Regularized Estimation in GINAR(p) Process

Haixiang Zhang[1*], Dehui Wang[2] and Liuquan Sun[3]

[1*]Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

[2]Mathematics School of Jilin University, Changchun 130012, China

[3]Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Abstract. This article is concerned with the regularized estimation methodology for generalized pth-order integer-valued autoregressive (GINAR(p)) process, especially when the regression coefficients are sparse. Under some mild regularity conditions, we show that the regularized estimators perform as well as if the correct submodel was known. The oracle properties of the estimators are established. Extensive Monte Carlo simulation studies demonstrate that the proposed procedure works well. To illustrate its usefulness, an application to a real data about epileptic patient is also provided.

Keywords. Integer-valued time series; Penalty function; Oracle property; Thinning operator; Regularized estimation.

1 Introduction

In recent years, there has been a growing interest in modelling integer-valued time series data, which often occur in many fields, such as the daily number of affected cases in epidemics, the daily number of transaction in stock market and the number of bases of DNA sequences, etc. Due to the particular structure of count time series, the traditional autoregressive process does not work. To model this kind of count data, one possible approach is based on the operator of Steutel and van Harn (1979), which established the foundation of thinning-based method. For example, Al-Osh and Alzaid (1987) proposed the first-order integer-valued autoregressive (INAR(1)) process; Drost, et al. (2009) considered the semiparametric efficient estimation for INAR(p) models; Zhang, et al. (2010) proposed a series of integer-valued autoregressive processes based on signed generalized power series thinning operator; Fokianos (2010) considered the penalized estimation for integer autoregressive models

*Corresponding author: zhx_math@163.com (H. Zhang)

using the ridge penalty; McCabe, et al. (2011) proposed an approach to forecasting count time series data; Pedeli, et al. (2015) used saddlepoint techniques to estimate the parameters of interest in high-order integer-valued autoregressive processes. Recently, Scotto, et al. (2015) reviewed some developments for the thinning-based count time series models.

The motivation of this work is that in some applications, most of the coefficients in the integer-valued autoregressive process are likely to be zero when the order p is large. For example, the daily epileptic seizure counts of an epileptic patient before treatment, which was also studied by Latour (1998). Thus, it is reasonable to describe this kind of data via a count time series model where many regression coefficients are exactly zero. However, some of the traditional estimation methods (e.g. CLS) do not lead to sparse estimators, and it is difficult to identify the structure of GINAR(p) process. The research topic of this article is different from such literatures that focus on the count time series data with plenty of zeros (Barreto-Souza, 2015). We are interested in the integer-valued autoregressive process with sparse dependence structure, and the count time series itself may not be zero-inflated. Also we do not need to require the marginal distribution of the count data in the proposed approach. To the best of our knowledge, very limited research has been done on the sparse estimation of count time series. Our key idea is to estimate the sparse regression coefficients with the help of penalty functions. Note that the simple LASSO penalty (Tibshirani, 1996) leads to biased estimator and does not possess the oracle properties (Fan and Li, 2001), so we only consider the SCAD (Fan and Li, 2001), adaptive LASSO (Zou, 2006), MCP (Zhang, 2010) and SELO (Dicker, et al., 2013) in our approach. Here we point out that Fokianos (2010)'s method does not possess the oracle properties (Fan and Li, 2001) and can not obtain sparse estimators, which result in the main differences with our proposed estimation procedure.

The rest of the paper is organized as follows. In Section 2, we introduce the notation and definition of the GINAR(p) process, also some statistical properties are provided. In Section 3, we describe the regularized estimation procedure for the parameters of interest and establish the oracle properties. Some simulation studies together with an application are given in Section 4. Conclusion remarks are presented in Section 5. All proof details are relegated to the Appendix.

2 Notation and model properties

In the literature, one of the main approaches for modelling count time series is based on the Binomial thinning operator "$\circ$" (Steutel and van Harn, 1979) with $\alpha \circ X = \sum_{i=1}^{X} B_i$, where $\alpha > 0$, X is an non-negative integer-valued random variable, $\{B_i\}$ are independent and identically distributed Bernoulli random variables with success probability $\alpha \in (0, 1)$. Afterwards, Latour (1998) extended it to the generalized thinning operator " $\circ^G$ " and proposed a pth-order generalized integer-valued autoregressive (GINAR(p)) process. First let us recall the definition of this process as follows.

Definition 1. *An integer-valued stochastic process* $\{X \}$ *is said to be a* GINAR(p) *process if it satisfies the following difference equation*

$$X \ = \sum_{i=1}^{p} \alpha_i \circ^G X_{\ -i} + \epsilon \ , \tag{2.1}$$

where
(i) the generalized thinning operator $\circ^G$ *is defined as*

$$\alpha_i \circ^G X_{\ -i} = \sum_{j=1}^{X_{t-i}} W_j^{(i)},$$

given $\alpha_i \geq 0$, $\{W_j^{(i)}\}$ *is some non-negative integer-valued random variable sequence with mean* α_i *and variance* β_i, $i = 1, \cdots, p$;
(ii) $\{\epsilon \}$ *is an i.i.d. integer-valued random variables sequence with mean* μ_ϵ *and variance* σ_ϵ^2;
(iii) the $\{\epsilon \}$ *are uncorrelated with* $X_{\ -i}$ *for* $i \geq 1$ *and all counting series* $\{W_j^{(i)}\}$ *in (2.1) are independent,* ϵ *is independent of* X_0.

The commonly used thinning operators $\circ^G$ include Binomial, Geometric, Poisson, dependent Bernoulli (Ristić, et al., 2013) and ρ-binomial (Borges, et al. 2016). If $0 < \sum_{i=1}^{p} \alpha_i < 1$, then there exists a unique strictly stationary and ergodic X that satisfies (2.1) (Zhang, et al., 2010). Let $\theta_0 = (\mu_\epsilon^0, \alpha_p^0, \cdots, \alpha_1^0)' = (\theta_{10}', \theta_{20}')'$ be the true value for the parameters of interest in (2.1). Without loss of generality, it is assumed that θ_{10} and θ_{20} denote the nonzero and zero components of θ_0, respectively. Let $Y = (1, X_{\ -p}, \cdots, X_{\ -1})'$ and define

$$\Theta = \lim_{n \to \infty} n^{-1} \sum_{=1}^{n} Y Y'. \tag{2.2}$$

Denote

$$\Gamma = (\sigma_{ij})_{(p+1) \times (p+1)}, \ \sigma_{ij} = \sigma_{ji}, \ 1 \leq i \leq j \leq p + 1, \tag{2.3}$$

where

$$
\begin{aligned}
\sigma_{11} &= E\left(X_{\boldsymbol{p}} - \sum_{k=1}^{\boldsymbol{p}} \alpha_k^0 X_{\boldsymbol{p}-k} - \mu_\epsilon^0\right)^2, \\
\sigma_{ii} &= E\left(X_{\boldsymbol{p}} - \sum_{k=1}^{\boldsymbol{p}} \alpha_k^0 X_{\boldsymbol{p}-k} - \mu_\epsilon^0\right)^2 X^2{}_{-(\boldsymbol{p}+2-i)}, 2 \leq i \leq p+1, \\
\sigma_{i1} &= E\left(X_{\boldsymbol{p}} - \sum_{k=1}^{\boldsymbol{p}} \alpha_k^0 X_{\boldsymbol{p}-k} - \mu_\epsilon^0\right)^2 X_{\ -(\boldsymbol{p}+2-i)}, 2 \leq i \leq p+1, \\
\sigma_{ij} &= E\left(X_{\boldsymbol{p}} - \sum_{k=1}^{\boldsymbol{p}} \alpha_k^0 X_{\boldsymbol{p}-k} - \mu_\epsilon^0\right)^2 X_{\ -(\boldsymbol{p}+2-i)} X_{\ -(\boldsymbol{p}+2-j)}, 2 \leq i < j \leq p+1.
\end{aligned}
$$

To prove the asymptotic properties of the regularized estimator in the next section, we will need the following lemma.

Lemma 1. *Suppose that the regularity conditions (C.1) and (C.2) in the Appendix hold. Then as $n \to \infty$, we have*

$$
\frac{1}{\sqrt{n}} B_n(\theta_0) \xrightarrow{D} N(0, \Gamma),
$$

where $\xrightarrow{D}$ denotes convergence in distribution, $B_n(\theta_0) = \sum_{\ =1}^{n} (X \ - \sum_{i=1}^{\boldsymbol{p}} \alpha_i^0 X_{\ -i} - \mu_\epsilon) Y$, and Γ is defined in (2.3).

3 Inference procedures

In this section, we will focus on the estimation for sparse regression coefficients in the GINAR(p) process. Our basic idea is to employ the penalty functions to shrink certain coefficients to be exactly zero. Note that $E(X \,|X_{\ -i}, 1 \leq i \leq p) = \sum_{i=1}^{\boldsymbol{p}} \alpha_i X_{\ -i} + \mu_\epsilon$, we can denote $S(\theta) = \frac{1}{2} \sum_{\ =1}^{n} (X \ - \sum_{i=1}^{\boldsymbol{p}} \alpha_i X_{\ -i} - \mu_\epsilon)^2$ as the conditional least squares (CLS) criterion function. Motivated by Fan and Li (2001), we propose the penalized estimating function $\mathscr{L}(\theta) = S(\theta) + n \sum_{i=1}^{\boldsymbol{p}+1} P_\lambda(|\theta_i|)$, where $P_\lambda(\cdot)$ is a penalty function. The regularized estimator for θ is defined as

$$
\hat{\theta} = \arg\min_{\theta} \mathscr{L}(\theta), \tag{3.1}
$$

where $P_\lambda(\cdot)$ is one of the following four penalty functions:

(P.1) The Adaptive LASSO, $P_\lambda(|\theta|) = \lambda\omega|\theta|$, where $\lambda > 0$ is the tuning parameter, and ω is a data-dependent weight (Zou, 2006).

