method even when faced with contaminated data. Drawing inspiration from these preceding studies, this paper endeavors to establish a novel model selection method akin to information criteria founded upon the estimating equations in conditional least squares (CLS) estimation. Furthermore, we attempt to demonstrate the consistency of this innovative model selection method in addressing variable selection problems within integer-valued time series. This approach circumvents the need for complex probability distribution assumptions while preserving effective variable selection capabilities.

The organization of this paper is as follows: In Section 2, we revisit the RCINAR(p) model, introduce the proposed information criterion, and outline its asymptotic properties. In Section 3, we carry out numerical simulation studies on variable selection utilizing this information criterion. In Section 4, we endeavor to apply this information criterion for variable selection in real data sets. Lastly, in Section 5, we engage in a discussion and offer concluding remarks.

2. RCINAR Model and Model Selection Procedure

In this section, we discuss the ergodic stationary RCINAR model and its associated model selection methods.

2.1. RCINAR(p) Model and Its Estimation

The INAR(p) model with constant coefficients, as introduced by Du and Li [3], is formulated as follows:

$$Y_t = \phi_1 \circ Y_{t-1} + \cdots + \phi_p \circ Y_{t-p} + Z_t \tag{3}$$

In this expression, given the vector $(Y_{t-1}, \cdots, Y_{t-p})'$, the elements $(\phi_1 \circ Y_{t-1}, \cdots, \phi_p \circ Y_{t-p})'$ are deemed to be mutually conditionally independent. This conditional independence ensures that the autocorrelation function of the INAR(p) model is congruent with that of its continuous-valued Autoregressive (AR(p)) counterpart. Moreover, Du and Li [3] substantiated that, under these model settings, the stationarity condition for the INAR(p) model necessitates that the roots of the polynomial $h(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0$ are located outside the unit circle. This implies that the INAR(p) model attains stationarity when the sum $\sum_{i=1}^{p} \phi_i$ is less than 1. Building upon these foundational insights, Zheng, Basawa, and Datta [6] extended the INAR(p) model under the constant coefficient assumption, giving rise to the Random Coefficient Integer-valued Autoregressive (RCINAR(p)) model.

Let $\{Y_t\}_{t=1}^{T}$ represent a non-negative integer-valued sequence. The RCINAR(p) model is defined by the following equation:

$$Y_t = \phi_1^{(t)} \circ Y_{t-1} + \cdots + \phi_p^{(t)} \circ Y_{t-p} + Z_t \tag{4}$$

where "$\circ$" denotes the thinning operator defined in Equation (1). Let $\theta_0 = (\phi_{10}, \ldots, \phi_{p0}, \lambda_0)'$ be the true parameter vector of this data-generating process, with $\theta_0 \in \Theta$, where Θ is a compact subset of $\mathbb{R}^{p+1}$. $\theta = (\phi_1, \ldots, \phi_p, \lambda)'$ represents the $p + 1$ dimensional parameter vector to be estimated. Here, $\left\{\phi_j^{(t)}\right\}$ are sequences of independent and identically distributed random variables defined on $[0, 1)$ with a mean of ϕ_j, and their probability density function $f_{\phi_j}(\phi) \geq 0, \forall \phi \in [0, 1)$, with $\sum_{j=1}^{p} \phi_j < 1$.

Moreover, we do not assume a specific parametric distribution for $\{Z_t\}$, only requiring that $\{Z_t\}$ be an independent and identically distributed non-negative integer-valued random variable sequence with a mean of λ and a probability mass function $f_Z(z) \geq 0$,

$\forall z \in \mathbb{N}$. In this context, we consider the semiparametric INAR model as described by Drost, Van den Akker, and Werker [23].

Remark 1. *As can be discerned from the preceding discussion, the INAR(p) model (3) represents a special case of the RCINAR(p) model (4). That is, when $\left\{\phi_j^{(t)}\right\}$ is a constant coefficient vector, the RCINAR(p) model reduces to the INAR(p) model. As demonstrated by Zheng, Basawa, and Datta [6], the statistical methods employed in the study of the RCINAR(p) model can also be directly applied to the INAR(p) model. Consequently, in order to cater to a wider range of application scenarios, the academic community tends to prioritize the study of the RCINAR model while investigating thinning operator models. For instance, Kang and Lee [24] investigated the problem of change-point detection in the RCINAR model by leveraging the Cumulative Sum (CUSUM) test. Similarly, Zhang, Wang, and Zhu [7] proposed an interval estimation method for the RCINAR model based on empirical likelihood estimation. Awale, Balakrishna, and Ramanathan [25], on the other hand, constructed a locally most powerful-type test devised specifically for examining structural changes within the framework of the RCINAR model. Therefore, this paper will center its research on the RCINAR model.*

To estimate the RCINAR(p) model and establish model selection criteria, we draw inspiration from the assumptions delineated by Zhang, Wang, and Zhu [7]. These assumptions are as follows:

(A1) $\{Y_t\}$ constitutes an ergodic and strictly stationary RCINAR(p) process.

(A2) There exists $\delta > 0$ such that $\mathbb{E}|Y_t|^{4+\delta} < \infty$.

Derived from Equation (4), the one-step-ahead transition probability is as follows:

$$
\mathbb{P}\left(Y_t = i \mid Y_{t-1} = i_1, \cdots, Y_{t-p} = i_p\right)
$$
$$
= \sum_{k=0}^{\min(i, \sum_{j=1}^{p} i_j)} f_Z(i-k) \prod_{0 \leq \sum_{j=1}^{p} k_j \leq k} \binom{i_j}{k_j}
$$
$$
\times \int_{0 \leq \phi_1^{(t)} \leq \cdots \leq \phi_p^{(t)} < 1} \prod_{j=1}^{p} \left(\phi_j^{(t)}\right)^{k_j} \left(1 - \phi_j^{(t)}\right)^{i_j - k_j} d\mathbb{P}_{\phi_1^{(t)}, \cdots, \phi_p^{(t)}}
$$

Here, $\mathbb{P}_{\phi_1^{(t)}, \cdots, \phi_p^{(t)}}$ represents the joint distribution function of $\phi_1^{(t)}, \cdots, \phi_p^{(t)}$. Utilizing this one-step-ahead transition probability function, we can construct the likelihood function:

$$
\mathfrak{L} = \mathbb{P}\left(Y_p = i_1^{(p+1)}, \cdots, Y_1 = i_p^{(p+1)}\right) \prod_{t=p+1}^{T} \mathbb{P}\left(Y_t = i^{(t)} \mid Y_{t-1} = i_1^{(t)}, \cdots, Y_{t-p} = i_p^{(t)}\right)
$$

The likelihood function $\mathfrak{L}$ for model (4) is notably complex, involving numerous multivariate numerical integrations within statistical computations, which demand substantial computational resources. Consequently, Zheng, Basawa, and Datta [6] advocated for estimating the model based on its conditional moments rather than employing the likelihood function. This preference also underlies the prevalent use of conditional least squares (CLS) estimation in the study of RCINAR(p) models within the scholarly community. In the subsequent section, we offer a concise introduction to the CLS estimation methodology for the RCINAR(p) model.

We can obtain the first-order conditional moment of model (4) as follows:

$$
\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \sum_{j=1}^{p} \phi_j Y_{t-j} + \lambda
$$

where $\mathcal{F}_{t-1} = \sigma(Y_{t-1}, Y_{t-2}, \cdots)$. This derivation allows us to compute the conditional least squares (CLS) estimation. Let

$$S(\theta) = \sum_{t=p+1}^{T} \left(Y_t - \sum_{j=1}^{p} \phi_j Y_{t-j} - \lambda \right)^2$$

represent the conditional least squares (CLS) objective function. The CLS estimator is then given by:

$$\hat{\theta} = argmin_\theta(S(\theta))$$

Let

$$S_t(\theta) = \left(Y_t - \sum_{j=1}^{p} \phi_j Y_{t-j} - \lambda \right)^2$$

Then the estimating equations are:

$$-\frac{1}{2}\frac{\partial S_t(\theta)}{\partial \theta} = 0 = \Psi_t(\theta) = \left(\psi_t^{(1)}(\theta), \psi_t^{(2)}(\theta), \ldots, \psi_t^{(p+1)}(\theta) \right)'$$

where

$$\psi_t^{(s)} = \left(Y_t - \sum_{j=1}^{p} \phi_j Y_{t-j} - \lambda \right) Y_{t-s}, \ 1 \leq s \leq p$$

$$\psi_t^{(p+1)} = Y_t - \sum_{j=1}^{p} \phi_j Y_{t-j} - \lambda$$

For the estimating equation $\Psi_t(\theta)$, we introduce an additional assumption:

(A3) $\Psi_t(\theta)$ is identifiable, that is, $\mathbb{E}(\Psi_t(\theta_0)) = 0$, and if θ is in the neighborhood of $\theta^* \neq \theta_0$, then $\|\mathbb{E}(\Psi_t(\theta))\|$ exists and $\|\mathbb{E}(\Psi_t(\theta))\| > 0$.

Assumption (A3) is the identifiability assumption, which further implies that the model (4) is identifiable if only the currently specified model satisfies $\mathbb{E}(\Psi_t(\theta_0)) = 0$. Based on these assumptions above, the following lemma can be deduced:

Lemma 1. *Based on assumptions (A1) to (A3), the subsequent conclusions are valid:*

(i) $\mathbb{E}\left(\Psi_t(\theta_0)\Psi_t(\theta_0)' \right)$ *constitutes a positive definite matrix.*

(ii) $\frac{\partial^2 \Psi_t(\theta)}{\partial\theta\partial\theta'}$ *remains continuous within the neighborhood of θ_0.*

(iii)   *Both* $\left\| \frac{\partial \Psi_t(\theta)}{\partial\theta'} \right\|$ *and* $\left\| \frac{\partial^2 \Psi_t(\theta)}{\partial\theta\partial\theta'} \right\|$ *possess upper bounds in the neighborhood of θ_0.*

Moreover, Zheng, Basawa, and Datta [6] established that $\hat{\theta}_{CLS}$ is a consistent estimator with an asymptotic distribution:

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(V^{-1}(\theta_0)W(\theta_0)V^{-1}(\theta_0)) \tag{5}$$

where:

$$W(\theta_0) = \mathbb{E}\left(\Psi_t(\theta_0)\Psi_t(\theta_0)' \right)$$

$$V(\theta_0) = \mathbb{E}\left(\frac{\partial\mathbb{E}(Y_t|Y_{t-1})}{\partial\theta} \cdot \frac{\partial\mathbb{E}(Y_t|Y_{t-1})}{\partial\theta'} \right) - \mathbb{E}\left(u_t(\theta_0)\frac{\partial^2\mathbb{E}(Y_t|Y_{t-1})}{\partial\theta\partial\theta'} \right)$$

$$u_t(\theta_0) = Y_t - \mathbb{E}\left(Y_t \Big| Y_{(t-1)} \right)$$

2.2. Model Selection Procedure

For the data-generating process defined by Equation (4), we establish the following settings:

1. A model m is a subset of $\mathcal{M} = \{1, 2, \ldots, p, p+1\}$, with its dimension denoted as $|m|$. Consequently, $p+1$ represents the maximum model dimension we consider, noted as the full model, while the minimum model dimension we consider is 1, corresponding to an independent and identically distributed non-negative integer-valued random variable sequence. Let the true model be m_0.

2. $\theta(m)$ is the parameter vector associated with model m, which can be extended to the p + 1 dimensional vector $\widetilde{\theta}(m) = \left\{ (\theta_j)_{1 \leq j \leq p+1} : \theta_j = \theta(m)_j, if\ j \in m; \theta_j = 0, if\ j \notin m \right\}$. For instance, if the considered model m is $Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_3^{(t)} \circ Y_{t-3} + Z_t$, then $m = \{1, 3, p+1\}$, $\theta(m) = (\phi_1, \phi_3, \lambda)$, and it can be extended to the $p+1$ dimensional vector $\widetilde{\theta}(m) = (\phi_1, 0, \phi_3, 0, \ldots, 0, \lambda)$.

3. Let $\Theta(m)$ be the compact parameter space of model m, $\widetilde{\Theta}(m) = \left\{ (\theta_j)_{1 \leq j \leq p+1} \in \mathbb{R}^{p+1} : \theta_j = 0, if\ j \notin m \right\}$ constitutes a compact subset of $\mathbb{R}^{p+1}$, and all possible $\widetilde{\theta}(m)$ values, when restricted to the $|m|$ dimensional vector $\theta(m)$, are interior points of its corresponding compact subset $\Theta(m)$. Furthermore, we denote $\widetilde{\theta} = \theta(\mathcal{M})$ as the parameter vector to be estimated in $\widetilde{\Theta}(\mathcal{M}) = \Theta(\mathcal{M})$, i.e., the parameter vector of the full model $\mathcal{M}$.