(P.2) Fan and Li (2001) proposed the SCAD penalty function defined as

$$\dot{P}_\lambda(|\theta|) = \lambda \mathrm{sgn}(\theta)\left\{ I(|\theta| \le \lambda) + \frac{\max(\tau\lambda - |\theta|, 0)}{(\tau - 1)\lambda} I(|\theta| > \lambda) \right\},$$

where $\lambda > 0$ is the tuning parameter, and $\tau > 0$ is the shape parameter; $P_\lambda(0) = 0$ with $\dot{P}(\cdot)$ denotes the first derivative of $P(\cdot)$.

(P.3) Zhang (2010) proposed the following MCP penalty,

$$P_\lambda(|\theta|) = \lambda\left\{ |\theta| - \frac{|\theta|^2}{2\tau\lambda} \right\} I(0 \le |\theta| < \tau\lambda) + \frac{\lambda^2\tau}{2} I(|\theta| \ge \tau\lambda),$$

where $\lambda > 0$ is the tuning parameter, and $\tau > 0$ is the shape parameter.

(P.4) Dicker, et al. (2013) gave the SELO penalty function defined as

$$P_\lambda(|\theta|) = \frac{\lambda}{\log(2)} \log\left(\frac{|\theta|}{|\theta| + \tau} + 1 \right),$$

where $\lambda > 0$ is the tuning parameter, and $\tau > 0$ is the shape parameter.

We note that for the class of models under consideration $\theta \ge 0$, hence $\mathrm{sgn}(\theta)$ can be omitted and $|\theta| = \theta$. Define $b = (\dot{P}_{\lambda_n}(\theta_1^0), \cdots, \dot{P}_{\lambda_n}(\theta_s^0))'$, $\Sigma = \mathrm{diag}\{\ddot{P}_{\lambda_n}(\theta_1^0), \cdots, \ddot{P}_{\lambda_n}(\theta_s^0)\}$, as well as $a_n = \max_{1 \le i \le p+1}\{\dot{P}_{\lambda_n}(\theta_i^0), \theta_i^0 \ne 0\}$ and $b_n = \max_{1 \le i \le p+1}\{\ddot{P}_{\lambda_n}(\theta_i^0), \theta_i^0 \ne 0\}$, where s denotes the number of components in θ_{10}. Here and below, we use $\ddot{P}(\cdot)$ to denote the second derivative of $P(\cdot)$ and λ_n to emphasize λ's dependence on the sample size n.

Theorem 1. *Suppose that (C.1) - (C.4) in the Appendix hold. Then there exists a local minimizer of $\mathscr{L}(\theta)$, $\hat{\theta}$, such that $||\hat{\theta} - \theta_0|| = O_p(n^{-1/2} + a_n)$, where $\hat{\theta}$ is defined in (3.1).*

The theorem above gives the existence of $\hat{\theta}$ and says that $\hat{\theta}$ is $\sqrt{n}$-consistent. To establish the asymptotic normality of $\hat{\theta}$, we need the following lemma.

Lemma 2. *Suppose that the regularity conditions (C.1) - (C.4) given in the Appendix hold. Then with probability tending to 1, for any given θ_1 satisfying $\|\theta_1 - \theta_{10}\| = O_p(n^{-1/2})$ and any constant $\sigma > 0$, we have*

$$\mathscr{L}\left\{ \begin{pmatrix} \theta_1 \\ \mathbf{0} \end{pmatrix} \right\} = \min_{\|\theta_2\| \le \sigma n^{-1/2}} \mathscr{L}\left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \right\}.$$

The following results establish the oracle properties of the regularized estimators for the parameters of interest in GINAR(p) process, where the technical method follows from Fan and Li (2001).

Theorem 2. *Assume that the regularity conditions (C.1) - (C.4) given in the Appendix hold. Then with probability tending to 1, the $\sqrt{n}$-consistent estimate in Theorem 1 satisfies:*

(i) Sparsity: $\hat{\theta}_2 = 0$*;*

(ii) Asymptotic normality:

$$\sqrt{n}(\Theta^{\blacktriangle} + \Sigma)\{(\hat{\theta}_1 - \theta_{10}) + (\Theta^{\blacktriangle} + \Sigma)^{-1} b\} \xrightarrow{D} N(0, \Gamma^{\blacktriangle}), \tag{3.2}$$

where $\xrightarrow{D}$ *stands for convergence in distribution,* $\Theta^{\blacktriangle}$ *and* $\Gamma^{\blacktriangle}$ *denote the upper-left* $s \times s$ *submatrix of* Θ *and* Γ*, which are defined in (2.2) and (2.3), respectively.*

The asymptotic covariance matrix of $\hat{\theta}_1$ can be consistently estimated by the sample version of $\Theta^{\blacktriangle}$, Σ and $\Gamma^{\blacktriangle}$, respectively. Note that the regularized estimator $\hat{\theta}$ is unbiased and sparse, while the traditional CLS estimator is dense. To obtain the regularized estimator $\hat{\theta}$, it is required to choose the tuning parameter λ and the shape parameter τ. For given λ and τ, following Fan and Li (2001), we propose to apply the Newton-Raphson algorithm, which is a commonly used approach in the literature of penalty-based methods. Let $\tilde{\theta}$ be some initial value of θ that is assumed to be very close to the solution of the estimating equation $Q(\theta) = \partial \mathscr{L}(\theta)/\partial\theta = 0$ (e.g. the CLS estimator). The penalty function $P_\lambda(\cdot)$ can be irregular at the origin and may not has a second derivative at the origin. To address this problem, we use the technique of linear function approximation. For each j, if $\tilde{\theta}_j$ is not close to zero, we can use $\dot{P}_\lambda(\theta_j) \approx \{\dot{P}_\lambda(\theta_j^{(0)})/\theta_j^{(0)}\}\theta_j$, and otherwise, we set $\hat{\theta}_j = 0$. Thus, if θ is close to $\tilde{\theta}$, we get that $Q(\theta) \approx Q(\tilde{\theta}) + \dot{Q}(\tilde{\theta})(\theta - \tilde{\theta})$, which suggests that $Q(\theta)$ can be locally approximated by $Q(\tilde{\theta}) + n\Theta(\theta - \tilde{\theta}) + n\Sigma_\lambda(\tilde{\theta})(\theta - \tilde{\theta})$. Here $\Sigma_\lambda(\theta) = \mathrm{diag}\{\dot{P}_\lambda(\theta_1)/\theta_1, \cdots, \dot{P}_\lambda(\theta_{p+1})/\theta_{p+1}\}$. Thus for given λ and τ, the estimate can be obtained through the iteration $\theta^{(k+1)} = \theta^{(k)} - \{n\Theta + n\Sigma_\lambda(\theta^{(k)})\}^{-1} Q(\theta^{(k)})$ until $\|\theta^{(k+1)} - \theta^{(k)}\| \leq 10^{-4}$, $k = 0, 1, 2, \cdots$.

Denote $\hat{\theta}_{\lambda,\tau}$ as the estimate given by the above algorithm for given λ and τ (the adaptive LASSO only has λ). The parameters λ and τ can be obtained by minimizing the BIC statistic

$$\mathrm{BIC}(\lambda, \tau) = \log\left(\frac{\sum_{=1}^{n} U\,(\hat{\theta}_{\lambda,\tau})^2}{n - \hat{s}_{\lambda,\tau}}\right) + \frac{\log(n)}{n}\hat{s}_{\lambda,\tau}, \tag{3.3}$$

where $U\,(\theta) = X \; - \sum_{i=1}^{p} \alpha_i X_{\;-i} - \mu_\epsilon$, $\hat{s}_{\lambda,\tau}$ is the number of the non-zero components of $\hat{\theta}_{\lambda,\tau}$ and $n - \hat{s}_{\lambda,\tau}$ is used to account for the degrees lost to estimation (Wang, et al. 2007). Then $(\hat{\lambda}, \hat{\tau}) = \arg\min_{\lambda,\tau} \mathrm{BIC}(\lambda, \tau)$, and the proposed regularized estimator is $\hat{\theta} = \hat{\theta}_{\hat{\lambda},\hat{\tau}}$.

4 Numerical studies

4.1 Simulation

In this part, we will conduct simulation to assess the finite-sample performance of the proposed procedure. We pay our attention to five kinds of AR(7)-type processes, which are defined as follows.

Model I (*Binomial thinning*): $X \ = \sum_{i=1}^{7} \alpha_i \circ^G X_{\ -i} + \epsilon$, where the probability mass function of the "counting series" $\{W_j^{(i)}\}$ is $P(W_j^{(i)} = x) = \alpha_i (1-\alpha_i)^{1-} \ , x = 0, 1$, and $\alpha_1 = 0.45$, $\alpha_2 = \cdots = \alpha_6 = 0$, $\alpha_7 = 0.40$; $\{\epsilon \ \}$ is generated from the Poisson distribution with mean $\mu_\epsilon = 0.15$.

Model II (*Geometric thinning*): $X \ = \sum_{i=1}^{7} \alpha_i \circ^G X_{\ -i} + \epsilon$, where the probability mass function of the "counting series" $\{W_j^{(i)}\}$ is $P(W_j^{(i)} = x) = \frac{\alpha_i^x}{(1+\alpha_i)^{1+x}}$, $x = 0, 1, 2, \cdots$, and $\alpha_1 = 0.45$, $\alpha_2 = \cdots = \alpha_6 = 0$, $\alpha_7 = 0.40$; $\{\epsilon \ \}$ is generated from the Poisson distribution with mean $\mu_\epsilon = 0.15$.

Model III (*Poisson thinning*): $X \ = \sum_{i=1}^{7} \alpha_i \circ^G X_{\ -i} + \epsilon$, where the probability mass function of the "counting series" $\{W_j^{(i)}\}$ is $P(W_j^{(i)} = x) = \frac{\alpha_i^x}{!} e^{-\alpha_i}$, $x = 0, 1, 2, \cdots$. and $\alpha_1 = 0.45$, $\alpha_2 = \cdots = \alpha_6 = 0$, $\alpha_7 = 0.40$; $\{\epsilon \ \}$ is generated from the Poisson distribution with mean $\mu_\epsilon = 0.15$.