For model m, we partition $\widetilde{\theta}$ into two components, i.e., $\widetilde{\theta} = \left(\widetilde{\theta}_{(1)}(m)', \widetilde{\theta}_{(2)}(m)' \right)'$, where $\widetilde{\theta}_{(1)}(m) = \{(\theta_j), j \in m\} = \theta(m)$ and $\widetilde{\theta}_{(2)}(m) = \{(\theta_j), j \notin m\}$. Correspondingly, it is evident that if the model m is correctly specified, denoted as $m = m_0$, $\widetilde{\theta}_{(2)}(m_0) = 0$, then $\widetilde{\theta}_0 = \left(\widetilde{\theta}_{(1)}(m_0)', \widetilde{\theta}_{(2)}(m_0)' \right)' = \left(\theta(m_0)', 0' \right)'$. We can then divide the estimating equation $\Psi_t(\widetilde{\theta})$ into two parts:

$$\Psi_t(\widetilde{\theta}) = \begin{cases} \Psi_{1t}(\widetilde{\theta}) \\ \Psi_{2t}(\widetilde{\theta}) \end{cases}$$

where

$$\Psi_{1t}(\widetilde{\theta}) = -\frac{1}{2} \frac{\partial S_t(\widetilde{\theta})}{\partial \widetilde{\theta}_{(1)}(m)}$$

$$\Psi_{2t}(\widetilde{\theta}) = -\frac{1}{2} \frac{\partial S_t(\widetilde{\theta})}{\partial \widetilde{\theta}_{(2)}(m)}$$

Let $\hat{\theta}_{CLS}(m) = \left(\hat{\theta}'_{(1),CLS}(m), 0' \right)'$, i.e., $\hat{\theta}_{(1),CLS}(m)$ is the solution to $\Psi_{1t}\left(\widetilde{\theta}_{(1)}(m), 0 \right) = 0$, where $\widetilde{\theta}_{(2)}(m)$ is constrained to be 0. Therefore $\hat{\theta}_{(1),CLS}(m)$ represents the CLS estimator of model m. Define the function:

$$H(\widetilde{\theta}) = \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}) \right)' \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}) \Psi_t(\widetilde{\theta})' \right)^{-1} \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}) \right)$$

We can then derive the following lemma:

Lemma 2. *Given assumptions (A1)–(A3), as* $T \to \infty$ *:*

$$H(\widetilde{\theta}_0) \to \chi^2_{p+1}$$

Because the proof of this lemma closely resembles the proof of Theorem 1 in Zhang, Wang, and Zhu [7], we omit the details. It is important to note that when $m = \mathcal{M}$, $\hat{\theta}_{CLS}(\mathcal{M})$ is the solution to the estimating equation $\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}) = 0$, and in this case, $H(\hat{\theta}_{CLS}(\mathcal{M})) = 0$. Furthermore, Lemma 2 suggests that $H(\widetilde{\theta}_0) = O_p(1)$.

Definition 1. *We propose the following penalized criteria:*

$$H(\hat{\theta}_{CLS}(m)) + P_T \cdot |m| \tag{6}$$

where the penalty term P_T *is an increasing sequence,* $P_T \to \infty$ *and satisfies* $P_T = O\left(T^{\frac{1}{2}}\right)$ *and* $\frac{log(T)}{P_T} = O(1)$.

Remark 2. *Intuitively, in this penalized criterion,* $H(\hat{\theta}_{CLS}(m))$ *serves as a measure of the model's fit to the data. If it can be demonstrated that the divergence rate of* $H(\hat{\theta}_{CLS}(m))$ *is slower when* $m_0 \subseteq m$ *compared to the divergence rate of* $H(\hat{\theta}_{CLS}(m_1))$ *when* $m_0 \not\subseteq m_1$, *then a smaller* $H(\hat{\theta}_{CLS}(m))$ *would suggest a superior fit of model m to the data. However, upon closer examination, it becomes evident that if we merely adopt model* $\mathcal{M}$, *then* $H(\hat{\theta}_{CLS}(\mathcal{M})) = 0$ *. Consequently, it is necessary to introduce a penalty term,* $P_T \cdot |m|$, *to constrain the number of lagged variables incorporated by model m. By striking a balance between the degree of data fitting* $H(\hat{\theta}_{CLS}(m))$ *and the number of lagged variables* $P_T \cdot |m|$, *Theorems 1–3 substantiate the ability to select the appropriate model.*

Under the correct model specification, the following theorem can be derived:

Theorem 1. *Given assumptions (A1) and (A2), under the correct model specification*

$$\hat{\theta}_{(1),CLS}(m) - \widetilde{\theta}_{(1)}(m_0) = -\left(\mathbb{E}\left(\frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'_{(1)}(m)}\right)\right)^{-1} \frac{1}{T} \sum_{t=p+1}^{T} \Psi_{1t}(\widetilde{\theta}_0) + o_p\left(T^{-\frac{1}{2}}\right)$$

and $H(\hat{\theta}_{CLS}(m))$ *converges in probability to* $\sum_{j=1}^{p+1} \Lambda_j \chi^2(1)$, *where* Λ_j *is the eigenvalue of the matrix* $\Sigma_{11}^{\frac{1}{2}} \Sigma'_* \Sigma_{11}^{-1} \Sigma_* \Sigma_{11}^{\frac{1}{2}}$, *where*

$$\Sigma_* = \left(I - \mathbb{E}\left(\frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'}\right) \left[\begin{pmatrix} \left(\mathbb{E}\left(\frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'_{(1)}(m)}\right)\right)^{-1} & 0 \\ 0 & 0 \end{pmatrix}\right]\right)$$

$$\Sigma_{11} = \mathbb{E}\left(\Psi_t(\widetilde{\theta}_0) \Psi_t(\widetilde{\theta}_0)'\right)$$

Theorem 1 establishes the asymptotic distribution of $\hat{\theta}_{(1),CLS}(m)$ and $H(\hat{\theta}_{CLS}(m))$ under the correct model specification, which serves as a crucial component in the derivation for the consistency of our penalized criteria (6). In the following, we discuss the performance of $H(\hat{\theta}_{CLS}(m))$ when the model specification m is incorrect.

Theorem 2. *Given assumptions (A1)–(A3), for any $\widetilde{\theta}_1$ in the neighborhood of $\widetilde{\theta}^* \neq \widetilde{\theta}_0$, we have:*

$$T^{-\frac{1}{2}} H\left(\widetilde{\theta}_1\right) \to \infty$$

Theorem 2 and assumption (A3) ensure that if the model m is misspecified, $H\left(\hat{\theta}_{CLS}(m)\right)$ will diverge to positive infinity at a rate of at least $T^{\frac{1}{2}}$. Combining Theorems 1 and 2, we can present the primary conclusion of this paper. When the model is specified as m, we have the following theorem.

Theorem 3. *Given assumptions (A1)–(A3), we have:*

$$\mathbb{P}\left(\min\left(H\left(\hat{\theta}_{CLS}(m)\right) + P_T \cdot |m| : m \neq m_o\right) > H\left(\hat{\theta}_{CLS}(m_0)\right) + P_T \cdot |m_0|\right) \to 1$$

From the proof of Theorem 3, and Lemma A.1, we can observe that the divergence rate of P_T needs to be at least as fast as $log(T)$. In practical applications, we may use settings such as $P_T = T^{\frac{1}{5}}$. In such settings, although $\frac{log(T)}{P_T} \to 0$, in finite samples, $P_T < log(T)$. In fact, in the interval $[4, 332, 106]$, $T^{\frac{1}{5}} < log(T)$, which may result in the performance of $P_T = T^{\frac{1}{5}}$ not being as effective as $P_T = log(T)$ in finite samples. Nevertheless, such penalty term settings still hold value, and we will discuss this situation in the numerical simulation section.

Theorem 3 provides the consistency of the penalized criteria (6) for model selection. It becomes evident that Theorem 3 holds under very relaxed assumptions and relies solely on the CLS estimation, which can be rapidly completed in any statistical software, and the estimating equation constructed by first-order conditional moments, which is easy to derive. This makes the penalized criteria (6) highly suitable for use in INAR models, particularly in RCINAR models. Now let $\hat{m}$ be the model selected by the criterion (6):

$$\hat{m} = argmin_{m \subseteq \mathcal{M}}(H(m) + P_T \cdot |m|)$$

We now present the asymptotic properties of the selected model:

Theorem 4. *Given assumptions (A1)–(A3), we have:*

$$\sqrt{T}(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0) \xrightarrow{d} N(V^{-1}(\widetilde{\theta}_0)W(\widetilde{\theta}_0)V^{-1}(\widetilde{\theta}_0))$$

where:

$$W(\widetilde{\theta}_0) = \mathbb{E}\left(\Psi_t(\widetilde{\theta}_0)\Psi_t(\widetilde{\theta}_0)'\right)$$

$$V(\widetilde{\theta}_0) = \mathbb{E}\left(\frac{\partial \mathbb{E}(Y_t|Y_{t-1})}{\partial \widetilde{\theta}} \cdot \frac{\partial \mathbb{E}(Y_t|Y_{t-1})}{\partial \widetilde{\theta}'}\right) - \mathbb{E}\left(u_t(\widetilde{\theta}_0)\frac{\partial^2 \mathbb{E}(Y_t|Y_{t-1})}{\partial \widetilde{\theta} \partial \widetilde{\theta}'}\right)$$

$$u_t(\widetilde{\theta}_0) = Y_t - \mathbb{E}\left(Y_t \middle| Y_{(t-1)}\right)$$

Remark 3. *From the inference process in this section, we can see that the estimating equation used in constructing the penalized criteria (6) actually utilizes the information of $\mathbb{E}(Y_t|\mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1} = \sigma(Y_{t-1}, Y_{t-2}, \cdots)$ and does not involve the information of thinning operators. Therefore, the penalized criteria (6) can be applied to models with the same linear form conditional expectations, such as INARCH models and continuous-valued AR models. The likelihood functions of INARCH and AR models can be established with relative ease, enabling us to compare the efficacy of the penalty criteria (6) with that of AIC and BIC across both models.*

3. Numerical Simulations

In this section, we first conduct a simulation study to evaluate the performance of the penalized criteria proposed in this paper for INAR models. Secondly, to compare the proposed penalized criteria with the traditional likelihood-based AIC and BIC, we apply these criteria to INARCH models and AR models. Finally, by utilizing innovation terms of different random distributions, we carry out a simulation study on the robustness of the penalized criteria proposed in this paper.

3.1. Performance of the Penalized Criteria in INAR Models

In this subsection, we consider the true data-generating process to be:

$$Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_3^{(t)} \circ Y_{t-3} + Z_t \tag{7}$$

where the mean of $\phi_1^{(t)}$ is 0.4, the mean of $\phi_3^{(t)}$ is 0.2, and $\lambda = 2$, i.e., $\widetilde{\theta}_0 = (0.4, 0, 0.2, 2)'$. By applying the penalized criteria (6), we attempt to select the true model from all RCINAR models up to the third order. In Table 1 below:

i.i.d. represents an *i.i.d. Poisson* random variable sequence,

y_{t-1} represents the model $Y_t = \phi_1^{(t)} \circ Y_{t-1} + Z_t$,

y_{t-2} represents the model $Y_t = \phi_2^{(t)} \circ Y_{t-2} + Z_t$,

y_{t-3} represents the model $Y_t = \phi_3^{(t)} \circ Y_{t-3} + Z_t$,

y_{t-1}, y_{t-2} represents the model $Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_2^{(t)} \circ Y_{t-2} + Z_t$,

y_{t-1}, y_{t-3} represents the model $Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_3^{(t)} \circ Y_{t-3} + Z_t$,

y_{t-1}, y_{t-3} represents the model $Y_t = \phi_2^{(t)} \circ Y_{t-2} + \phi_3^{(t)} \circ Y_{t-3} + Z_t$,

$y_{t-1}, y_{t-2}, y_{t-3}$ represents the model $Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_2^{(t)} \circ Y_{t-2} + \phi_3^{(t)} \circ Y_{t-3} + Z_t$.

In addition, "Coef" denotes the random distribution of the coefficient. In this subsection, we focus on the performance of penalized criteria in INAR models. We use boldface to highlight the true model, i.e., y_{t-1}, y_{t-3}. We compare three different penalty term settings $P_T = log(T)$, $P_T = T^{\frac{1}{3}}$, and $P_T = T^{\frac{1}{5}}$ and consider three different distributions for $\phi_1^{(t)}$ and $\phi_3^{(t)}$:

(i)   Fixed coefficients, i.e., $\phi_1^{(t)} = 0.4$, $\phi_3^{(t)} = 0.2$, regardless of t;

(ii) $\phi_1^{(t)}$ follows a uniform distribution on the interval $[0, 0.8]$, $\phi_3^{(t)}$ follows a uniform distribution on the interval $[0, 0.4]$;

(iii) $\phi_1^{(t)}$ follows a beta distribution with a mean of 0.4, $\phi_3^{(t)}$ follows a beta distribution with a mean of 0.2. In this scenario, we fix the parameter vector (a, b) for the beta distribution with $a = 4$ and control the parameter b to achieve different means.