Model IV (*Dependent Bernoulli thinning*): $X \ = \sum_{i=1}^{7} \alpha_i \circ^G X_{\ -i} + \epsilon$, where the "counting series" $W_j^{(i)} = (1 - V_j^{(i)}) D_j^{(i)} + V_j^{(i)} Z^{(i)}$ (Ristić, et al., 2013) with $\{D_j^{(i)}\}_{j \in \mathbb{N}}$ is a sequence of i.i.d. Bernoulli(α_i) variables, $\alpha_i \in [0, 1)$; $\{V_j^{(i)}\}_{j \in \mathbb{N}}$ is a sequence of i.i.d. Bernoulli(γ_i) variables, $\gamma_i \in [0, 1)$; $Z^{(i)}$ is a random variable with Bernoulli(α_i) distribution. Here $\alpha_1 = 0.4$, $\alpha_2 = \cdots = \alpha_6 = 0$, $\alpha_7 = 0.35$; $\gamma_1 = 0.4$, $\gamma_2 = \cdots = \gamma_6 = 0$, $\gamma_7 = 0.45$; $\{\epsilon \ \}$ is generated from the Poisson distribution with mean $\mu_\epsilon = 0.15$.

Model V (*ρ-Binomial thinning*): $X \ = \sum_{i=1}^{7} \alpha_{\rho,i} \circ^G X_{\ -i} + \epsilon$, where the probability mass function of the "counting series" $\{W_j^{(i)}\}$ is $P(W_j^{(i)} = 0) = 1 - \alpha_i$; $P(W_j^{(i)} = x) = \alpha_i (\frac{\rho_i}{1+\rho_i})^{\ -1} (\frac{1}{1+\rho_i})$, $x = 1, 2, \cdots$ (Borges, et al. 2016). Here $\alpha_1 = 0.30$, $\alpha_2 = \cdots = \alpha_6 = 0$, $\alpha_7 = 0.35$; $\rho_1 = 0.12$, $\rho_2 = \cdots = \rho_6 = 0$, $\rho_7 = 0.15$; $\{\epsilon \ \}$ is generated from the Poisson distribution with mean $\mu_\epsilon = 0.15$.

We generate time series data from the above five models with the help of MATLAB software. Let $\Omega = \{1, 7, 8\}$ denote the index of nonzero elements in θ_0. Then we use the traditional CLS and the regularized method to estimate the parameters of interest, where the penalty functions include adaptive LASSO (denoted as ALASSO), SCAD, MCP and SELO. The weights for ALASSO are defined as $w_j = 1/|\tilde{\theta}_j|$, $j = 1, \cdots, p+1$, where $\tilde{\theta}$ is the traditional CLS for θ. We use a data-driven procedure to choose the tuning parameter λ and the shape parameter τ. This requires the computation of solution surface over a two-dimensional grid of (λ, τ). As suggested by one referee, we define the grid values in $[\lambda_{\min}, \lambda_{\max}] \times [\tau_{\min}, \tau_{\max}]$ to be $\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M = \lambda_{\min}$ and $\tau_{\min} = \tau_1 \leq \tau_2 \leq \cdots \leq \tau_K = \tau_{\max}$. The number of grid points $M = 50$ and $K = 5$, respectively. In our implementation, we choose some $\lambda_{\max} > 0$ (e.g. $\lambda_{\max} = 1.2$) which ensures that $\hat{\theta} = 0$ and set $\lambda_{\min} = \epsilon \lambda_{\max}$ for some $\epsilon > 0$ (e.g. $\epsilon = 0.001$) with $\lambda_k = \lambda_{\max} (\lambda_{\min}/\lambda_{\max})^{k-1/M-1}$ for $k = 1, \cdots, M$. We try and compare a small range of possible values for the shape parameter: $\tau \in \{2.5, 3, 3.7, 4.5, 5\}$ for the SCAD; $\tau \in \{1, 1.5, 2, 2.5, 3\}$ for the MCP; $\tau \in \{0.001, 0.005, 0.01, 0.05, 0.1\}$ for the SELO. To ensure the convergence of the algorithm, we use the solution from the previous position as a warm start. Finally, we choose the optimal $(\lambda,\, \tau)$ by minimizing the BIC criterion in (3.3). In Figure 1, we report the optimal value of λ in Model I and other cases are omitted here. All the results are based on 1000 replications with sample sizes $n = 500$, 1000 and 2000, respectively.

Tables 1-5 report the estimated bias (Bias) given by the average of the proposed estimate of θ minus the true value, the sample standard error (SSE) of the proposed estimate, the average of the estimated standard error (ESE) in Theorem 2. Because the traditional CLS estimates of the zero parameters $(\hat{\alpha}_2 - \hat{\alpha}_6)$ may be negative, so we adjust the CLS with negative estimates as zeros. From the results we can see that the standard error estimates corresponding to non-zero entries in θ_0 agreed well with the empirical standard error. However, the standard error estimators corresponding to zero entries of θ_0 appeared to underestimate the true variability of the regularized estimators. Similar phenomena were also observed in Table 2 of Dicker, et al. (2013).

The variable selection results are reported in Tables 6 - 10, which include the estimated average model size $\hat{\Omega} = \{j; \hat{\theta}_j \neq 0\}$ (MS), the rate that the true model was selected $I\{\hat{\Omega} = \Omega\}$ (CMR), the false positive rate $|\hat{\Omega}\backslash\Omega|/|\hat{\Omega}|$ $(F+)$, and the false negative rate $|\Omega\backslash\hat{\Omega}|/(p - |\hat{\Omega}|)$ $(F-)$. Although all methods tend to overestimate the true model with respect to the model size, the SELO-based method seems to have the highest selection of the correct model and the smallest model size, which are in line with the performances of false positive rate and false negative rate.

4.2 Application

In this section, we apply the proposed methodology to the daily epileptic seizure counts analyzed by Latour (1998). For the analysis, let $\{X\ \}$ represents the daily number of seizures of an epileptic patient before treatment, $t = 1, \cdots, 120$. The data are available from Latour (1998) and are presented in Figure 2. The plots of autocorrelation function (ACF) and partial autocorrelation function (PACF) are given in Figures 3 and 4, respectively. It is easy to see that we may describe the count time series $\{X\ \}$ using the GINAR(14) process.

We use the regularized method in Section 3 with four different penalty functions to estimate the parameters of interest. The selection process of (λ, τ) is similar to those described in Section 4.1. The estimators, standard errors (SE) and their optimal values of λ and τ are reported in Table 11, which suggest that $X\ = \alpha_6 \circ^G X_{\ -6} + \alpha_{14} \circ^G X_{\ -14} + \epsilon\ $ is the fitted model. Note that the SEs of CLS are much larger than those of the regularized estimators, which indicate that the proposed approach provides an improvement over the traditional CLS estimator. Moreover, the fitted values are defined as $\hat{X} = \langle \sum_{i=1}^{14} \hat{\alpha}_i X_{-i} + \hat{\mu}_\epsilon \rangle$, where $\langle \cdot \rangle$ denotes the rounding operator to the nearest integer (Kachour and Yao, 2009). Figure 2 reports the fitted count values with the SELO estimators (Other cases are similar and omitted here).

To check the model adequacy, we also study the relevant properties of the fitted model and compare with the corresponding empirical ones (mean, ACF and PACF) in Tables 12 and 13, where the theoretical mean of X is $\mu = \mu_\epsilon / (1 - \sum_{i=1}^{p} \alpha_i)$. Also the ACF and PACF of the residual for the fitted model with SELO procedure are reported in Figures 5 and 6, respectively (Other cases are similar, so we omit those figures). From these results we can see that the GINAR (14) process is suitable to model the daily epileptic seizure number of this patient. This conclusion is in line with Latour (1998), which suggests that the proposed method is acceptable in practice.

5 Conclusion and remarks

In this article, we used the penalized approach to estimate the parameters of interest in the GINAR(p) process. We have derived the oracle properties and illustrated the usefulness of the proposed estimation methodology via some simulation studies and a real data example.

There exist some topics for future study. As one referee suggested, we can penalize the likelihood function for the GINAR(p) process, instead of using the conditional least squares function in (3.1). However, the likelihood function is very complicated especially for high-order integer-valued autoregressive processes. A possible way is to use the saddlepoint approximation technique for the likelihood (Pedeli, et al. 2015). This is an important and challenging problem that deserves further careful study, but is beyond the scope of the current paper. Furthermore, the proposed approach can be extended to other kinds of integer-valued time series data (Zhang, et al., 2010; Scotto, et al., 2014).

Acknowledgements

The authors would like to thank the Editor, the Associate Editor and three reviewers for their constructive and insightful comments and suggestions that greatly improved the manuscript. Haixiang Zhang's work is supported by National Natural Science Foundation of China(Nos. 11301212 and 11401146). Dehui Wang's work is supported by National Natural Science Foundation of China (Nos. 11271155 and 11001105). Liuquan Sun's work is partially supported by the National Natural Science Foundation of China Grants (Nos. 11231010 and 11690015) and Key Laboratory of RCSDS, CAS (No.2008DP173182).

6 Appendix

For the proofs, we need the following regularity conditions:

(C.1) $\{X\ \}$ is a strictly stationary and ergodic process.

(C.2) $E(|X\ |^4) < \infty$.

(C.3) $a_n = O(n^{-1/2})$.

(C.4) $b_n = o(1)$.

Here (C.1) and (C.2) are used for the asymptotic properties of $\hat{\theta}$ (Zhang, et al., 2010). (C.3) is to ensure that the estimators are $\sqrt{n}$-consistent. We use (C.4) to make sure that the influence of penalty function does not exceed that of CLS criterion function on the resulting estimator.

Proof of Lemma 1. Define

$$\begin{aligned} M_{n1} &= \sum_{=1}^{n} \left(X\ - \sum_{i=1}^{p} \alpha_i^0 X_{\ -i} - \mu_\epsilon^0 \right), \\ M_{ni} &= \sum_{=1}^{n} \left(X\ - \sum_{i=1}^{p} \alpha_i^0 X_{\ -i} - \mu_\epsilon^0 \right) X_{\ -(p+2-i)}, \quad 2 \leq i \leq p+1. \end{aligned}$$

Let $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$. Then it is easy to derive that

$$\begin{aligned} E(M_{n1}|\mathcal{F}_{n-1}) &= E\left(M_{(n-1)1} + X_n - \sum_{i=1}^{p} \alpha_i^0 X_{n-i} - \mu_\epsilon^0 | \mathcal{F}_{n-1} \right) \\ &= M_{(n-1)1} + E\left(X_n - \sum_{i=1}^{p} \alpha_i^0 X_{n-i} - \mu_\epsilon^0 | \mathcal{F}_{n-1} \right) \\ &= M_{(n-1)1}, \end{aligned}$$

so $\{M_{n1}, \mathcal{F}_n, n \geq 1\}$ is a martingale. Using $E|X\ |^4 < \infty$, the strict stationarity of $\{X\ \}$ and the ergodic theorem, we have

$$\frac{1}{n} \sum_{=1}^{n} \left(X\ - \sum_{i=1}^{p} \alpha_i^0 X_{\ -i} - \mu_\epsilon^0 \right)^2 \xrightarrow{a.s.} \sigma_{11}.$$

Then, by Corollary 3.2 in Hall and Heyde(1980),

$$\frac{1}{\sqrt{n}} M_{n1} \xrightarrow{D} N(0, \sigma_{11}).$$

Similarly, we can prove that $\{M_{ni}, \mathcal{F}_n, n \geq 1\}$ is a martingale, and for $i = 2, \cdots, p+1$,

$$\frac{1}{\sqrt{n}} M_{ni} \xrightarrow{D} N(0, \sigma_{ii}).$$

Also, for any $c = (c_1, \cdots, c_{p+1})' \in \mathbb{R}^{p+1} \setminus (0, \cdots, 0)$,

$$\begin{aligned} \frac{1}{\sqrt{n}} c' \begin{pmatrix} M_{n1} \\ \vdots \\ M_{n(p+1)} \end{pmatrix} &= \frac{1}{\sqrt{n}} \sum_{=1}^{n} (c_1 + c_2 X_{\ -p} + \cdots + c_{p+1} X_{\ -1})(X_{\ } - \sum_{i=1}^{p} \alpha_i^0 X_{\ -i} - \mu_\epsilon^0) \\ &\xrightarrow{D} N\left(0, E(c_1 + c_2 X_0 + \cdots + c_{p+1} X_{p-1})^2 (X_p - \sum_{i=1}^{p} \alpha_i^0 X_{p-i} - \mu_\epsilon^0)^2\right). \end{aligned}$$

Thus, by the Cramer-Wold device, we can obtain the desired results. This ends the proof. □

Proof of Lemma 2. To prove the lemma, it suffices to show that with probability tending to 1 as $n \to \infty$, for given θ_1 satisfying $\|\theta_1 - \theta_{10}\| = O_p(n^{-1/2})$ and for some small $\epsilon_n = \sigma n^{-1/2}$ and $j = s+1, \cdots, p+1$,

$$\begin{aligned} \frac{\partial \mathscr{L}(\theta)}{\partial \theta_j} &> 0, \quad \text{for } 0 < \theta_j < \epsilon_n, \\ \frac{\partial \mathscr{L}(\theta)}{\partial \theta_j} &< 0, \quad \text{for } -\epsilon_n < \theta_j < 0. \end{aligned} \tag{6.1}$$

Let $U_{\ }(\theta) = X_{\ } - \sum_{i=1}^{p} \alpha_i X_{\ -i} - \mu_\epsilon$ and $Y_{\ } = (1, X_{\ -p}, \cdots, X_{\ -1})'$. Then, by the Taylor's expansion, we can derive that

$$\begin{aligned} \frac{\partial \mathscr{L}(\theta)}{\partial \theta_j} &= -\sum_{=1}^{n} U_{\ }(\theta) X_{\ -(p+2-j)} + n \dot{P}_{\lambda_n}(\theta_j) \\ &= I_1 + I_2 + n \dot{P}_{\lambda_n}(\theta_j), \end{aligned}$$

where $I_1 = -\sum_{=1}^{n} U_{\ }(\theta_0) X_{\ -(p+2-j)}$, and $I_2 = \sum_{=1}^{n} X_{\ -(p+2-j)} Y'_{\ }(\theta - \theta_0)$. Similar to Lemma 1, applying the martingale central limit theorem, we obtain that $I_1 = O_p(n^{1/2})$. By the strict stationarity and the law of large numbers, together with $\theta - \theta_0 = O_p(n^{-1/2})$, we get that $I_2 = O_p(n^{1/2})$. Thus, $\frac{\partial \mathscr{L}(\theta)}{\partial \theta_j} = n\lambda_n \{O_p(n^{-1/2}/\lambda_n) + \lambda_n^{-1} \dot{P}_{\lambda_n}(\theta_j)$. Since $n^{-1/2}/\lambda_n = o(1)$ and $\lambda_n^{-1} \dot{P}_{\lambda_n}(\theta_j) > 0$, the sign of $\frac{\partial \mathscr{L}(\theta)}{\partial \theta_j}$ is completely determined by that of θ_j. Hence, (6.1) follows. This completes the proof. □

Proof of Theorem 1. Let $\alpha_n = n^{-1/2} + a_n$. To prove the theorem, we need to show that for any given $\epsilon > 0$, there exists a constant C such that

$$P\Big\{\mathscr{L}(\theta_0) < \inf_{\| \ \| = C} \mathscr{L}(\theta_0 + \alpha_n u)\Big\} \geq 1 - \epsilon. \tag{6.2}$$

Then, there exists a local minima in the ball $\{\theta_0 + \alpha_n u : \|u\| \leq C\}$ with probability at least $1 - \epsilon$. Thus, there exists a local minimizer $\hat{\theta}$ such that $\|\hat{\theta} - \theta_0\| = O_p(\alpha_n)$.

Define $D_n(u) = \mathscr{L}(\theta_0 + \alpha_n u) - \mathscr{L}(\beta_0)$, using $p_{\lambda_n}(0) = 0$, we get that

$$D_n(u) \geq S(\theta_0 + \alpha_n u) - S(\theta_0) + n \sum_{j=1}^{s} \Big\{ P_{\lambda_n}(\theta_j^0 + \alpha_n u_j) - P_{\lambda_n}(\theta_j^0) \Big\},$$

where s is the number of components of θ_{10}. By the Taylor expansion, we have

$$\begin{aligned} D_n(u) \;\; \geq \;\; & -\alpha_n B'(\theta_0) u + \frac{1}{2} u' A(\theta_0) u n \alpha_n^2 \{1 + o_p(1)\} \\ & + \sum_{j=1}^{s} \Big[n\alpha_n \dot{P}_{\lambda_n}(\theta_j^0) u_j + n\alpha_n^2 \ddot{P}_{\lambda_n}(\theta_j^0) u_j^2 \{1 + o(1)\} \Big]. \end{aligned} \tag{6.3}$$

It follows from Lemma 1 that $n^{-1/2} B(\theta_0) = O_p(1)$. Then the first term on the right-hand side of (6.3) is on the order $O_p(n^{1/2}\alpha_n) = O_p(n\alpha_n^2)$. Furthermore, the second term dominates the first term uniformly in $\|u\| = C$, where C is sufficiently large. Note that the third term of (6.3) is bounded by

$$\sqrt{s} n \alpha_n a_n \|u\| + n\alpha_n^2 \max_{1 \leq j \leq p+1} \{\ddot{P}_{\lambda_n}(\theta_j^0), \theta_j^0 \neq 0\} \|u\|^2. \tag{6.4}$$

It is easy to see that (6.4) is also dominated by the second term of (6.3). Thus, by choosing a sufficiently large C, we conclude that (6.2) holds. This completes the proof. □

Proof of Theorem 2. Part (i) directly follows from Lemma 1. For part (ii), let $Q(\theta) = \partial \mathscr{L}(\theta)/\partial\theta$. It can be easily shown that there exists a $\hat{\theta}_1$ in Theorem 1 that is a $\sqrt{n}$-consistent local minimizer of $\mathscr{L}\{(\theta_1', 0)'\}$, which satisfies the equation $Q\{(\hat{\theta}_1', 0)'\} = 0$. Let $Y_t^s = (1, X_{t-p}, \cdots, X_{t-p-s+1})'$. Since $\hat{\theta}_1$ is a consistent estimator, by the Taylor expansion, we get

$$\begin{aligned} Q^s(\hat{\theta}_1) \;\; &= \;\; -\sum_{t=1}^{n} U_t(\hat{\theta}) Y_t^s + n(\dot{P}_{\lambda_n}(\hat{\theta}_{11}), \cdots, \dot{P}_{\lambda_n}(\hat{\theta}_{1s}))' \\ &= \;\; -\sum_{t=1}^{n} U_t(\theta_0) Y_t^s + \sum_{t=1}^{n} Y_t^s Y_t^{s\prime} \times (\hat{\theta}_1 - \theta_{10}) \\ & \quad\;\; + n[b + \{\mathrm{diag}\{\ddot{P}_{\lambda_n}(\theta_{10}), \cdots, \ddot{P}_{\lambda_n}(\theta_{s0})\} + o_p(1)\} \times (\hat{\theta}_1 - \theta_{10})], \end{aligned}$$

which yields that

$$\sqrt{n}(\Theta^s + \Sigma)\{(\hat{\theta}_1 - \theta_{10}) + (\Theta^s + \Sigma)^{-1} b\} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t(\theta_0) Y_t^s + o_p(1).$$

Similar to the proof of Lemma 1, by the Slutsky's theorem and the martingale central limit theorem, the desired results are obtained. This completes the proof. □

References

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

Barreto-Souza, W. (2015). Zero-modified geometric INAR(1) process for modelling count time series with deflation or inflation of zeros. *Journal of Time Series Analysis*, **36**, 839 - 852.

Borges, P., Molinares, F. and Bourguignon, M. (2016). A geometric time series model with inflated-parameter Bernoulli counting series. *Statistics and Probability Letters*, **119**, 264 - 272.

Dicker, L. Huang, B. and Lin, X. (2013). Variable selection and estimation with the seamless-L_0 penalty. *Statistica Sinica*, **23**, 929 - 962.

Drost, F. C., Akker, R. V., Werker, B. J. (2009). Efficient estimation of auto-regression parameters and innovation distributions for semi-parametric integer-valued AR(p) models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 467 - 485.

Fan, J. and Li, R. (2001). Variable selection via nonconvace penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348 - 1360.