We consider sample sizes $T = 100, 200, 300, 500, 1000$, and for each sample size T and parameter setting, we perform 1000 independent repeated experiments.

As shown in Table 1, for the three penalty terms, the accuracy of model selection using the penalized criteria (6) increases with the sample size T, consistent with the asymptotic conclusion described in Theorem 3. However, when the sample size is large, we find that the accuracy of $P_T = T^{\frac{1}{5}}$ is slightly worse than $P_T = T^{\frac{1}{3}}$ and $P_T = log(T)$. This is because

$$\frac{T^{1/5}}{log(T)} \to \infty$$

However, in the interval $[4, 332, 106]$, $T^{\frac{1}{5}} < log(T)$, which may cause the performance of $P_T = T^{\frac{1}{5}}$ in larger finite samples to be not as good as $P_T = log(T)$. Nonetheless, the penalty term setting $P_T = T^{\frac{1}{5}}$ is not entirely without merit. As shown in Table 1, when the sample size is small, i.e., $T \leq 500$, the performance of $P_T = T^{\frac{1}{5}}$ is better.

Table 1. Frequency of model selection for INAR model of order 2 by the penalized criterion (6).

			$Y_t=\phi_1^{(t)}\circ Y_{t-1}+\phi_3^{(t)}\circ Y_{t-3}+Z_t$
\multicolumn{3}{l	}{$\phi_1=0.4$, $\phi_3=0.2$}	\multicolumn{8}{c	}{**Models to be Selected**}
T	**Coef**	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
100	Fixed	$T^{1/3}$	0.054	0.601	0.003	0.019	0.015	**0.296**	0.002	0.01
		$log(T)$	0.05	0.599	0.003	0.018	0.016	**0.301**	0.002	0.011
		$T^{1/5}$	0.006	0.361	0.001	0.012	0.047	**0.493**	0.001	0.079
	Uniform	$T^{1/3}$	0.059	0.596	0.005	0.037	0.015	**0.275**	0	0.013
		$log(T)$	0.055	0.592	0.005	0.037	0.015	**0.283**	0	0.013
		$T^{1/5}$	0.01	0.366	0.001	0.018	0.049	**0.475**	0.004	0.077
	Beta	$T^{1/3}$	0.072	0.585	0.002	0.029	0.026	**0.281**	0.002	0.003
		$log(T)$	0.069	0.582	0.002	0.029	0.027	**0.285**	0.002	0.004
		$T^{1/5}$	0.013	0.369	0.002	0.017	0.056	**0.479**	0.003	0.061
200	Fixed	$T^{1/3}$	0	0.368	0	0.002	0.016	**0.607**	0	0.007
		$log(T)$	0	0.326	0	0.002	0.017	**0.644**	0	0.011
		$T^{1/5}$	0	0.126	0	0	0.03	**0.781**	0	0.063
	Uniform	$T^{1/3}$	0.001	0.429	0	0.002	0.016	**0.545**	0	0.007
		$log(T)$	0	0.37	0	0.001	0.02	**0.594**	0	0.015
		$T^{1/5}$	0	0.159	0	0	0.032	**0.721**	0	0.088
	Beta	$T^{1/3}$	0.002	0.363	0	0.001	0.021	**0.602**	0	0.011
		$log(T)$	0.002	0.314	0	0	0.025	**0.645**	0	0.014
		$T^{1/5}$	0	0.122	0	0	0.029	**0.768**	0	0.081
300	Fixed	$T^{1/3}$	0	0.183	0	0	0.008	**0.802**	0	0.007
		$log(T)$	0	0.132	0	0	0.007	**0.845**	0	0.016
		$T^{1/5}$	0	0.037	0	0	0.009	**0.88**	0	0.074
	Uniform	$T^{1/3}$	0	0.252	0	0	0.01	**0.725**	0	0.013
		$log(T)$	0	0.176	0	0	0.015	**0.79**	0	0.019
		$T^{1/5}$	0	0.06	0	0	0.02	**0.842**	0	0.078
	Beta	$T^{1/3}$	0	0.218	0	0	0.012	**0.766**	0	0.004
		$log(T)$	0	0.15	0	0	0.016	**0.825**	0	0.009
		$T^{1/5}$	0	0.06	0	0	0.021	**0.859**	0	0.06
500	Fixed	$T^{1/3}$	0	0.04	0	0	0.002	**0.955**	0	0.003
		$log(T)$	0	0.014	0	0	0.003	**0.974**	0	0.009
		$T^{1/5}$	0	0.002	0	0	0.002	**0.95**	0	0.046
	Uniform	$T^{1/3}$	0	0.062	0	0	0.004	**0.932**	0	0.002
		$log(T)$	0	0.03	0	0	0.007	**0.955**	0	0.008
		$T^{1/5}$	0	0.007	0	0	0.006	**0.919**	0	0.068
	Beta	$T^{1/3}$	0	0.046	0	0	0.003	**0.936**	0	0.015
		$log(T)$	0	0.026	0	0	0.003	**0.96**	0	0.011
		$T^{1/5}$	0	0.005	0	0	0.003	**0.932**	0	0.06
1000	Fixed	$T^{1/3}$	0	0	0	0	0	**0.999**	0	0.001
		$log(T)$	0	0	0	0	0	**0.989**	0	0.011
		$T^{1/5}$	0	0	0	0	0	**0.964**	0	0.036

Table 1. *Cont.*

						$Y_t=\phi_1^{(t)}\circ Y_{t-1}+\phi_3^{(t)}\circ Y_{t-3}+Z_t$			
$\phi_1=0.4$, $\phi_3=0.2$						**Models to be Selected**			
T	**Coef**	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
	Uniform	$T^{1/3}$	0	0	0	0	0	**0.997**	0	0.003
		$log(T)$	0	0	0	0	0	**0.99**	0	0.01
		$T^{1/5}$	0	0	0	0	0	**0.952**	0	0.048
	Beta	$T^{1/3}$	0	0	0	0	0	**0.998**	0	0.002
		$log(T)$	0	0	0	0	0	**0.992**	0	0.008
		$T^{1/5}$	0	0	0	0	0	**0.94**	0	0.06

To investigate the performance of the three penalty terms under varying sample sizes and coefficient mean settings, we continue to consider model (7), where $\phi_1^{(t)}$ follows a beta distribution with a mean of 0.4, and $\phi_3^{(t)}$ follows a beta distribution with a mean of ϕ_3. In Figure 1, we report the impact of sample size on the accuracy of the penalized criteria using the three penalty terms under different ϕ_3 settings. In Figures 1 and 2, the red line represents $P_T = T^{\frac{1}{3}}$, the black line represents $P_T = log(T)$, and the blue line represents $P_T = T^{\frac{1}{5}}$, and the vertical axis of both figures represents the frequency of the penalized criteria (6) selecting the correct model. It can be observed that when ϕ_3 is small or the sample size is small, the performance of $P_T = T^{\frac{1}{5}}$ is superior. However, as ϕ_3 gradually moves further from 0 and the sample size increases, the performance of $P_T = T^{\frac{1}{5}}$ becomes slightly worse than $P_T = log(T)$ and $P_T = T^{\frac{1}{3}}$.
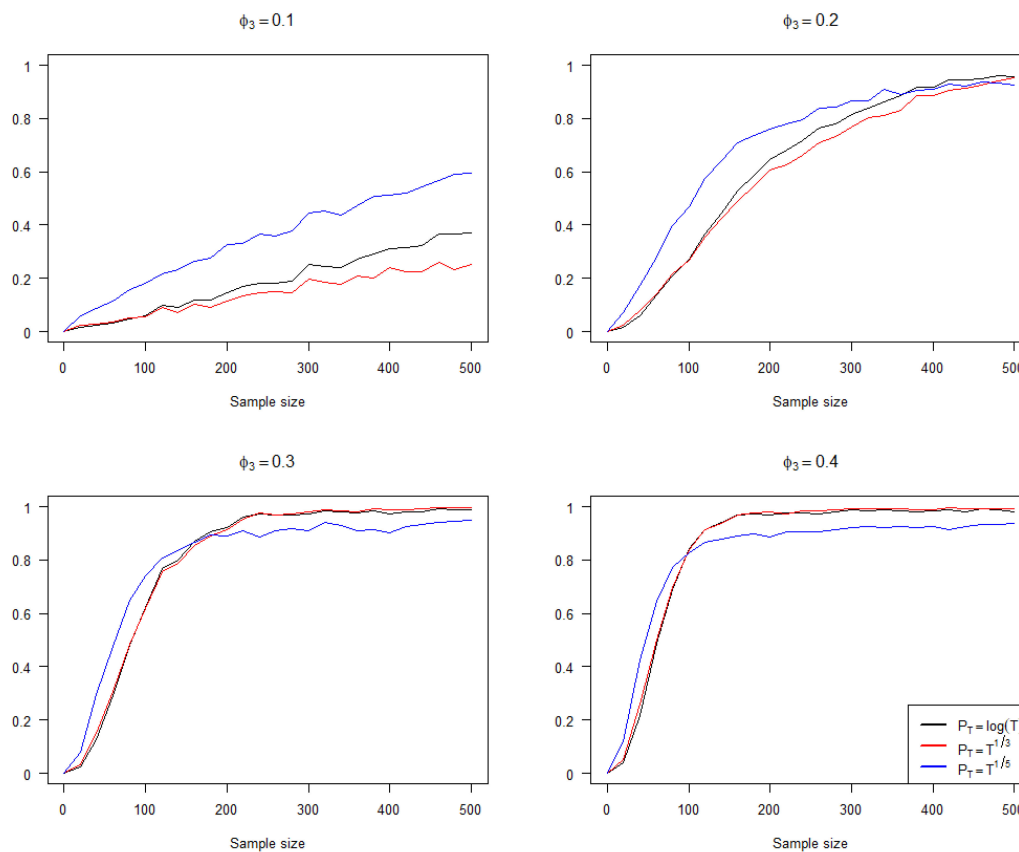
Figure 1. The impact of sample size on accuracy under different ϕ_3 settings.
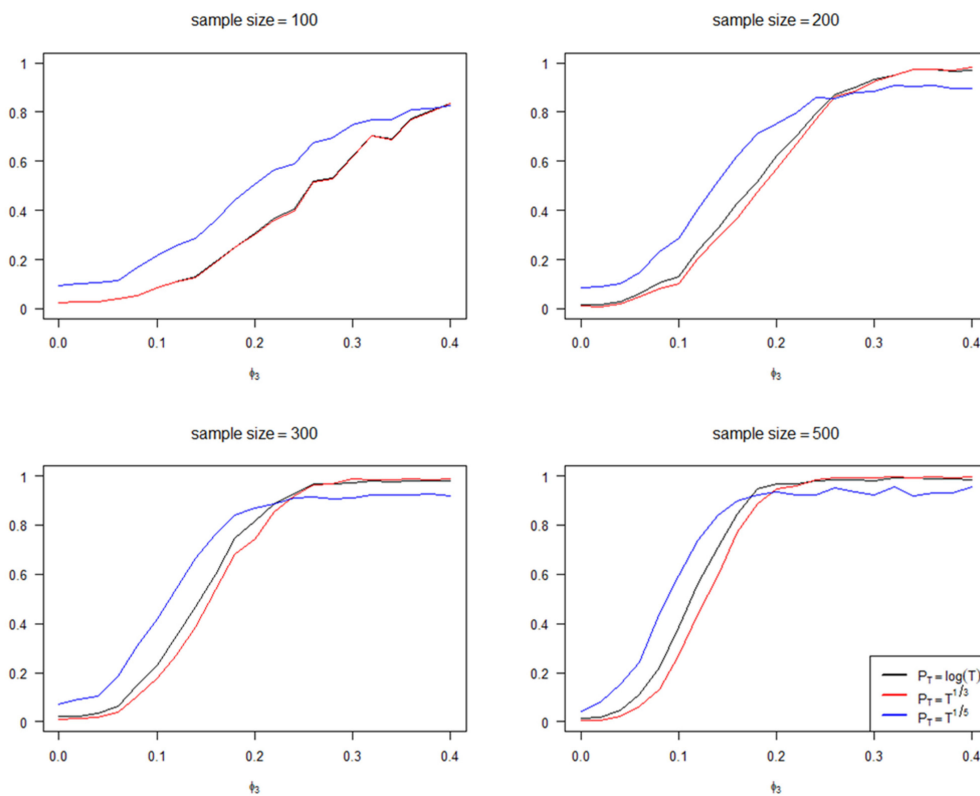
Figure 2. The impact of ϕ_3 settings on accuracy under different sample sizes.

In Figure 2, we report the frequency of selecting the model $Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_3^{(t)} \circ Y_{t-3} + Z_t$ using the penalized criteria (6) as ϕ_3 gradually varies from 0 to 0.4 under different sample size conditions. It should be noted that when $\phi_3 = 0$, $Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_3^{(t)} \circ Y_{t-3} + Z_t$ represents an incorrect model setting and the correct model setting, in this case, should be $Y_t = \phi_1^{(t)} \circ Y_{t-1} + Z_t$. As shown in Figure 2, when the sample size is small, particularly when the sample size is 100, the performance of $P_T = T^{\frac{1}{5}}$ is notably improved compared to $P_T = log(T)$ and $P_T = T^{\frac{1}{3}}$. As the sample size increases, this advantage gradually diminishes, but the penalty term setting $P_T = T^{\frac{1}{5}}$ still maintains an advantage when ϕ_3 is relatively close to 0.

Based on the numerical simulation results presented in this subsection, we can offer recommendations for applying the penalized criteria (6): when the sample size is small, or some coefficients in the true model are relatively close to 0, we can employ the penalty term setting $P_T = T^{\frac{1}{5}}$. In other cases, the performance of the penalty term settings $P_T = T^{\frac{1}{3}}$ and $P_T = log(T)$ is comparable and slightly better than $P_T = T^{\frac{1}{5}}$. Furthermore, we also conducted a simulation study on lag variable selection for the data-generating process:

$$Y_t = \phi_2^{(t)} \circ Y_{t-2} + Z_t$$

where the mean of $\phi_2^{(t)}$ is 0.3. The results can be found in Table A1 in Appendix A.

3.2. Performance of Penalized Criteria in INARCH Models and AR Models

As stated in the Remark of Section 2, we can apply the penalty criteria (6) to both INARCH and AR models. Because the likelihood functions for these two models can be easily established, we can compare the performance of the penalty criteria (6) with that of AIC and BIC for both these models.

3.2.1. INARCH Model

In this subsection, we consider the true data-generating process as follows:

$$Y_t | \mathcal{F}_{t-1} \sim Poisson(\lambda_t)$$

$$\lambda_t = \phi_0 + \phi_1 Y_{t-1} + \phi_3 Y_{t-3}$$

where $\phi_1 = 0.4$, $\phi_3 = 0.2$, and $\phi_0 = 2$. Fokianos, Rahbek, and Tjøstheim [26] proposed this model and derived the conditions for its stationarity and ergodicity. By applying the penalized criteria (6) alongside AIC and BIC, we attempt to select the true model from all INARCH models up to the third order. In Table 2:

i.i.d. represents an *i.i.d. Poisson* random variable sequence,
y_{t-1} represents the model $Y_t | \mathcal{F}_{t-1} \sim Poisson(\lambda_t) \lambda_t = \phi_0 + \phi_1 Y_{t-1}$,
y_{t-2} represents the model $Y_t | \mathcal{F}_{t-1} \sim Poisson(\lambda_t) \lambda_t = \phi_0 + \phi_2 Y_{t-2}$,
y_{t-3} represents the model $Y_t | \mathcal{F}_{t-1} \sim Poisson(\lambda_t) \lambda_t = \phi_0 + \phi_3 Y_{t-3}$,
y_{t-1}, y_{t-2} represents the model $Y_t | \mathcal{F}_{t-1} \sim Poisson(\lambda_t) \lambda_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2}$,
y_{t-1}, y_{t-3} represents the model $Y_t | \mathcal{F}_{t-1} \sim Poisson(\lambda_t) \lambda_t = \phi_0 + \phi_1 Y_{t-1} + \phi_3 Y_{t-3}$,
y_{t-1}, y_{t-3} represents the model $Y_t | \mathcal{F}_{t-1} \sim Poisson(\lambda_t) \lambda_t = \phi_0 + \phi_2 Y_{t-2} + \phi_3 Y_{t-3}$,
$y_{t-1}, y_{t-2}, y_{t-3}$ represents the model $Y_t | \mathcal{F}_{t-1} \sim Poisson(\lambda_t)$,

$$\lambda_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3}.$$

"Criterion" denotes the model selection criteria we use, and we use $H + P_T \cdot |m|$ to denote penalized criteria (6). Furthermore, we have bolded the true model y_{t-1}, y_{t-3}. We consider sample sizes $T = 100, 200, 300, 500, 1000$, and for each sample size T and parameter setting, we conduct 1000 independent repeated experiments.

From Table 2, we can observe that, similar to the INAR case, the accuracy of $P_T = T^{\frac{1}{5}}$ is slightly worse than $P_T = T^{\frac{1}{3}}$ and $P_T = log(T)$ in larger sample sizes, but in smaller sample sizes, i.e., $T \leq 500$, the performance of $P_T = T^{\frac{1}{5}}$ is superior. In addition, from Table 2, we can observe that the accuracy of the penalized criteria proposed in this paper is roughly equivalent to BIC when $P_T = T^{\frac{1}{3}}$ and $P_T = log(T)$, while the accuracy is roughly equivalent to AIC in small samples when $P_T = T^{\frac{1}{5}}$, but $P_T = T^{\frac{1}{5}}$ is far better than AIC when the sample size is large.

Table 2. Frequency of model selection for INARCH model of order 2 by the penalized criterion (6).

			$Y_t \| \mathcal{F}_{t-1} \sim Poisson(\lambda_t)$							
			$\lambda_t = \phi_0 + \phi_1 Y_{t-1} + \phi_3 Y_{t-3}$							
	$\phi_1 = 0.4, \phi_3 = 0.2$		**Models to Be Selected**							
T	Criterion	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1}, y_{t-2}	y_{t-1}, y_{t-3}	y_{t-2}, y_{t-3}	$y_{t-1}, y_{t-2}, y_{t-3}$
100	$H + P_T \cdot \|m\|$	$T^{1/3}$	0.059	0.576	0.001	0.019	0.021	**0.309**	0.001	0.014
		$log(T)$	0.057	0.568	0.001	0.017	0.021	**0.32**	0.001	0.015
		$T^{1/5}$	0.012	0.337	0	0.007	0.054	**0.509**	0.003	0.078
	AIC		0.003	0.27	0	0.006	0.054	**0.559**	0.001	0.107
	BIC		0.032	0.551	0.002	0.021	0.022	**0.361**	0.001	0.01
200	$H + P_T \cdot \|m\|$	$T^{1/3}$	0	0.406	0	0.002	0.013	**0.574**	0	0.005
		$log(T)$	0	0.359	0	0.001	0.017	**0.612**	0	0.011
		$T^{1/5}$	0	0.138	0	0	0.028	**0.756**	0	0.078
	AIC		0	0.068	0	0	0.024	**0.774**	0	0.134
	BIC		0	0.296	0	0.001	0.019	**0.673**	0	0.011

Table 2. *Cont.*

			$Y_t \mid \mathcal{F}_{t-1} \sim Poisson(\lambda_t)$							
			$\lambda_t = \phi_0 + \phi_1 Y_{t-1} + \phi_3 Y_{t-3}$							
$\phi_1 = 0.4, \phi_3 = 0.2$				Models to Be Selected						
T	Criterion	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1}, y_{t-2}	y_{t-1}, y_{t-3}	y_{t-2}, y_{t-3}	$y_{t-1}, y_{t-2}, y_{t-3}$
300	$H + P_T \cdot \lvert m \rvert$	$T^{1/3}$	0	0.22	0	0	0.07	**0.767**	0	0.006
		$log(T)$	0	0.153	0	0	0.011	**0.826**	0	0.01
		$T^{1/5}$	0	0.041	0	0	0.01	**0.874**	0	0.075
	AIC		0	0.016	0	0	0.006	**0.832**	0	0.146
	BIC		0	0.127	0	0	0.008	**0.855**	0	0.01
500	$H + P_T \cdot \lvert m \rvert$	$T^{1/3}$	0	0.035	0	0	0.03	**0.958**	0	0.004
		$log(T)$	0	0.017	0	0	0.002	**0.971**	0	0.01
		$T^{1/5}$	0	0.001	0	0	0.002	**0.934**	0	0.063
	AIC		0	0	0	0	0.001	**0.841**	0	0.158
	BIC		0	0.012	0	0	0.004	**0.976**	0	0.008
1000	$H + P_T \cdot \lvert m \rvert$	$T^{1/3}$	0	0	0	0	0	**1**	0	0
		$log(T)$	0	0	0	0	0	**0.991**	0	0.009
		$T^{1/5}$	0	0	0	0	0	**0.956**	0	0.044
	AIC		0	0	0	0	0	**0.848**	0	0.152
	BIC		0	0	0	0	0	**0.995**	0	0.005

Additionally, we provide a simulation study on lag variable selection for the data-generating process:

$$Y_t \mid \mathcal{F}_{t-1} \sim Poisson(\lambda_t)$$

$$\lambda_t = \phi_0 + \phi_1 Y_{t-1}.$$

The results can be found in Table A2 in the Appendix A.

3.2.2. AR Model

In this subsection, we consider the true data-generating process as follows:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_3 Y_{t-3} + Z_t \tag{8}$$

where $\phi_1 = 0.4$, $\phi_3 = 0.2$, $\phi_0 = 1$, and Z_t follows a normal distribution with a mean of 0 and a standard deviation of 2. By applying the penalized criteria (6) alongside AIC and BIC, we attempt to select the true model from all AR models up to the third order. In Table 3:

i.i.d. represents an i.i.d. Normal random variable sequence,
y_{t-1} represents the model $Y_t = \phi_0 + \phi_1 Y_{t-1} + Z_t$,
y_{t-2} represents the model $Y_t = \phi_0 + \phi_2 Y_{t-2} + Z_t$,
y_{t-3} represents the model $Y_t = \phi_0 + \phi_3 Y_{t-3} + Z_t$,
y_{t-1}, y_{t-2} represents the model $Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t$,
y_{t-1}, y_{t-3} represents the model $Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_3 Y_{t-3} + Z_t$,
y_{t-1}, y_{t-3} represents the model $Y_t = \phi_0 + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + Z_t$,
$y_{t-1}, y_{t-2}, y_{t-3}$ represents the model $Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + Z_t$,

"Criterion" denotes the model selection criteria we use, and we use $H + P_T \cdot \lvert m \rvert$ to denote penalized criteria (6). We use boldface to highlight the true model:

Table 3. Frequency of model selection AR model of order 1 by the penalized criterion (6).

			$Y_t=\phi_0+\phi_1 Y_{t-1}+\phi_3 Y_{t-3}+Z_t$							
ϕ_1=0.4, ϕ_3=0.2			**Models to Be Selected**							
T	Criterion	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
100	$H+P_T\cdot\lvert m\rvert$	$T^{1/3}$	0.048	0.577	0.002	0.02	0.018	**0.324**	0.001	0.01
		$log(T)$	0.048	0.564	0.002	0.02	0.018	**0.336**	0.001	0.011
		$T^{1/5}$	0.009	0.34	0.002	0.012	0.044	**0.524**	0.001	0.068
	AIC		0.004	0.255	0.002	0.005	0.059	**0.578**	0.002	0.095
	BIC		0.032	0.564	0.001	0.016	0.018	**0.352**	0.001	0.016
200	$H+P_T\cdot\lvert m\rvert$	$T^{1/3}$	0.001	0.323	0	0.001	0.011	**0.654**	0	0.01
		$log(T)$	0.001	0.279	0	0	0.014	**0.693**	0	0.013
		$T^{1/5}$	0	0.11	0	0	0.025	**0.79**	0	0.075
	AIC		0	0.062	0	0	0.021	**0.771**	0	0.146
	BIC		0.001	0.261	0	0	0.013	**0.713**	0	0.012
300	$H+P_T\cdot\lvert m\rvert$	$T^{1/3}$	0	0.167	0	0	0.004	**0.825**	0	0.004
		$log(T)$	0	0.116	0	0	0.004	**0.874**	0	0.006
		$T^{1/5}$	0	0.042	0	0	0.007	**0.893**	0	0.058
	AIC		0	0.017	0	0	0.007	**0.824**	0	0.152
	BIC		0	0.107	0	0	0.005	**0.88**	0	0.008
500	$H+P_T\cdot\lvert m\rvert$	$T^{1/3}$	0	0.034	0	0	0.003	**0.959**	0	0.004
		$log(T)$	0	0.013	0	0	0.004	**0.975**	0	0.008
		$T^{1/5}$	0	0.002	0	0	0.001	**0.937**	0	0.06
	AIC		0	0	0	0	0	**0.837**	0	0.163
	BIC		0	0.011	0	0	0.003	**0.977**	0	0.009
1000	$H+P_T\cdot\lvert m\rvert$	$T^{1/3}$	0	0	0	0	0	**0.996**	0	0.004
		$log(T)$	0	0	0	0	0	**0.989**	0	0.011
		$T^{1/5}$	0	0	0	0	0	**0.951**	0	0.046
	AIC		0	0	0	0	0	**0.846**	0	0.154
	BIC		0	0	0	0	0	**0.988**	0	0.012

From Table 3, we can observe that, similar to the INAR case, the accuracy of $P_T = T^{\frac{1}{5}}$ is slightly worse than $P_T = T^{\frac{1}{3}}$ and $P_T = log(T)$ in larger sample sizes, but in smaller sample sizes, i.e., $T \leq 500$, the performance of $P_T = T^{\frac{1}{5}}$ is superior. The comparison of the penalized criteria proposed in this paper with AIC and BIC in the AR model is analogous to that in the INARCH model; thus further elaboration is not required.