Fokianos, K. (2010). *Penalized estimation for integer autoregressive models*. In Thomas Kneib and Gerhard Tutz, editors, Statistical Modelling and Regression Structures, pages 337-352. Physica-Verlag HD. ISBN 978-3-7908-2413-1.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Application*. Academic Press: New York.

Kachour, M. and Yao, J. (2009). First-order rounded integer-valued autoregressive (RI-NAR(1)) process. *Journal of Time Series Analysis*, **30**, 417-448.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, **19**, 439-455.

McCabe, B., Martin, G. and Harris, D. (2011). Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**, 253 - 272.

Pedeli, X., Davison, A. and Fokianos K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association*, **110**, 1229-1238.

Ristić, M., Nastić, A. and Miletić Ilić, A. (2013). A geometric time series model with dependent Bernoulli counting series. *Journal of Time Series Analysis*, **34**, 466 - 476.

Scotto, M., Weiß, C., Silva, M. and Pereira, I. (2014). Bivariate binomial autoregressive models. *Journal of Multivariate Analysis*, **125**, 233 - 251.

Scotto, M., Weiß, C., Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, **15**, 590 - 618.

Steutel, F. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, **7**, 893-899.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **58**, 267 - 288.

Wang, H., Li, R., and Tsai, C. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, **94**, 553 - 568.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, **38**, 894 - 942.

Zhang, H., Wang, D. and Zhu, F. (2010). Inference for INAR(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference*, **140**, 667 - 683.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418 - 1429.

Table 1.

Bias, (SSE) and [ESE] of the parameters in Model I.

Sample size	Method	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$	$\hat{\alpha}_7$	$\hat{\mu}_\epsilon$
$n = 500$	CLS	-0.0099	0.0178	0.0199	0.0173	0.0193	0.0184	-0.0119	0.0319
		(0.0522)	(0.0288)	(0.0292)	(0.0272)	(0.0285)	(0.0301)	(0.0502)	(0.0600)
		[0.0517]	[0.0528]	[0.0512]	[0.0508]	[0.0513]	[0.0528]	[0.0516]	[0.0587]
	ALASSO	-0.0106	0.0104	0.0084	0.0075	0.0089	0.0114	-0.0135	0.0018
		(0.0515)	(0.0254)	(0.0205)	(0.0201)	(0.0222)	(0.0271)	(0.0521)	(0.0681)
		[0.0385]	[0.0027]	[0.0020]	[0.0019]	[0.0021]	[0.0033]	[0.0381]	[0.0312]
	SCAD	-0.0022	0.0054	0.0045	0.0035	0.0039	0.0057	-0.0055	0.0147
		(0.0555)	(0.0251)	(0.0194)	(0.0203)	(0.0157)	(0.0239)	(0.0534)	(0.0788)
		[0.0406]	[0.0027]	[0.0019]	[0.0019]	[0.0017]	[0.0029]	[0.0402]	[0.0302]
	MCP	-0.0001	0.0057	0.0045	0.0040	0.0047	0.0065	-0.0037	0.0187
		(0.0561)	(0.0191)	(0.0166)	(0.0161)	(0.0162)	(0.0215)	(0.0528)	(0.0775)
		[0.0396]	[0.0018]	[0.0012]	[0.0011]	[0.0013]	[0.0022]	[0.0392]	[0.0302]
	SELO	-0.0041	0.0046	0.0039	0.0040	0.0041	0.0051	-0.0048	0.0155
		(0.0520)	(0.0195)	(0.0170)	(0.0171)	(0.0167)	(0.0208)	(0.0517)	(0.0621)
		[0.0403]	[0.0012]	[0.0010]	[0.0009]	[0.0009]	[0.0014]	[0.0398]	[0.0314]
$n = 1000$	CLS	-0.0033	0.0130	0.0143	0.0132	0.0130	0.0145	-0.0085	0.0156
		(0.0349)	(0.0198)	(0.0211)	(0.0201)	(0.0202)	(0.0217)	(0.0368)	(0.0394)
		[0.0363]	[0.0370]	[0.0360]	[0.0356]	[0.0360]	[0.0370]	[0.0361]	[0.0400]
	ALASSO	-0.0046	0.0057	0.0055	0.0062	0.0051	0.0068	-0.0071	-0.0039
		(0.0347)	(0.0150)	(0.0141)	(0.0152)	(0.0128)	(0.0163)	(0.0366)	(0.0477)
		[0.0289]	[0.0019]	[0.0015]	[0.0016]	[0.0016]	[0.0021]	[0.0287]	[0.0246]
	SCAD	0.0003	0.0029	0.0024	0.0024	0.0018	0.0032	-0.0038	0.0080
		(0.0377)	(0.0140)	(0.0114)	(0.0099)	(0.0088)	(0.0134)	(0.0375)	(0.0534)
		[0.0295]	[0.0016]	[0.0010]	[0.0011]	[0.0007]	[0.0016]	[0.0293]	[0.0240]
	MCP	0.0014	0.0030	0.0026	0.0031	0.0022	0.0032	-0.0019	0.0066
		(0.0373)	(0.0116)	(0.0100)	(0.0117)	(0.0090)	(0.0125)	(0.0380)	(0.0519)
		[0.0290]	[0.0011]	[0.0009]	[0.0009]	[0.0008]	[0.0013]	[0.0289]	[0.0236]
	SELO	-0.0004	0.0024	0.0021	0.0020	0.0015	0.0030	-0.0009	0.0020
		(0.0357)	(0.0111)	(0.0101)	(0.0096)	(0.0080)	(0.0127)	(0.0362)	(0.0417)
		[0.0295]	[0.0007]	[0.0007]	[0.0007]	[0.0005]	[0.0010]	[0.0294]	[0.0251]
$n = 2000$	CLS	-0.0023	0.0097	0.0097	0.0096	0.0094	0.0102	-0.0039	0.0081
		(0.0259)	(0.0145)	(0.0145)	(0.0149)	(0.0145)	(0.0153)	(0.0257)	(0.0276)
		[0.0256]	[0.0261]	[0.0253]	[0.0250]	[0.0253]	[0.0261]	[0.0256]	[0.0280]
	ALASSO	-0.0019	0.0046	0.0037	0.0039	0.0034	0.0048	-0.0055	-0.0054
		(0.0253)	(0.0113)	(0.0097)	(0.0103)	(0.0089)	(0.0123)	(0.0250)	(0.0338)
		[0.0208]	[0.0013]	[0.0010]	[0.0009]	[0.0009]	[0.0014]	[0.0207]	[0.0188]
	SCAD	0.0014	0.0017	0.0015	0.0016	0.0012	0.0021	-0.0034	0.0059
		(0.0262)	(0.0098)	(0.0072)	(0.0076)	(0.0071)	(0.0103)	(0.0263)	(0.0332)
		[0.0213]	[0.0012]	[0.0008]	[0.0006]	[0.0008]	[0.0013]	[0.0211]	[0.0189]
	MCP	0.0016	0.0025	0.0018	0.0016	0.0016	0.0028	-0.0029	0.0049
		(0.0261)	(0.0085)	(0.0070)	(0.0068)	(0.0063)	(0.0101)	(0.0270)	(0.0348)
		[0.0211]	[0.0010]	[0.0007]	[0.0007]	[0.0007]	[0.0011]	[0.0208]	[0.0188]
	SELO	0.0008	0.0015	0.0011	0.0014	0.0011	0.0020	-0.0021	0.0004
		(0.0255)	(0.0076)	(0.0060)	(0.0071)	(0.0063)	(0.0092)	(0.0261)	(0.0297)
		[0.0213]	[0.0005]	[0.0003]	[0.0005]	[0.0003]	[0.0006]	[0.0212]	[0.0196]

Table 2.

Bias, (SSE) and [ESE] of the parameters in Model II.