3.3. Robustness of Variable Selection Procedure

In this section, we investigate the robustness of the penalized criteria (6) for different distributions of the innovation term Z_t in model (7). Specifically, we consider Z_t to follow a Poisson distribution, a geometric distribution with a mean of 2, and a uniform distribution over $\{0, 1, 2, 3, 4\}$. In Table 4, "Z_t" denotes the random distribution of the innovation term, whereas "geom" denotes the geometric distribution.

Through Table 4, we observe that the penalized criteria (6) remain robust for various distributions of the innovation term Z_t. This finding suggests that the criteria proposed

in this paper can effectively select the correct lag order even when the innovation term adheres to different distributions. We use boldface to highlight the true model:

Table 4. Frequency of model selection of INAR model by the penalized criterion (6) with Z_t misspecification.

			$Y_t=\phi_1^{(t)}\circ Y_{t-1}+\phi_3^{(t)}\circ Y_{t-3}+Z_t$							
	$\phi_1=0.4, \phi_3=0.2$		**Models to Be Selected**							
T	Z_t	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
100	Poisson	$T^{1/3}$	0.045	0.587	0.003	0.018	0.018	**0.319**	0	0.01
		$log(T)$	0.043	0.584	0.003	0.019	0.018	**0.323**	0	0.01
		$T^{1/5}$	0.005	0.349	0.002	0.008	0.053	**0.514**	0.003	0.066
	Uniform	$T^{1/3}$	0.048	0.564	0.001	0.02	0.018	**0.336**	0.001	0.048
		$log(T)$	0.043	0.559	0	0.019	0.02	**0.345**	0.001	0.043
		$T^{1/5}$	0.014	0.332	0	0.004	0.047	**0.519**	0	0.014
	Geom	$T^{1/3}$	0.07	0.575	0.004	0.032	0.023	**0.285**	0.001	0.02
		$log(T)$	0.067	0.57	0.004	0.032	0.024	**0.292**	0.001	0.02
		$T^{1/5}$	0.009	0.327	0.002	0.011	0.052	**0.511**	0.002	0.086
200	Poisson	$T^{1/3}$	0	0.37	0	0.001	0.008	**0.612**	0	0.008
		$log(T)$	0	0.319	0	0.002	0.012	**0.655**	0	0.012
		$T^{1/5}$	0	0.109	0	0.001	0.02	**0.795**	0	0.075
	Uniform	$T^{1/3}$	0	0.343	0	0	0.012	**0.636**	0	0.009
		$log(T)$	0	0.29	0	0	0.016	**0.681**	0	0.013
		$T^{1/5}$	0	0.13	0	0	0.026	**0.776**	0	0.068
	Geom	$T^{1/3}$	0.005	0.358	0	0	0.018	**0.603**	0	0.016
		$log(T)$	0.004	0.312	0	0	0.02	**0.643**	0	0.021
		$T^{1/5}$	0	0.108	0	0	0.034	**0.752**	0	0.106
300	Poisson	$T^{1/3}$	0	0.193	0	0	0.003	**0.801**	0	0.003
		$log(T)$	0	0.138	0	0	0.004	**0.852**	0	0.006
		$T^{1/5}$	0	0.044	0	0	0.005	**0.878**	0	0.073
	Uniform	$T^{1/3}$	0	0.184	0	0	0.01	**0.802**	0	0.004
		$log(T)$	0	0.122	0	0	0.012	**0.851**	0	0.015
		$T^{1/5}$	0	0.03	0	0	0.011	**0.885**	0	0.074
	Geom	$T^{1/3}$	0	0.188	0	0	0.012	**0.796**	0	0.004
		$log(T)$	0	0.133	0	0	0.015	**0.834**	0	0.018
		$T^{1/5}$	0	0.027	0	0	0.013	**0.88**	0	0.08
500	Poisson	$T^{1/3}$	0	0.027	0	0	0.005	**0.962**	0	0.006
		$log(T)$	0	0.008	0	0	0.005	**0.975**	0	0.012
		$T^{1/5}$	0	0.002	0	0	0.003	**0.923**	0	0.072
	Uniform	$T^{1/3}$	0	0.04	0	0	0.004	**0.95**	0	0.006
		$log(T)$	0	0.02	0	0	0.002	**0.964**	0	0.014
		$T^{1/5}$	0	0.003	0	0	0	**0.926**	0	0.071
	Geom	$T^{1/3}$	0	0.035	0	0	0.008	**0.954**	0	0.003
		$log(T)$	0	0.018	0	0	0.007	**0.964**	0	0.011
		$T^{1/5}$	0	0.002	0	0	0.003	**0.928**	0	0.067

Table 4. *Cont.*

				$Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_3^{(t)} \circ Y_{t-3} + Z_t$						
	$\phi_1 = 0.4, \phi_3 = 0.2$			**Models to Be Selected**						
T	Z_t	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1}, y_{t-2}	y_{t-1}, y_{t-3}	y_{t-2}, y_{t-3}	$y_{t-1}, y_{t-2}, y_{t-3}$
1000	Poisson	$T^{1/3}$	0	0	0	0	0	**0.998**	0	0.002
		$log(T)$	0	0	0	0	0	**0.989**	0	0.011
		$T^{1/5}$	0	0	0	0	0	**0.952**	0	0.048
	Uniform	$T^{1/3}$	0	0	0	0	0	**1**	0	0
		$log(T)$	0	0	0	0	0	**0.995**	0	0.05
		$T^{1/5}$	0	0	0	0	0	**0.944**	0	0.056
	Geom	$T^{1/3}$	0	0	0	0	0	**0.994**	0	0.006
		$log(T)$	0	0	0	0	0	**0.982**	0	0.018
		$T^{1/5}$	0	0	0	0	0	**0.947**	0	0.053

Furthermore, we compare the performance of the penalized criteria proposed in this paper, AIC, and BIC when the innovation term Z_t in AR model (8) follows a uniform distribution over $[-2, 2]$ while the assumption of Z_t is a normal distribution with mean 0 and unknown variance σ_Z^2. In Appendix A, Table A3 shows that regardless of the distribution of the innovation term, when the conditional mean is set correctly, the performance and robustness of the penalized criteria proposed in this paper are generally equivalent to those of AIC and BIC.

4. Real Data Application

4.1. COVID-19 Infection Data

The investigation of data related to infectious diseases constitutes a crucial application of integer-valued time series models within the public health domain. In May 2020, the Ministry of Health in Cyprus disseminated a national epidemic surveillance report, which displayed the temporal data pertaining to the number of infections during the initial phase of the COVID-19 outbreak. Conducting research on this data is instrumental for the public health academia in uncovering the intrinsic mechanisms governing epidemic propagation. Owing to the incubation period associated with the coronavirus, individuals who contract the virus typically disclose their infection status to governmental statistical departments after a lapse of several days. As a result, it becomes imperative to scrutinize the matter of lag variable selection within this time series dataset, see Figure 3 below.
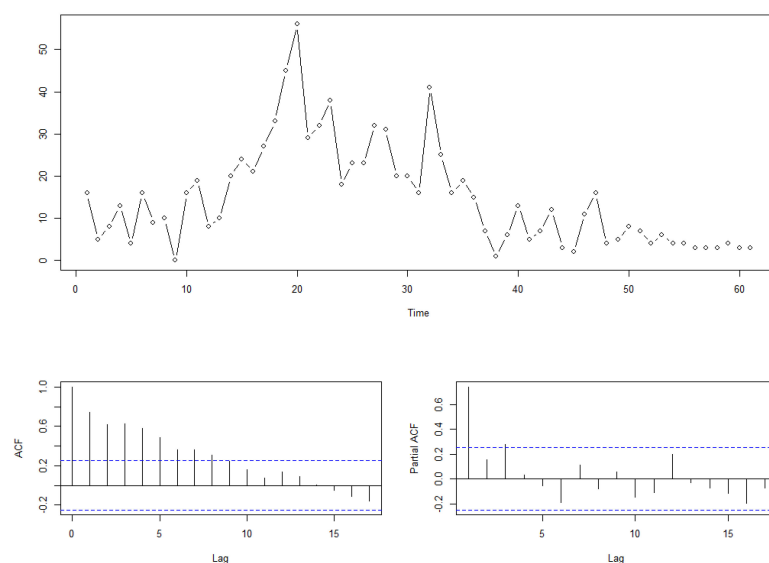

Figure 3. Number of COVID-19 infections in Cyprus, 13 March to 12 May 2020.
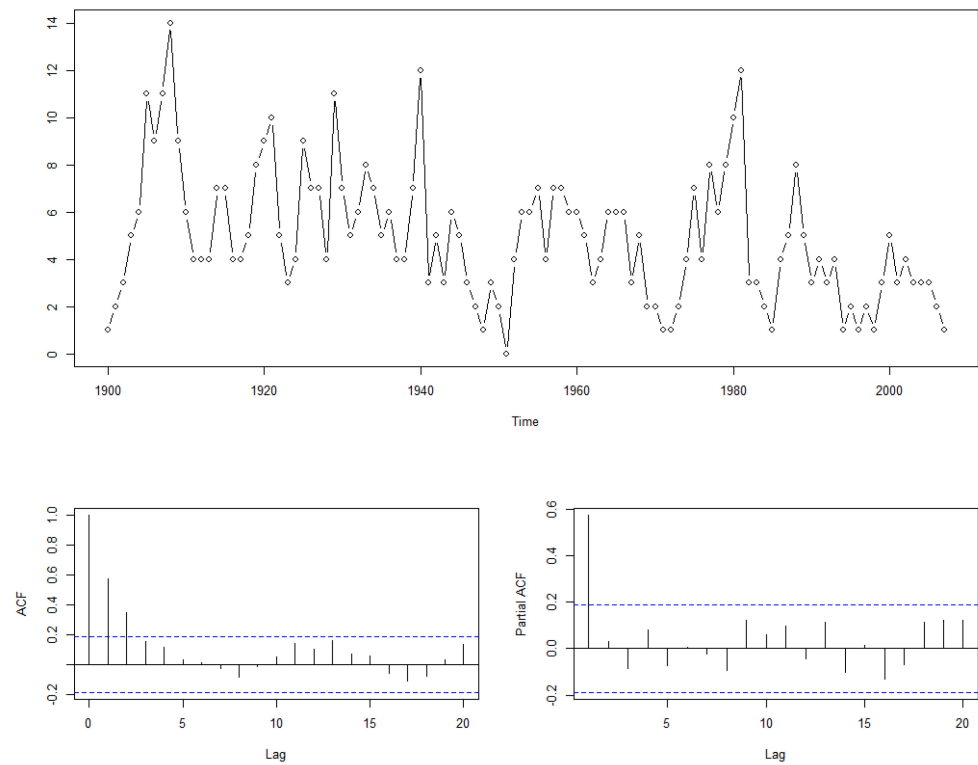
Based on the ACF plot, it can be inferred that the data may stem from an autoregressive data-generating process. The PACF plot suggests that selecting either the model:

$$Y_t = \phi_1^{(t)} \circ Y_{t-1} + Z_t$$

or

$$Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_3^{(t)} \circ Y_{t-3} + Z_t$$

is reasonable, as the partial autocorrelation function for a lag of three periods does not significantly exceed the critical value. Consequently, we employ the model selection procedure (6) for variable selection, and we provide the result in the following Table 5.

Table 5. Result of model selection with COVID-19 data.

P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
$T^{1/3}$	28.6	15.9	22.9	20.5	17.9	**11.8**	22.9	15.7
$log(T)$	28.7	16.3	23.3	20.8	18.4	**12.3**	23.4	16.4
$T^{1/5}$	26.9	12.6	19.6	17.1	12.9	**6.8**	17.9	9.1

Given that the penalized criteria (6) favor the model

$$Y_t = \phi_1^{(t)} \circ Y_{t-1} + \phi_3^{(t)} \circ Y_{t-3} + Z_t$$

under all three penalty settings, we adopt this model. The estimated results for this model are:

$$Y_t = \begin{array}{cc} 0.5736 \circ Y_{t-1} & + \quad 0.2933 \circ Y_{t-3} + Z_t \\ (0.1081) & (0.1092) \end{array}$$

where the mean of Z_t is 1.8567; $\phi_1^{(t)}$ and $\phi_3^{(t)}$ as two non-negative random variables, have expected values of 0.5736 and 0.2933, respectively. This finding suggests that during the initial stages of the outbreak in Cyprus, the number of infections on a given day may have been influenced by the number of infections one day and three days prior.

4.2. Seismic Frequency Data

The exploration of earthquake frequency constitutes a significant application frontier for integer-valued time series models. As documented by Zucchini, MacDonald, and Langrock [27], comprehensive annual data delineating global seismic occurrences of magnitude seven or above, encompassing the period from 1900 to 2007, has been provided. This wealth of data offers a promising platform for scholars seeking to unravel the intricate mechanisms underpinning the mutual interactions among earthquakes. It is envisaged that through a meticulous investigation of this time-series data associated with seismic activities, one might gain insights into whether the interplay is mediated by crustal stress dynamics or alternative conduits, see Figure 4 below.

Informed by the ACF, we hypothesize that the underlying data generation process might be suitably modeled by an autoregressive construct. On the other hand, insights gleaned from the PACF advocate for the application of a first-order autoregressive model. To substantiate this conjecture further, we will proceed to invoke the penalized criterion (6) as our analytical tool in the ensuing discourse, and we provide the result in the following Table 6.

Figure 4. Global frequency of earthquakes of magnitude seven or greater between 1900 and 2007.

Table 6. Result of model selection with seismic frequency data.

P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1}, y_{t-2}	y_{t-1}, y_{t-3}	y_{t-2}, y_{t-3}	$y_{t-1}, y_{t-2}, y_{t-3}$
$T^{1/3}$	29.8	**10.7**	29.5	35.9	15.3	15.1	33.1	19.0
$log(T)$	29.7	**10.6**	29.3	35.8	15.0	12.9	32.8	18.7
$T^{1/5}$	27.5	**6.3**	25.1	31.6	8.6	8.5	26.5	10.2

Given that under the three penalty settings, the penalized criterion (6) exhibits a preference for the model

$$Y_t = \phi_1^{(t)} \circ Y_{t-1} + Z_t$$

we opt to adopt this model. The estimated results for this model are as follows:

$$Y_t = 0.5799 \circ Y_{t-1} + Z_t(0.0812)$$

In this model, the mean value of Z_t is identified as 2.1014. The derived estimations posit that every occurrence of a magnitude seven or higher earthquake in the preceding year induces a count of similar-intensity earthquakes in the subsequent year, which manifests as a discrete random variable with an expected value of 0.5799. Simultaneously, the number of major earthquakes occurring independently each year is approximately two. These results substantiate the existence of a year-on-year time-varying dependency mechanism in the frequency of major seismic disasters.

5. Discussion and Conclusions

In this paper, we propose a model selection criterion based on an estimation equation established in Conditional Least Squares estimation. This penalized method does not rely on detailed distributional assumptions for the data-generating process. It circumvents the complex likelihood function construction in Random Coefficient Integer-Valued

Autoregressive models and can consistently select the correct variables under relatively mild assumptions.

In our numerical simulations, we compared the impact of three penalty term settings on the performance of the penalty criteria. We found that the impact of these penalty terms on the performance of the information criteria varies as partial coefficients in the RCINAR model move farther away from 0 or as the sample size increases. Moreover, we applied the model selection method proposed in this paper to both the INARCH and traditional continuous-valued AR models. We discovered that in both scenarios where likelihood functions can be easily constructed, the proposed model selection criteria and the traditional likelihood-based information criteria, AIC and BIC, exhibit similar model selection efficiency. Specifically, under the settings of $P_T = T^{\frac{1}{3}}$ and $P_T = log(T)$, the accuracy of the proposed model selection method is similar to that of BIC. However, in cases with smaller sample sizes, the proposed model selection method with $P_T = T^{\frac{1}{5}}$ performs similarly to AIC while outperforming AIC with larger sample sizes.

In the future, model selection methods based on estimation equations have considerable potential for development. In this discussion section, we briefly introduce three aspects:

(1) Distinguishing between different thinning operators or innovation terms with varying distributions: The criterion (6) provided in this paper is primarily used for lag variable selection but lacks the ability to differentiate between various thinning operators and distinct distributions of innovation terms. It is well known that INAR models can describe scenarios such as zero inflation, variance inflation, and extreme values by flexibly selecting thinning operators and innovation terms. Therefore, if a model selection criterion can distinguish between different thinning operators and varying distributions of innovation terms, it will have a more extensive application scope.

(2) Incorporating higher-order conditional moments from the data-generating process into the information criterion. Through the form of the $H(\widetilde{\theta})$ function:

$$H(\widetilde{\theta}) = \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}) \right)' \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}) \Psi_t(\widetilde{\theta})' \right)^{-1} \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}) \right)$$

It is evident that criterion (6) only contains the mean structure information of the model and lacks the ability to describe higher-order moment information. Since many variants of the INAR model exhibit differences in higher-order moments, incorporating higher-order moment information into the model selection criterion would enable criterion (6) to perform model selection within a broader context.

(3) Detecting change points. In the field of time series data research, the change point detection problem has a long history. Specifically, within the integer-valued time series domain, the change point problem refers to the existence of positive integers $\tau_1, \tau_2, \cdots, \tau_m$, such that:

$$Y_t = \begin{cases} \phi^{(1)} \circ Y_{\{t-1\}} + Z_t^{(1)} & 0 < t \leq \tau_1 \\ \phi^{(2)} \circ Y_{\{t-1\}} + Z_t^{(2)} & \tau_1 < t \leq \tau_2 \\ \quad\vdots & \quad\vdots \\ \phi^{(m)} \circ Y_{\{t-1\}} + Z_t^{(m)} & \tau_m < t \leq T \end{cases}$$

For continuous-valued time series models, Chen and Gupta [28] introduced a method for change point detection using AIC and BIC. Since parameter changes are prominently reflected in the mean structure of INAR models, it is likely feasible to perform change point detection using the criterion (6) based on the estimation equations.

Appendix A

Appendix A.1 Proofs

Proof of Lemma 1.

(1) From Lemma 1 in Zhang, Wang, and Zhu [7], it can be proved immediately.

(2) Due to the construction of $\Psi_t(\theta)$, we have $\frac{\partial^2 \Psi_t(\theta)}{\partial \theta \partial \theta'} = 0$, which proves the statement.

(3) Because the construction of $\Psi_t(\theta)$ ensures that $\frac{\partial \Psi_t(\theta)}{\partial \theta'}$ is a constant with respect to the parameter vector θ and is only related to $\{Y_t\}$, the conclusion holds under Assumption (A1). $\square$

Proof of Theorem 1. When model m is correctly specified, by Taylor expansion, we have:

$$\sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m)) = \sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_0) + \sum_{t=p+1}^{T} \frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'} \left(\hat{\theta}_{CLS}(m) - \widetilde{\theta}_0\right) + o_p\left(T^{-\frac{1}{2}}\right)$$

where, according to Equation (5), $o_p\left(\left\|\hat{\theta}_{CLS}(m) - \widetilde{\theta}_0\right\|\right) = o_p\left(T^{-\frac{1}{2}}\right)$.

Since $\hat{\theta}_{(2),CLS}(m) = \widetilde{\theta}_{(2)}(m) = 0$, then:

$$0 = \frac{1}{T} \sum_{t=p+1}^{T} \Psi_{1t}(\hat{\theta}_{CLS}(m))$$

$$= \frac{1}{T} \sum_{t=p+1}^{T} \Psi_{1t}(\widetilde{\theta}_0) + \frac{1}{T} \sum_{t=p+1}^{T} \frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \theta'_{(1)}(m)} \left(\hat{\theta}_{(1),CLS}(m) - \widetilde{\theta}_{(1)}(m_0)\right) + o_p\left(T^{-\frac{1}{2}}\right)$$

Thus, we have:

$$\hat{\theta}_{(1),CLS}(m) - \widetilde{\theta}_{(1)}(m_0) = -\left(\frac{1}{T} \sum_{t=p+1}^{T} \frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}_{(1)}(m)}\right)^{-1} \frac{1}{T} \sum_{t=p+1}^{T} \Psi_{1t}(\widetilde{\theta}_0) + o_p\left(T^{-\frac{1}{2}}\right)$$

$$= -\left(\mathbb{E}\left(\frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'_{(1)}(m)}\right)\right)^{-1} \frac{1}{T} \sum_{t=p+1}^{T} \Psi_{1t}(\widetilde{\theta}_0) + o_p\left(T^{-\frac{1}{2}}\right)$$

Therefore:

$$\hat{\theta}_{CLS}(m) - \widetilde{\theta}_0 = \begin{bmatrix} \hat{\theta}_{(1),CLS}(m) - \widetilde{\theta}_{(1)}(m_0) \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\left(\mathbb{E}\left(\frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'_{(1)}(m)} \right) \right)^{-1} \frac{1}{T} \sum_{t=p+1}^{T} \Psi_{1t}(\widetilde{\theta}_0) + o_p\left(T^{-\frac{1}{2}} \right) \\ 0 \end{bmatrix}$$

Hence:

$$\frac{1}{T} \sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m))$$

$$= \left(I - \mathbb{E}\left(\frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'} \right) \left[\begin{matrix} \left(\mathbb{E}\left(\frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'_{(1)}(m)} \right) \right)^{-1} & 0 \\ 0 & 0 \end{matrix} \right] \right) \frac{1}{T} \sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_0) + o_p\left(T^{-\frac{1}{2}} \right)$$

$$\overset{\text{def}}{=} \Sigma_* \frac{1}{T} \sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_0) + op\left(T^{-\frac{1}{2}} \right)$$

where I is the identity matrix, and let $\Sigma_{11} = \mathbb{E}\left(\Psi_t(\widetilde{\theta}_0) \Psi_t(\widetilde{\theta}_0)' \right)$, then by Lemma 1, we have:

$$H(\hat{\theta}_{CLS}(m)) = \left(\sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m)) \right)' \left(\sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m)) \Psi_t(\hat{\theta}_{CLS}(m))' \right)^{-1} \left(\sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m)) \right)$$

$$= \left(T^{-\frac{1}{2}} \sum_{t=p+1}^{T} \Sigma_{11}^{-\frac{1}{2}} \Psi_t(\widetilde{\theta}_0) \right)' \Sigma_{11}^{\frac{1}{2}} \Sigma_*' \Sigma_{11}^{-1} \Sigma_* \Sigma_{11}^{\frac{1}{2}} \left(T^{-\frac{1}{2}} \sum_{t=p+1}^{T} \Sigma_{11}^{-\frac{1}{2}} \Psi_t(\widetilde{\theta}_0) \right) + O_p(1)$$

Let $\Omega = \Sigma_{11}^{\frac{1}{2}} \Sigma_*' \Sigma_{11}^{-1} \Sigma_* \Sigma_{11}^{\frac{1}{2}} = \Sigma_{11}^{\frac{1}{2}} \Sigma_*' \Sigma_{11}^{-\frac{1}{2}} \Sigma_{11}^{-\frac{1}{2}} \Sigma_* \Sigma_{11}^{\frac{1}{2}}$, which implies that Ω is a positive semi-definite matrix. Consequently, according to Johnston [29]'s Theorem 2.1.6, we have $\Omega = U\Lambda U'$, where U is an orthogonal matrix, Λ is a diagonal matrix, and the diagonal elements of Λ are the eigenvalues of Ω, denoted as $\Lambda_1, \ldots, \Lambda_{p+1}$. Thus,

$$H(\hat{\theta}_{CLS}(m)) = \sum_{j=1}^{p+1} \Lambda_j \left[T^{-\frac{1}{2}} \sum_{t=p+1}^{T} \Sigma_{11}^{-\frac{1}{2}} \Psi_t(\widetilde{\theta}_0) \right]_j^2 + O_p(1)$$

where $[\mathcal{J}]_j$ denotes the j-th element of vector $\mathcal{J}$. Therefore, we obtain:

$$H(\hat{\theta}_{CLS}(m)) \overset{d}{\to} \sum_{j=1}^{p+1} \Lambda_j \chi^2(1)$$

Thus, it is known that Λ_j is the eigenvalue of the matrix $\Sigma_{11}^{\frac{1}{2}} \Sigma_*' \Sigma_{11}^{-1} \Sigma_* \Sigma_{11}^{\frac{1}{2}}$, and $\sum_{j=1}^{p+1} \Lambda_j = trace(\Omega) = trace\left(\Sigma_{11}^{\frac{1}{2}} \Sigma_*' \Sigma_{11}^{-1} \Sigma_* \Sigma_{11}^{\frac{1}{2}} \right)$. □

Proof of Theorem 2. Due to assumptions (A1) and (A2), following steps similar to those in Lemma 1 of Zhang, Wang, and Zhu [7], we know that $\mathbb{E}\left(\left\| \Psi_t(\widetilde{\theta}) \right\|^{2+\frac{\delta}{2}} \right)$ is bounded above.