Sample size	Method	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$	$\hat{\alpha}_7$	$\hat{\mu}_\epsilon$
$n = 500$	CLS	-0.0142	0.0214	0.0208	0.0187	0.0192	0.0216	-0.0221	0.0418
		(0.0678)	(0.0351)	(0.0335)	(0.0299)	(0.0311)	(0.0329)	(0.0656)	(0.0644)
		[0.0671]	[0.0621]	[0.0571]	[0.0550]	[0.0557]	[0.0599]	[0.0649]	[0.0660]
	ALASSO	-0.0202	0.0113	0.0086	0.0080	0.0089	0.0106	-0.0240	0.0098
		(0.0677)	(0.0268)	(0.0239)	(0.0229)	(0.0235)	(0.0273)	(0.0648)	(0.0749)
		[0.0518]	[0.0044]	[0.0028]	[0.0023]	[0.0024]	[0.0036]	[0.0500]	[0.0339]
	SCAD	-0.0139	0.0074	0.0060	0.0065	0.0051	0.0069	-0.0193	0.0105
		(0.0714)	(0.0272)	(0.0233)	(0.0241)	(0.0223)	(0.0279)	(0.0657)	(0.0850)
		[0.0541]	[0.0058]	[0.0042]	[0.0037]	[0.0039]	[0.0052]	[0.0525]	[0.0322]
	MCP	-0.0129	0.0078	0.0054	0.0050	0.0049	0.0069	-0.0185	0.0255
		(0.0704)	(0.0227)	(0.0196)	(0.0200)	(0.0184)	(0.0236)	(0.0668)	(0.0843)
		[0.0535]	[0.0033]	[0.0022]	[0.0019]	[0.0017]	[0.0027]	[0.0518]	[0.0321]
	SELO	-0.0122	0.0057	0.0050	0.0048	0.0040	0.0064	-0.0146	0.0246
		(0.0680)	(0.0210)	(0.0223)	(0.0221)	(0.0176)	(0.0257)	(0.0657)	(0.0636)
		[0.0535]	[0.0021]	[0.0014]	[0.0012]	[0.0011]	[0.0019]	[0.0519]	[0.0354]
$n = 1000$	CLS	-0.0106	0.0166	0.0158	0.0143	0.0153	0.0157	-0.0104	0.0238
		(0.0486)	(0.0249)	(0.0239)	(0.0229)	(0.0238)	(0.0246)	(0.0443)	(0.0440)
		[0.0486]	[0.0450]	[0.0409]	[0.0394]	[0.0399]	[0.0430]	[0.0466]	[0.0464]
	ALASSO	-0.0110	0.0084	0.0059	0.0060	0.0053	0.0080	-0.0133	0.0034
		(0.0485)	(0.0212)	(0.0157)	(0.0160)	(0.0144)	(0.0193)	(0.0461)	(0.0559)
		[0.0388]	[0.0034]	[0.0021]	[0.0019]	[0.0016]	[0.0030]	[0.0373]	[0.0277]
	SCAD	-0.0075	0.0049	0.0043	0.0037	0.0031	0.0045	-0.0116	0.0085
		(0.0509)	(0.0224)	(0.0169)	(0.0151)	(0.0167)	(0.0176)	(0.0480)	(0.0626)
		[0.0402]	[0.0038]	[0.0027]	[0.0023]	[0.0023]	[0.0032]	[0.0387]	[0.0261]
	MCP	-0.0066	0.0061	0.0033	0.0033	0.0032	0.0045	-0.0111	0.0196
		(0.0503)	(0.0199)	(0.0127)	(0.0129)	(0.0126)	(0.0157)	(0.0484)	(0.0580)
		[0.0400]	[0.0023]	[0.0014]	[0.0012]	[0.0011]	[0.0019]	[0.0384]	[0.0266]
	SELO	-0.0054	0.0048	0.0024	0.0020	0.0022	0.0041	-0.0066	0.0102
		(0.0491)	(0.0180)	(0.0117)	(0.0117)	(0.0116)	(0.0154)	(0.0468)	(0.0483)
		[0.0403]	[0.0019]	[0.0009]	[0.0007]	[0.0007]	[0.0014]	[0.0388]	[0.0297]
$n = 2000$	CLS	-0.0054	0.0122	0.0121	0.0099	0.0108	0.0107	-0.0063	0.0129
		(0.0352)	(0.0186)	(0.0178)	(0.0150)	(0.0165)	(0.0173)	(0.0327)	(0.0310)
		[0.0347]	[0.0322]	[0.0294]	[0.0282]	[0.0285]	[0.0305]	[0.0334]	[0.0329]
	ALASSO	-0.0055	0.0055	0.0044	0.0040	0.0039	0.0052	-0.0071	-0.0029
		(0.0339)	(0.0135)	(0.0108)	(0.0107)	(0.0102)	(0.0133)	(0.0321)	(0.0403)
		[0.0287]	[0.0022]	[0.0015]	[0.0014]	[0.0013]	[0.0019]	[0.0278]	[0.0220]
	SCAD	-0.0030	0.0025	0.0029	0.0019	0.0015	0.0037	-0.0067	0.0058
		(0.0355)	(0.0156)	(0.0109)	(0.0098)	(0.0098)	(0.0134)	(0.0324)	(0.0426)
		[0.0295]	[0.0024]	[0.0017]	[0.0014]	[0.0014]	[0.0022]	[0.0285]	[0.0215]
	MCP	-0.0029	0.0034	0.0026	0.0020	0.0020	0.0034	-0.0056	0.0100
		(0.0352)	(0.0112)	(0.0093)	(0.0076)	(0.0076)	(0.0115)	(0.0322)	(0.0398)
		[0.0296]	[0.0017]	[0.0012]	[0.0009]	[0.0009]	[0.0015]	[0.0286]	[0.0223]
	SELO	-0.0020	0.0028	0.0017	0.0012	0.0011	0.0029	-0.0025	0.0013
		(0.0339)	(0.0115)	(0.0081)	(0.0065)	(0.0067)	(0.0112)	(0.0322)	(0.0341)
		[0.0295]	[0.0011]	[0.0007]	[0.0004]	[0.0004]	[0.0011]	[0.0287]	[0.0234]

Table 3.

Bias, (SSE) and [ESE] of the parameters in Model III.

Sample size	Method	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$	$\hat{\alpha}_7$	$\hat{\mu}_\epsilon$
$n = 500$	CLS	-0.0092	0.0206	0.0202	0.0181	0.0203	0.0216	-0.0210	0.0366
		(0.0616)	(0.0330)	(0.0321)	(0.0294)	(0.0312)	(0.0324)	(0.0574)	(0.0634)
		[0.0599]	[0.0579]	[0.0546]	[0.0532]	[0.0539]	[0.0567]	[0.0585]	[0.0621]
	ALASSO	-0.0153	0.0107	0.0078	0.0073	0.0080	0.0117	-0.0206	0.0048
		(0.0606)	(0.0262)	(0.0199)	(0.0200)	(0.0203)	(0.0282)	(0.0593)	(0.0695)
		[0.0463]	[0.0037]	[0.0024]	[0.0020]	[0.0021]	[0.0038]	[0.0449]	[0.0323]
	SCAD	-0.0073	0.0068	0.0045	0.0048	0.0050	0.0065	-0.0146	0.0132
		(0.0647)	(0.0243)	(0.0191)	(0.0212)	(0.0177)	(0.0230)	(0.0606)	(0.0847)
		[0.0483]	[0.0035]	[0.0024]	[0.0025]	[0.0023]	[0.0032]	[0.0469]	[0.0308]
	MCP	-0.0049	0.0059	0.0049	0.0044	0.0044	0.0076	-0.0137	0.0227
		(0.0648)	(0.0223)	(0.0179)	(0.0161)	(0.0163)	(0.0238)	(0.0614)	(0.0819)
		[0.0480]	[0.0021]	[0.0019]	[0.0014]	[0.0015]	[0.0030]	[0.0466]	[0.0316]
	SELO	-0.0071	0.0061	0.0041	0.0031	0.0039	0.0059	-0.0098	0.0203
		(0.0621)	(0.0223)	(0.0175)	(0.0152)	(0.0174)	(0.0219)	(0.0592)	(0.0613)
		[0.0484]	[0.0019]	[0.0013]	[0.0009]	[0.0011]	[0.0019]	[0.0470]	[0.0350]
$n = 1000$	CLS	-0.0060	0.0129	0.0148	0.0143	0.0131	0.0140	-0.0096	0.0230
		(0.0428)	(0.0210)	(0.0222)	(0.0223)	(0.0205)	(0.0226)	(0.0422)	(0.0438)
		[0.0426]	[0.0409]	[0.0385]	[0.0374]	[0.0380]	[0.0401]	[0.0417]	[0.0434]
	ALASSO	-0.0064	0.0069	0.0053	0.0058	0.0065	0.0072	-0.0111	-0.0015
		(0.0415)	(0.0172)	(0.0149)	(0.0158)	(0.0161)	(0.0177)	(0.0400)	(0.0525)
		[0.0344]	[0.0025]	[0.0015]	[0.0018]	[0.0019]	[0.0025]	[0.0334]	[0.0263]
	SCAD	-0.0028	0.0036	0.0035	0.0029	0.0037	0.0040	-0.0102	0.0090
		(0.0441)	(0.0166)	(0.0143)	(0.0119)	(0.0140)	(0.0162)	(0.0400)	(0.0580)
		[0.0356]	[0.0024]	[0.0015]	[0.0014]	[0.0016]	[0.0025]	[0.0346]	[0.0248]
	MCP	-0.0022	0.0042	0.0029	0.0033	0.0030	0.0041	-0.0086	0.0148
		(0.0446)	(0.0140)	(0.0120)	(0.0130)	(0.0119)	(0.0142)	(0.0406)	(0.0557)
		[0.0354]	[0.0017]	[0.0010]	[0.0010]	[0.0011]	[0.0018]	[0.0343]	[0.0256]
	SELO	-0.0024	0.0032	0.0024	0.0024	0.0026	0.0033	-0.0051	0.0073
		(0.0423)	(0.0131)	(0.0123)	(0.0119)	(0.0114)	(0.0126)	(0.0402)	(0.0450)
		[0.0355]	[0.0013]	[0.0007]	[0.0007]	[0.0008]	[0.0012]	[0.0345]	[0.0276]
$n = 2000$	CLS	-0.0027	0.0109	0.0110	0.0098	0.0110	0.0100	-0.0048	0.0097
		(0.0305)	(0.0166)	(0.0165)	(0.0148)	(0.0165)	(0.0154)	(0.0301)	(0.0294)
		[0.0305]	[0.0293]	[0.0276]	[0.0268]	[0.0272]	[0.0287]	[0.0297]	[0.0309]
	ALASSO	-0.0046	0.0052	0.0040	0.0048	0.0041	0.0052	-0.0081	-0.0033
		(0.0293)	(0.0129)	(0.0110)	(0.0118)	(0.0106)	(0.0127)	(0.0293)	(0.0382)
		[0.0250]	[0.0018]	[0.0014]	[0.0014]	[0.0012]	[0.0018]	[0.0244]	[0.0204]
	SCAD	-0.0012	0.0017	0.0020	0.0026	0.0019	0.0031	-0.0067	0.0074
		(0.0306)	(0.0130)	(0.0087)	(0.0109)	(0.0092)	(0.0114)	(0.0290)	(0.0364)
		[0.0258]	[0.0017]	[0.0012]	[0.0013]	[0.0012]	[0.0018]	[0.0252]	[0.0203]
	MCP	-0.0019	0.0034	0.0019	0.0023	0.0021	0.0029	-0.0068	0.0096
		(0.0307)	(0.0119)	(0.0078)	(0.0086)	(0.0082)	(0.0101)	(0.0293)	(0.0378)
		[0.0257]	[0.0015]	[0.0008]	[0.0010]	[0.0009]	[0.0013]	[0.0251]	[0.0208]
	SELO	-0.0015	0.0028	0.0014	0.0018	0.0013	0.0017	-0.0031	0.0016
		(0.0292)	(0.0109)	(0.0075)	(0.0087)	(0.0072)	(0.0086)	(0.0284)	(0.0328)
		[0.0258]	[0.0011]	[0.0005]	[0.0007]	[0.0005]	[0.0006]	[0.0252]	[0.0216]

Table 4.

Bias, (SSE) and [ESE] of the parameters in Model IV.