For any $\widetilde{\theta}_1$ in the neighborhood of $\widetilde{\theta}^* \neq \widetilde{\theta}_0$, by applying the Markov inequality, we have:

$$\sum_{i=1}^{\infty} \mathbb{P}\left(\|\Psi_t(\widetilde{\theta}_1)\|^2 > i \right) \leq \sum_{i=1}^{\infty} \frac{\mathbb{E}\left(\|\Psi_t(\widetilde{\theta}_1)\|^{2+\frac{\delta}{2}} \right)}{i^{1+\frac{\delta}{4}}} < \infty$$

By applying the Borel–Cantelli lemma, we can always find a sufficiently large natural number N such that for any $i > N$, $\|\Psi_t(\theta_1)\| \leq i^{-\frac{1}{2}}$ holds with probability 1. This further implies that $\max_{1\leq t\leq T} \|\Psi_t(\widetilde{\theta}_1)\| = o_p\left(T^{\frac{1}{2}}\right)$. Therefore:

$$
\begin{aligned}
H(\widetilde{\theta}_1) &= \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_1) \right)' \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_1)\Psi_t(\widetilde{\theta}_1)' \right)^{-1} \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_1) \right) \\
&\geq \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_1) \right)' \left(\max_t \|\Psi_t(\widetilde{\theta}_1)\| \mathbf{1}' \sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_1)' \right)^{-1} \left(\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_1) \right)
\end{aligned}
$$

where $\mathbf{1} = (1,1,\ldots,1)'$. Due to assumption (A3), $\sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_1) = O_p(T)$, we can deduce that:

$$T^{-\frac{1}{2}} H(\widetilde{\theta}_1) \to \infty$$

□

Lemma A1. $H(m_0) = o_p(log(T))$.

Given:

$$H(m_0) = \left(\sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m_0)) \right)' \left(\sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m_0))\Psi_t(\hat{\theta}_{CLS}(m_0))' \right)^{-1} \left(\sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m_0)) \right)$$

and

$$
\begin{aligned}
\sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m_0)) &= \sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_0) + \left(\sum_{t=p+1}^{T} \frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'_{(1)}(m)} \right)^{-1} \left(\hat{\theta}_{(1),CLS}(m) - \widetilde{\theta}_{(1)}(m_0) \right) + o_p(1) \\
&\overset{\text{def}}{=} Q_t(\widetilde{\theta}_0)
\end{aligned}
$$

$$\frac{1}{T} \sum_{t=p+1}^{T} \Psi_t(\hat{\theta}_{CLS}(m_0))\Psi_t(\hat{\theta}_{CLS}(m_0))' = \Sigma_{11} + o_p(1)$$

We can deduce that:

$$H(m_0) = \frac{1}{T} Q_t(\widetilde{\theta}_0) \Sigma_{11}^{-1} Q_t(\widetilde{\theta}_0) + o_p(1)$$

Let $V = Cov\left(\frac{1}{T} \sum_{t=p+1}^{T} \Psi_t(\widetilde{\theta}_0) + \mathbb{E}\left(\frac{\partial \Psi_t(\widetilde{\theta}_0)}{\partial \widetilde{\theta}'_{(1)}(m)} \right) \left(\hat{\theta}_{(1),CLS}(m) - \widetilde{\theta}_{(1),0}(m) \right) \right)$ Notice that:

$$\mathbb{P}(H(m_0) \geq log(T)) \leq \frac{\mathbb{E}\left(\frac{1}{T} Q_t(\widetilde{\theta}_0) \Sigma_{11}^{-1} Q_t(\widetilde{\theta}_0) \right)}{log(T)} = \frac{trace\left(\Sigma_{11}^{-1} V \right)}{log(T)}$$

Using Lemma A2 below, we can conclude that $H(m_0) = o_p(logT)$.

Lemma A2. $trace\left(\Sigma_{11}^{-1}V\right) = O_p(1).$

Because both Σ_{11}^{-1} and V are semi-positive definite matrices:

$$0 \leq trace\left(\Sigma_{11}^{-1}V\right) \leq trace\left(\Sigma_{11}^{-1}\right)trace(V)$$

As $\Sigma_{11}^{-1} = O_p(1)$. Note that $\left(\hat{\theta}_{(1),CLS}(m) - \widetilde{\theta}_{(1),0}(m)\right) = O_p\left(T^{-\frac{1}{2}}\right)$, and according to Lemma 1, $\frac{1}{T}\sum_{t=p+1}^{T}\Psi_t\left(\widetilde{\theta}_0\right) = O_p(1)$. Therefore, the proof is complete.

Proof of Theorem 3. We divide the proof of this theorem into two parts:

(1) If $m_0 \nsubseteq m$, by applying Theorem 1 and Lemma A1 to m_0, and Theorem 2 to m, we know that for $P_T = O\left(T^{\frac{1}{2}}\right)$ and $\frac{log(T)}{P_T} \to 0$:

$$H\left(\hat{\theta}_{CLS}(m)\right) + P_T \cdot |m| - H\left(\hat{\theta}_{CLS}(m_0)\right) - P_T \cdot |m_0| \to \infty$$

(2) If $m_0 \subseteq m$, then by applying Theorem 1 and Lemma A1 to both m_0 and m, we know that for $P_T \to \infty$:

$$H\left(\hat{\theta}_{CLS}(m)\right) + P_T \cdot |m| - H\left(\hat{\theta}_{CLS}(m_0)\right) - P_T \cdot |m_0|$$
$$= P_T(|m| - |m_0|) + o_p(log(T)) \to \infty$$

Therefore,

$$\mathbb{P}\left(\min\left(H\left(\hat{\theta}_{CLS}(m)\right) + P_T \cdot |m| : m \neq m_o\right) > H\left(\hat{\theta}_{CLS}(m_0)\right) + P_T \cdot |m_0|\right) \to 1$$

$\square$

Proof of Theorem 4. Following the steps in Diop and Kengne [21], we have:
For $x = (x_i)_{1 \leq i \leq p+1}$, $x_i \in \mathbb{R}$, define:

$$F_T(x) = \mathbb{P}\left(\bigcap_{1 \leq i \leq p+1} \sqrt{T}\left(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0\right)_i \leq x_i\right)$$

Then we have:

$$F_T(x) = \mathbb{P}\left(\bigcap_{1 \leq i \leq p+1} \sqrt{T}\left(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0\right)_i \leq x_i | \hat{m} = m_o\right)\mathbb{P}(\hat{m} = m_o)$$
$$+ \mathbb{P}\left(\bigcap_{1 \leq i \leq p+1} \sqrt{T}\left(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0\right)_i \leq x_i | \hat{m} \neq m_o\right)\mathbb{P}(\hat{m} \neq m_o)$$

According to Theorem 3, as $T \to \infty$:

$$\mathbb{P}(\hat{m} = m_0) \to 1 \, , \, \mathbb{P}(\hat{m} \neq m_0) \to 0$$

Therefore:

$$\mathbb{P}\left(\bigcap_{1 \leq i \leq p+1} \sqrt{T}\left(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0\right)_i \leq x_i | \hat{m} \neq m_o\right)\mathbb{P}(\hat{m} \neq m_o) \to 0$$

Hence:

$$F_T(x) = \mathbb{P}\left(\left\{\bigcap_{i \in m_o} \sqrt{T}\left(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0\right)_i \le x_i\right\} \bigcap \left\{\bigcap_{i \notin m_o} \sqrt{T}\left(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0\right)_i \le x_i\right\}\right)$$

Since $\widetilde{\theta}(m_0) \in \widetilde{\Theta}(m_0)$, $\left((\hat{\theta}_{CLS}(m_0))_i\right)_{i \notin m_0} = \left(\widetilde{\theta}_i\right)_{i \notin m_0} = 0$ and $(x_i)_{i \notin m_0}$ is a set of real numbers, i.e., $(x_i)_{i \notin m_0} < \infty$, then by Lemma 1:

$$\mathbb{P}\left(\left\{\bigcap_{i \in m_o} \sqrt{T}(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0)_i \le x_i\right\} \bigcap \left\{\bigcap_{i \notin m_o} \sqrt{T}(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0)_i \le x_i\right\}\right)$$

$$= \mathbb{P}\left(\bigcap_{i \in m_o} \sqrt{T}(\hat{\theta}_{CLS}(\hat{m}) - \widetilde{\theta}_0)_i \le x_i\right) + o_p(1)$$

$$\to \mathbb{P}\left(\left(\Sigma(\widetilde{\theta}_0)\right)^{-\frac{1}{2}} Z \le (x_i)_{i \in m_0}\right)$$

where $\Sigma(\widetilde{\theta}_0) = V^{-1}(\widetilde{\theta}_0) W(\widetilde{\theta}_0) V^{-1}(\widetilde{\theta}_0)$, and Z is a standard normal random vector of dimension $|m_0|$ $\square$

Appendix A.2 Complementary Tables

Table A1. Frequency of model selection for INAR model of order 1 by the penalized criterion (6).

					$Y_t = \phi_2^{(t)} \circ Y_{t-2} + Z_t$					
	$\phi_2 = 0.3$				**Models to Be Selected**					
T	Coef	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1}, y_{t-2}	y_{t-1}, y_{t-3}	y_{t-2}, y_{t-3}	$y_{t-1}, y_{t-2}, y_{t-3}$
100	Fixed	$T^{1/3}$	0.283	0.025	**0.647**	0.006	0.016	0.001	0.022	0
		$log(T)$	0.281	0.025	**0.648**	0.006	0.016	0.001	0.023	0
		$T^{1/5}$	0.08	0.025	**0.691**	0.015	0.088	0.008	0.085	0.008
	Uniform	$T^{1/3}$	0.304	0.033	**0.618**	0.01	0.011	0	0.023	0.001
		$log(T)$	0.297	0.034	**0.623**	0.011	0.011	0	0.023	0.001
		$T^{1/5}$	0.104	0.038	**0.652**	0.015	0.087	0.003	0.085	0.016
	Beta	$T^{1/3}$	0.287	0.029	**0.64**	0.011	0.011	0.001	0.021	0
		$log(T)$	0.281	0.029	**0.645**	0.011	0.011	0.001	0.022	0
		$T^{1/5}$	0.118	0.032	**0.666**	0.01	0.076	0.005	0.081	0.012
200	Fixed	$T^{1/3}$	0.058	0.004	**0.902**	0.001	0.017	0	0.018	0
		$log(T)$	0.045	0.003	**0.902**	0.001	0.021	0	0.024	0.001
		$T^{1/5}$	0.007	0	**0.804**	0.003	0.081	0	0.092	0.013
	Uniform	$T^{1/3}$	0.091	0.006	**0.882**	0.002	0.006	0	0.013	0
		$log(T)$	0.071	0.007	**0.894**	0.002	0.009	0	0.013	0
		$T^{1/5}$	0.017	0.002	**0.809**	0	0.074	0	0.09	0.008
	Beta	$T^{1/3}$	0.059	0.004	**0.912**	0.001	0.011	0	0.012	0.001
		$log(T)$	0.044	0.003	**0.911**	0.001	0.018	0	0.021	0.002
		$T^{1/5}$	0.009	0.001	**0.803**	0.001	0.081	0	0.092	0.013

Table A1. *Cont.*

			$Y_t=\phi_2^{(t)}\circ Y_{t-2}+Z_t$							
ϕ_2=0.3			**Models to Be Selected**							
T	**Coef**	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
300	Fixed	$T^{1/3}$	0.013	0	**0.97**	0.001	0.008	0	0.008	0
		$log(T)$	0.007	0	**0.96**	0.001	0.015	0	0.017	0
		$T^{1/5}$	0.001	0	**0.836**	0	0.076	0	0.077	0.01
	Uniform	$T^{1/3}$	0.017	0	**0.956**	0	0.016	0	0.01	0.001
		$log(T)$	0.009	0	**0.94**	0	0.025	0	0.024	0.002
		$T^{1/5}$	0	0	**0.839**	0	0.081	0	0.264	0.016
	Beta	$T^{1/3}$	0.011	0	**0.967**	0	0.012	0	0.01	0
		$log(T)$	0.008	0	**0.956**	0	0.019	0	0.017	0
		$T^{1/5}$	0	0	**0.857**	0	0.078	0	0.058	0.007
500	Fixed	$T^{1/3}$	0	0	**0.99**	0	0.008	0	0.002	0
		$log(T)$	0	0	**0.977**	0	0.012	0	0.011	0
		$T^{1/5}$	0	0	**0.868**	0	0.054	0	0.066	0.012
	Uniform	$T^{1/3}$	0	0	**0.991**	0	0.003	0	0.006	0
		$log(T)$	0	0	**0.974**	0	0.01	0	0.016	0
		$T^{1/5}$	0	0	**0.874**	0	0.047	0	0.073	0.006
	Beta	$T^{1/3}$	0	0	**0.993**	0	0.004	0	0.003	0
		$log(T)$	0	0	**0.976**	0	0.011	0	0.013	0
		$T^{1/5}$	0	0	**0.879**	0	0.056	0	0.058	0.007
1000	Fixed	$T^{1/3}$	0	0	**0.998**	0	0	0	0.002	0
		$log(T)$	0	0	**0.987**	0	0.006	0	0.007	0
		$T^{1/5}$	0	0	**0.916**	0	0.04	0	0.043	0
	Uniform	$T^{1/3}$	0	0	**0.996**	0	0.001	0	0.003	0
		$log(T)$	0	0	**0.978**	0	0.012	0	0.01	0
		$T^{1/5}$	0	0	**0.914**	0	0.045	0	0.036	0.005
	Beta	$T^{1/3}$	0	0	**0.998**	0	0.001	0	0.001	0
		$log(T)$	0	0	**0.987**	0	0.005	0	0.008	0
		$T^{1/5}$	0	0	**0.901**	0	0.041	0	0.055	0.003

Table A2. Frequency of model selection for INARCH model of order 1 by the penalized criterion (6).