Sample size	Method	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$	$\hat{\alpha}_7$	$\hat{\mu}_\epsilon$
$n = 500$	CLS	-0.0109	0.0191	0.0202	0.0174	0.0186	0.0187	-0.0139	0.0183
		(0.0561)	(0.0288)	(0.0297)	(0.0278)	(0.0286)	(0.0298)	(0.0596)	(0.0448)
		[0.0582]	[0.0530]	[0.0505]	[0.0498]	[0.0506]	[0.0535]	[0.0584]	[0.0459]
	ALASSO	-0.0105	0.0087	0.0086	0.0073	0.0076	0.0097	-0.0178	0.0030
		(0.0572)	(0.0226)	(0.0224)	(0.0187)	(0.0213)	(0.0246)	(0.0604)	(0.0509)
		[0.0486]	[0.0034]	[0.0029]	[0.0024]	[0.0022]	[0.0037]	[0.0478]	[0.0284]
	SCAD	-0.0032	0.0057	0.0047	0.0037	0.0046	0.0065	-0.0099	0.0134
		(0.0600)	(0.0212)	(0.0205)	(0.0142)	(0.0194)	(0.0238)	(0.0613)	(0.0562)
		[0.0495]	[0.0026]	[0.0019]	[0.0016]	[0.0021]	[0.0030]	[0.0491]	[0.0282]
	MCP	-0.0021	0.0048	0.0042	0.0035	0.0041	0.0056	-0.0071	0.0136
		(0.0593)	(0.0191)	(0.0184)	(0.0142)	(0.0184)	(0.0214)	(0.0612)	(0.0550)
		[0.0468]	[0.0017]	[0.0012]	[0.0010]	[0.0013]	[0.0022]	[0.0460]	[0.0250]
	SELO	-0.0066	0.0037	0.0036	0.0023	0.0040	0.0038	-0.0122	0.0092
		(0.0585)	(0.0178)	(0.0174)	(0.0130)	(0.0182)	(0.0190)	(0.0632)	(0.0467)
		[0.0488]	[0.0011]	[0.0013]	[0.0007]	[0.0012]	[0.0012]	[0.0478]	[0.0262]
$n = 1000$	CLS	-0.0042	0.0126	0.0138	0.0124	0.0141	0.0134	-0.0082	0.0114
		(0.0403)	(0.0203)	(0.0198)	(0.0195)	(0.0210)	(0.0209)	(0.0419)	(0.0317)
		[0.0417]	[0.0375]	[0.0356]	[0.0349]	[0.0355]	[0.0379]	[0.0419]	[0.0325]
	ALASSO	-0.0045	0.0070	0.0056	0.0057	0.0054	0.0066	-0.0077	-0.0056
		(0.0408)	(0.0163)	(0.0141)	(0.0141)	(0.0141)	(0.0160)	(0.0414)	(0.0340)
		[0.0361]	[0.0025]	[0.0018]	[0.0017]	[0.0016]	[0.0025]	[0.0361]	[0.0217]
	SCAD	-0.0005	0.0031	0.0029	0.0020	0.0028	0.0035	-0.0035	0.0068
		(0.0419)	(0.0141)	(0.0124)	(0.0108)	(0.0133)	(0.0154)	(0.0410)	(0.0343)
		[0.0356]	[0.0017]	[0.0010]	[0.0010]	[0.0012]	[0.0020]	[0.0358]	[0.0207]
	MCP	-0.0010	0.0034	0.0025	0.0026	0.0024	0.0043	-0.0042	0.0068
		(0.0421)	(0.0132)	(0.0107)	(0.0114)	(0.0114)	(0.0148)	(0.0415)	(0.0341)
		[0.0337]	[0.0012]	[0.0007]	[0.0007]	[0.0007]	[0.0016]	[0.0335]	[0.0181]
	SELO	-0.0027	0.0025	0.0018	0.0016	0.0019	0.0025	-0.0051	0.0011
		(0.0416)	(0.0116)	(0.0094)	(0.0094)	(0.0111)	(0.0123)	(0.0423)	(0.0325)
		[0.0359]	[0.0008]	[0.0005]	[0.0005]	[0.0007]	[0.0009]	[0.0358]	[0.0204]
$n = 2000$	CLS	-0.0030	0.0105	0.0087	0.0095	0.0093	0.0100	-0.0052	0.0069
		(0.0307)	(0.0154)	(0.0139)	(0.0142)	(0.0147)	(0.0152)	(0.0299)	(0.0223)
		[0.0294]	[0.0265]	[0.0250]	[0.0246]	[0.0251]	[0.0268]	[0.0297]	[0.0231]
	ALASSO	-0.0031	0.0047	0.0046	0.0036	0.0042	0.0058	-0.0046	-0.0034
		(0.0289)	(0.0108)	(0.0107)	(0.0092)	(0.0099)	(0.0133)	(0.0291)	(0.0231)
		[0.0262]	[0.0016]	[0.0012]	[0.0010]	[0.0013]	[0.0020]	[0.0263]	[0.0168]
	SCAD	-0.0018	0.0021	0.0021	0.0010	0.0014	0.0034	-0.0034	0.0063
		(0.0292)	(0.0102)	(0.0095)	(0.0069)	(0.0080)	(0.0136)	(0.0292)	(0.0229)
		[0.0252]	[0.0014]	[0.0008]	[0.0006]	[0.0007]	[0.0016]	[0.0252]	[0.0150]
	MCP	-0.0028	0.0029	0.0015	0.0012	0.0019	0.0035	-0.0034	0.0061
		(0.0292)	(0.0097)	(0.0071)	(0.0063)	(0.0079)	(0.0120)	(0.0284)	(0.0225)
		[0.0235]	[0.0010]	[0.0004]	[0.0003]	[0.0005]	[0.0011]	[0.0233]	[0.0130]
	SELO	-0.0029	0.0015	0.0009	0.0009	0.0012	0.0027	-0.0039	0.0026
		(0.0294)	(0.0073)	(0.0062)	(0.0058)	(0.0069)	(0.0112)	(0.0292)	(0.0222)
		[0.0262]	[0.0005]	[0.0003]	[0.0003]	[0.0004]	[0.0010]	[0.0263]	[0.0162]

Table 5.

Bias, (SSE) and [ESE] of the parameters in Model V.

Sample size	Method	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$	$\hat{\alpha}_7$	$\hat{\mu}_\epsilon$
$n = 500$	CLS	-0.0094	0.0175	0.0188	0.0168	0.0168	0.0170	-0.0180	0.0221
		(0.0582)	(0.0271)	(0.0287)	(0.0261)	(0.0255)	(0.0271)	(0.0565)	(0.0469)
		[0.0572]	[0.0513]	[0.0478]	[0.0464]	[0.0473]	[0.0509]	[0.0587]	[0.0466]
	ALASSO	-0.0140	0.0093	0.0076	0.0071	0.0079	0.0097	-0.0178	0.0060
		(0.0606)	(0.0221)	(0.0194)	(0.0186)	(0.0213)	(0.0250)	(0.0600)	(0.0510)
		[0.0478]	[0.0040]	[0.0030]	[0.0026]	[0.0029]	[0.0039]	[0.0500]	[0.0283]
	SCAD	-0.0064	0.0059	0.0049	0.0044	0.0054	0.0071	-0.0096	0.0137
		(0.0613)	(0.0235)	(0.0200)	(0.0181)	(0.0217)	(0.0255)	(0.0604)	(0.0578)
		[0.0493]	[0.0031]	[0.0023]	[0.0020]	[0.0022]	[0.0037]	[0.0514]	[0.0283]
	MCP	-0.0056	0.0052	0.0037	0.0027	0.0052	0.0062	-0.0095	0.0137
		(0.0628)	(0.0199)	(0.0165)	(0.0127)	(0.0213)	(0.0227)	(0.0617)	(0.0575)
		[0.0468]	[0.0019]	[0.0011]	[0.0009]	[0.0016]	[0.0023]	[0.0493]	[0.0244]
	SELO	-0.0090	0.0040	0.0027	0.0026	0.0036	0.0046	-0.0118	0.0082
		(0.0616)	(0.0183)	(0.0159)	(0.0136)	(0.0189)	(0.0202)	(0.0611)	(0.0453)
		[0.0489]	[0.0014]	[0.0009]	[0.0010]	[0.0012]	[0.0016]	[0.0513]	[0.0266]
$n = 1000$	CLS	-0.0048	0.0130	0.0123	0.0126	0.0138	0.0126	-0.0091	0.0113
		(0.0408)	(0.0204)	(0.0198)	(0.0195)	(0.0202)	(0.0206)	(0.0405)	(0.0316)
		[0.0411]	[0.0362]	[0.0336]	[0.0328]	[0.0334]	[0.0361]	[0.0420]	[0.0326]
	ALASSO	-0.0062	0.0058	0.0052	0.0049	0.0051	0.0073	-0.0077	-0.0008
		(0.0403)	(0.0154)	(0.0137)	(0.0129)	(0.0138)	(0.0169)	(0.0414)	(0.0357)
		[0.0357]	[0.0023]	[0.0018]	[0.0019]	[0.0015]	[0.0026]	[0.0371]	[0.0216]
	SCAD	-0.0031	0.0036	0.0027	0.0023	0.0019	0.0040	-0.0039	0.0097
		(0.0406)	(0.0143)	(0.0121)	(0.0110)	(0.0098)	(0.0143)	(0.0416)	(0.0367)
		[0.0355]	[0.0018]	[0.0010]	[0.0013]	[0.0010]	[0.0019]	[0.0368]	[0.0207]
	MCP	-0.0027	0.0028	0.0022	0.0024	0.0018	0.0036	-0.0039	0.0092
		(0.0409)	(0.0118)	(0.0116)	(0.0104)	(0.0085)	(0.0134)	(0.0416)	(0.0364)
		[0.0335]	[0.0011]	[0.0006]	[0.0009]	[0.0005]	[0.0014]	[0.0351]	[0.0181]
	SELO	-0.0039	0.0023	0.0018	0.0015	0.0014	0.0029	-0.0042	0.0018
		(0.0420)	(0.0114)	(0.0101)	(0.0091)	(0.0088)	(0.0130)	(0.0417)	(0.0337)
		[0.0359]	[0.0008]	[0.0007]	[0.0006]	[0.0004]	[0.0010]	[0.0375]	[0.0205]
$n = 2000$	CLS	-0.0034	0.0099	0.0089	0.0095	0.0098	0.0097	-0.0048	0.0050
		(0.0296)	(0.0147)	(0.0137)	(0.0137)	(0.0145)	(0.0148)	(0.0306)	(0.0227)
		[0.0293]	[0.0256]	[0.0237]	[0.0232]	[0.0237]	[0.0257]	[0.0299]	[0.0230]
	ALASSO	-0.0019	0.0039	0.0038	0.0041	0.0034	0.0040	-0.0032	-0.0035
		(0.0297)	(0.0103)	(0.0097)	(0.0101)	(0.0088)	(0.0101)	(0.0325)	(0.0237)
		[0.0262]	[0.0014]	[0.0012]	[0.0011]	[0.0011]	[0.0015]	[0.0273]	[0.0165]
	SCAD	-0.0009	0.0017	0.0022	0.0018	0.0013	0.0016	-0.0019	0.0058
		(0.0298)	(0.0095)	(0.0095)	(0.0081)	(0.0069)	(0.0083)	(0.0323)	(0.0238)
		[0.0250]	[0.0010]	[0.0008]	[0.0007]	[0.0006]	[0.0008]	[0.0261]	[0.0147]
	MCP	-0.0015	0.0018	0.0021	0.0018	0.0015	0.0016	-0.0026	0.0056
		(0.0299)	(0.0082)	(0.0090)	(0.0080)	(0.0072)	(0.0074)	(0.0323)	(0.0233)
		[0.0233]	0.0007]	[0.0007]	[0.0004]	[0.0005]	[0.0006]	[0.0245]	[0.0128]
	SELO	-0.0018	0.0015	0.0014	0.0013	0.0013	0.0011	-0.0027	0.0014
		(0.0305)	(0.0082)	(0.0078)	(0.0073)	(0.0070)	(0.0065)	(0.0326)	(0.0222)
		[0.0260]	[0.0005]	[0.0006]	[0.0004]	[0.0004]	[0.0004]	[0.0271]	[0.0157]