			$Y_t\,|\,\mathcal{F}_{t-1}\sim Poisson(\lambda_t)$							
			$\lambda_t=\phi_0+\phi_1 Y_{t-1}$							
ϕ_1=0.3			**Models to Be Selected**							
T	**Criterion**	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
100	$H+P_T\cdot|m|$	$T^{1/3}$	0.09	**0.849**	0.003	0.02	0.032	0.022	0	0.002
		$log(T)$	0.087	**0.85**	0.003	0.002	0.033	0.023	0	0.002
		$T^{1/5}$	0.024	**0.746**	0.006	0.002	0.109	0.094	0	0.019
	AIC		0.011	**0.713**	0.005	0.001	0.125	0.116	0	0.029
	BIC		0.06	**0.888**	0.002	0.003	0.023	0.023	0	0.001

Table A2. *Cont.*

				$Y_t\,|\,\mathcal{F}_{t-1} \sim Poisson(\lambda_t)$						
				$\lambda_t=\phi_0+\phi_1 Y_{t-1}$						
	$\phi_1=0.3$			**Models to Be Selected**						
T	Criterion	P_T	i.i.d.	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
200	$H+P_T\cdot|m|$	$T^{1/3}$	0.001	**0.958**	0	0	0.017	0.023	0	0.001
		$log(T)$	0.001	**0.947**	0	0	0.022	0.029	0	0.001
		$T^{1/5}$	0	**0.821**	0	0	0.071	0.09	0	0.018
	AIC		0	**0.719**	0	0	0.113	0.129	0	0.039
	BIC		0	**0.959**	0	0	0.019	0.021	0	0.001
300	$H+P_T\cdot|m|$	$T^{1/3}$	0	**0.986**	0	0	0.005	0.009	0	0
		$log(T)$	0	**0.978**	0	0	0.01	0.012	0	0
		$T^{1/5}$	0	**0.854**	0	0	0.065	0.075	0	0
	AIC		0	**0.707**	0	0	0.124	0.148	0	0.021
	BIC		0	**0.977**	0	0	0.009	0.014	0	0
500	$H+P_T\cdot|m|$	$T^{1/3}$	0	**0.994**	0	0	0.004	0.002	0	0
		$log(T)$	0	**0.978**	0	0	0.012	0.01	0	0
		$T^{1/5}$	0	**0.878**	0	0	0.06	0.057	0	0.005
	AIC		0	**0.704**	0	0	0.135	0.126	0	0.035
	BIC		0	**0.973**	0	0	0.017	0.01	0	0
1000	$H+P_T\cdot|m|$	$T^{1/3}$	0	**0.997**	0	0	0.02	0.01	0	0
		$log(T)$	0	**0.979**	0	0	0.008	0.013	0	0
		$T^{1/5}$	0	**0.904**	0	0	0.042	0.05	0	0.004
	AIC		0	**0.701**	0	0	0.13	0.142	0	0.027
	BIC		0	**0.976**	0	0	0.008	0.016	0	0

Table A3. Frequency of model selection for AR model by the penalized criterion (6) with Z_t misspecification.

				$Y_t=\phi_0+\phi_1 Y_{t-1}+\phi_3 Y_{t-3}+Z_t$						
				$Z_t \sim Uniform([-2,2])$						
	$\phi_1=0.4,\ \phi_3=0.2$			**Models to Be Selected**						
T	Criterion	P_T	i.i.d.	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
100	$H+P_T\cdot|m|$	$T^{1/3}$	0.046	0.595	0.001	0.016	0.016	**0.317**	0.001	0.008
		$log(T)$	0.046	0.59	0.001	0.015	0.017	**0.322**	0.001	0.008
		$T^{1/5}$	0.003	0.366	0.001	0.008	0.039	**0.516**	0	0.067
	AIC		0.003	0.269	0.001	0.003	0.059	**0.554**	0	0.111
	BIC		0.026	0.578	0.001	0.012	0.021	**0.349**	0.001	0.012
200	$H+P_T\cdot|m|$	$T^{1/3}$	0	0.356	0	0	0.008	**0.63**	0	0.006
		$log(T)$	0	0.308	0	0	0.012	**0.669**	0	0.011
		$T^{1/5}$	0	0.115	0	0	0.026	**0.794**	0	0.065
	AIC		0	0.069	0	0	0.027	**0.787**	0	0.117
	BIC		0	0.283	0	0	0.014	**0.691**	0	0.012
300	$H+P_T\cdot|m|$	$T^{1/3}$	0	0.157	0	0	0.002	**0.832**	0	0.009
		$log(T)$	0	0.105	0	0	0.005	**0.875**	0	0.015
		$T^{1/5}$	0	0.029	0	0	0.009	**0.901**	0	0.061
	AIC		0	0.01	0	0	0.008	**0.851**	0	0.131
	BIC		0	0.098	0	0	0.006	**0.883**	0	0.013

Table A3. *Cont.*

				$Y_t=\phi_0+\phi_1Y_{t-1}+\phi_3Y_{t-3}+Z_t$						
				$Z_t \sim Uniform([-2,2])$						
	ϕ_1=0.4, ϕ_3=0.2			**Models to Be Selected**						
T	**Criterion**	P_T	*i.i.d.*	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-1},y_{t-2}	y_{t-1},y_{t-3}	y_{t-2},y_{t-3}	y_{t-1},y_{t-2},y_{t-3}
500	$H + P_T\cdot\|m\|$	$T^{1/3}$	0	0.029	0	0	0.01	**0.966**	0	0.004
		$log(T)$	0	0.013	0	0	0.01	**0.975**	0	0.011
		$T^{1/5}$	0	0.002	0	0	0.01	**0.933**	0	0.064
	AIC		0	0	0	0	0.01	**0.841**	0	0.158
	BIC		0	0.011	0	0	0.02	**0.977**	0	0.01
1000	$H + P_T\cdot\|m\|$	$T^{1/3}$	0	0	0	0	0	**0.999**	0	0.001
		$log(T)$	0	0	0	0	0	**0.992**	0	0.008
		$T^{1/5}$	0	0	0	0	0	**0.962**	0	0.038
	AIC		0	0	0	0	0	**0.864**	0	0.136
	BIC		0	0	0	0	0	**0.992**	0	0.008

References

1. Steutel, F.W.; van Harn, K. Discrete analogues of self-decomposability and stability. *Ann. Probab.* **1979**, *7*, 893–899. [CrossRef]
2. Al-Osh, M.A.; Alzaid, A.A. First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [CrossRef]
3. Du, J.G.; Li, Y. The integer-valued autoregressive (INAR(p)) model. *J. Time Ser. Anal.* **1991**, *12*, 129–142.
4. Joe, H. Time series models with univariate margins in the convolution-closed infinitely divisible class. *J. Appl. Probab.* **1996**, *33*, 664–677. [CrossRef]
5. Zheng, H.T.; Basawa, I.V.; Datta, S. First-order random coefficient integer-valued autoregressive processes. *J. Stat. Plan. Inference* **2007**, *137*, 212–229. [CrossRef]
6. Zheng, H.T.; Basawa, I.V.; Datta, S. Inference for pth-order random coefficient integer-valued autoregressive processes. *J. Time Ser. Anal.* **2007**, *23*, 411–440. [CrossRef]
7. Zhang, H.X.; Wang, D.H.; Zhu, F.K. Empirical likelihood inference for random coefficient INAR(p) process. *J. Time Ser. Anal.* **2011**, *32*, 195–203. [CrossRef]
8. Ristić, M.M.; Bakouch, H.S.; Nastić, A.S. A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *J. Stat. Plan. Inference* **2009**, *139*, 2218–2226. [CrossRef]
9. Tian, S.Q.; Wang, D.H.; Shuai, C. A seasonal geometric INAR(1) process based on negative binomial thinning operator. *Stat. Pap.* **2018**, *61*, 2561–2581. [CrossRef]
10. Lu, Y. The predictive distributions of thinning-based count processes. *Scand. J. Stat.* **2019**, *48*, 42–67. [CrossRef]
11. Weiß, C.H. *An Introduction to Discrete-Valued Time Series*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2018.
12. Latour, A. Existence and stochastic structure of a non-negative integer-valued autoregressive process. *J. Time Ser. Anal.* **1998**, *719*, 439–455. [CrossRef]
13. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **1974**, *19*, 716–723. [CrossRef]
14. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
15. Ding, J.; Tarokh, V.; Yang, Y.H. Bridging AIC and BIC: A New Criterion for Autoregression. *IEEE Trans. Inf. Theory* **2018**, *64*, 4024–4043. [CrossRef]
16. Variyath, A.M.; Chen, J.H.; Abraham, B. Empirical likelihood based variable selection. *J. Stat. Plan. Inference* **2010**, *140*, 971–981. [CrossRef]
17. Chen, C.X.; Wang, M.; Wu, R.L.; Li, R.Z. A robust consistent information criterion for model selection based on empirical like-lihood. *Stat. Sin.* **2022**, *32*, 1205–1223.
18. Konishi, S.; Kitagawa, G. *Information Criteria and Statistical Modeling*; Springer: New York, NY, USA, 2007.
19. Ding, J.; Tarokh, V.; Yang, Y.H. Model selection techniques. *IEEE Signal Process. Mag.* **2018**, *11*, 16–34. [CrossRef]
20. Weiß, C.H.; Feld, M.H.-J. On the performance of information criteria for model identification of count time series. *Stud. Nonlinear Dyn. Econ.* **2019**, *24*, 20180012. [CrossRef]
21. Diop, M.L.; Kengne, W. Consistent model selection procedure for general integer-valued time series. *Statistics* **2022**, *55*, 1207–1230. [CrossRef]
22. Wang, X.Y.; Wang, D.H.; Yang, K. Integer-valued time series model order shrinkage and selection via penalized quasi-likelihood approach. *Metrika* **2021**, *84*, 713–750. [CrossRef]
23. Drost, F.C.; Van den Akker, R.; Werker, B.J.M. Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *J. R. Stat. Soc. Ser. B* **2009**, *71*, 467–485. [CrossRef]

24. Kang, J.W.; Lee, S.Y. Parameter Change Test for Random Coefficient Integer-Valued Autoregressive Process with Application to Polio Data Analysis. *J. Time Ser. Anal.* **2009**, *30*, 239–258. [CrossRef]
25. Awale, M.; Balakrishna, N.; Ramanathan, T.V. Testing the constancy of the thinning parameter in a random coefficient integer autoregressive model. *Stat. Pap.* **2019**, *60*, 1515–1539. [CrossRef]
26. Fokianos, K.; Rahbek, A.; Tjøstheim, D. Poisson autoregression. *J. Am. Stat. Assoc.* **2009**, *104*, 1430–1439. [CrossRef]
27. Zucchini, W.; MacDonald, I.L.; Langrock, R. *Hidden Markov Models for Time Series an Introduction Using R*; CRC Press: New York, NY, USA, 2016.
28. Chen, J.; Gupta, A.K. *Parametric Statistical Change Point Analysis*; Birkhäuser: New York, NY, USA, 2013.
29. Johnston, N. *Advanced Linear and Matrix Algebra*; Springer Nature: Cham, Switzerland, 2021.