Table 6.

Simulation results for model selection in Model I.

Sample size	Method	MS	CMR	F+	F−
$n = 500$	ALASSO	3.8770	0.2140	0.2359	0.0246
	SCAD	3.4540	0.4990	0.1413	0.0242
	MCP	3.3560	0.4610	0.1416	0.0348
	SELO	3.1610	0.6120	0.0946	0.0343
$n = 1000$	ALASSO	3.8920	0.2660	0.2134	0.0060
	SCAD	3.4190	0.5810	0.1087	0.0055
	MCP	3.3860	0.5670	0.1108	0.0100
	SELO	3.2420	0.7030	0.0729	0.0079
$n = 2000$	ALASSO	3.8190	0.3250	0.1911	0.0009
	SCAD	3.3820	0.6540	0.0916	0.0004
	MCP	3.4180	0.5900	0.1051	0.0013
	SELO	3.2180	0.7790	0.0560	0.0011

Table 7.

Simulation results for model selection in Model II.

Sample size	Method	MS	CMR	F+	F−
$n = 500$	ALASSO	3.8000	0.2430	0.2239	0.0279
	SCAD	3.7370	0.4000	0.1852	0.0200
	MCP	3.4180	0.4690	0.1499	0.0344
	SELO	3.1350	0.5680	0.1059	0.0462
$n = 1000$	ALASSO	3.8520	0.2550	0.2166	0.0144
	SCAD	3.6730	0.4520	0.1595	0.0066
	MCP	3.4270	0.5150	0.1309	0.0189
	SELO	3.2240	0.6520	0.0856	0.0192
$n = 2000$	ALASSO	3.8320	0.2890	0.1985	0.0030
	SCAD	3.5770	0.5120	0.1342	0.0011
	MCP	3.4800	0.5280	0.1217	0.0030
	SELO	3.2590	0.7210	0.0698	0.0038

Table 8.

Simulation results for model selection in Model III.

Sample size	Method	MS	CMR	F+	F−
$n = 500$	ALASSO	3.8240	0.2410	0.2264	0.0263
	SCAD	3.5400	0.4510	0.1570	0.0217
	MCP	3.3700	0.4660	0.1431	0.0350
	SELO	3.1670	0.5720	0.1027	0.0388
$n = 1000$	ALASSO	3.8220	0.2730	0.2078	0.0121
	SCAD	3.5310	0.5060	0.1329	0.0057
	MCP	3.3920	0.5310	0.1202	0.0153
	SELO	3.2390	0.6600	0.0833	0.0153
$n = 2000$	ALASSO	3.8580	0.2940	0.2009	0.0016
	SCAD	3.4880	0.5840	0.1136	0.0006
	MCP	3.4480	0.5720	0.1114	0.0012
	SELO	3.2580	0.7470	0.0647	0.0006

Table 9.

Simulation results for model selection in Model IV.

Sample size	Method	MS	CMR	F+	F−
$n = 500$	ALASSO	3.9090	0.2600	0.2158	0.0056
	SCAD	3.4710	0.5500	0.1210	0.0059
	MCP	3.3360	0.6020	0.1009	0.0118
	SELO	3.2030	0.7350	0.0657	0.0095
$n = 1000$	ALASSO	3.9250	0.2610	0.2130	0.0009
	SCAD	3.3980	0.6300	0.0973	0.0014
	MCP	3.3850	0.6230	0.0971	0.0016
	SELO	3.2150	0.7860	0.0544	0.0006
$n = 2000$	ALASSO	3.9300	0.2900	0.2098	0
	SCAD	3.3990	0.6520	0.0938	0
	MCP	3.3730	0.6520	0.0909	0.0002
	SELO	3.2070	0.7950	0.0515	0

Table 10.

Simulation results for model selection in Model V.

Sample size	Method	MS	CMR	F+	F−
$n = 500$	ALASSO	3.9010	0.2730	0.2126	0.0058
	SCAD	3.4930	0.5300	0.1296	0.0092
	MCP	3.3450	0.6040	0.1007	0.0104
	SELO	3.2070	0.7400	0.0640	0.0081
$n = 1000$	ALASSO	3.8760	0.3000	0.2024	0.0014
	SCAD	3.4130	0.6280	0.0993	0.0004
	MCP	3.3360	0.6710	0.0845	0.0012
	SELO	3.2160	0.7840	0.0543	0.0004
$n = 2000$	ALASSO	3.8130	0.3380	0.1877	0
	SCAD	3.3100	0.7100	0.0753	0
	MCP	3.3020	0.7160	0.0737	0
	SELO	3.1840	0.8190	0.0457	0

Table 11.

Estimated regression coefficients and standard errors(SE)[‡].

Method	ALASSO	SCAD	MCP	SELO	CLS
$\hat{\alpha}_1$	0	0	0	0	0.0135
(SE)	(−)	(−)	(−)	(−)	(0.1089)
$\hat{\alpha}_2$	0	0	0	0	0
(SE)	(−)	(−)	(−)	(−)	(0.0989)
$\hat{\alpha}_3$	0	0	0	0	0.0141
(SE)	(−)	(−)	(−)	(−)	(0.0983)
$\hat{\alpha}_4$	0	0	0	0	0
(SE)	(−)	(−)	(−)	(−)	(0.0985)
$\hat{\alpha}_5$	0	0	0	0	0.0099
(SE)	(−)	(−)	(−)	(−)	(0.1177)
$\hat{\alpha}_6$	0.2509	0.2600	0.2757	0.2570	0.2508
(SE)	(0.1218)	(0.1277)	(0.0949)	(0.1255)	(0.1467)
$\hat{\alpha}_7$	0	0	0	0	0
(SE)	(−)	(−)	(−)	(−)	(0.0908)
$\hat{\alpha}_8$	0	0	0	0	0.0279
(SE)	(−)	(−)	(−)	(−)	(0.0995)
$\hat{\alpha}_9$	0	0	0	0	0
(SE)	(−)	(−)	(−)	(−)	(0.0897)
$\hat{\alpha}_{10}$	0	0	0	0	0.0308
(SE)	(−)	(−)	(−)	(−)	(0.1043)
$\hat{\alpha}_{11}$	0	0	0	0	0.0065
(SE)	(−)	(−)	(−)	(−)	(0.1015)
$\hat{\alpha}_{12}$	0	0	0	0	0.0603
(SE)	(−)	(−)	(−)	(−)	(0.0938)
$\hat{\alpha}_{13}$	0	0	0	0	0.0095
(SE)	(−)	(−)	(−)	(−)	(0.0965)
$\hat{\alpha}_{14}$	0.2369	0.2487	0.1292	0.2439	0.2495
(SE)	(0.0900)	(0.0971)	(0.0364)	(0.0945)	(0.1217)
$\hat{\mu}_\epsilon$	0.4108	0.4540	0.5140	0.4490	0.4327
(SE)	(0.0928)	(0.0981)	(0.0861)	(0.0965)	(0.2298)
λ	0.0029	0.0337	0.2668	0.0062	−
τ	−	3	1.5	0.1	−

‡ The CLS estimates were adjusted by the procedure in Section 4.1.

Table 12.

The ACF and mean of the fitted GINAR model.

Lag	observation	ALASSO	SCAD	MCP	SELO	CLS
1	0.0184	0	0	0	0	0.0395
2	0.0400	0.0181	0.0209	0.0110	0.0198	0.0730
3	0.0162	0	0	0	0	0.0406
4	-0.0141	0.0051	0.0062	0.0032	0.0058	0.0559
5	0.0206	0	0	0	0	0.0369
6	0.2955	0.2669	0.2785	0.2807	0.2745	0.3076
7	-0.0368	0	0	0	0	0.0312
8	0.1082	0.0678	0.0747	0.0393	0.0721	0.1305
9	-0.0682	0	0	0	0	0.0318
10	0.0216	0.0025	0.0031	0.0013	0.0029	0.0671
11	0.0173	0	0	0	0	0.0371
12	0.1136	0.0712	0.0776	0.0788	0.0754	0.1613
13	-0.0173	0	0	0	0	0.0367
14	0.2511	0.2539	0.2681	0.1400	0.2624	0.2989
15	0.0649	0	0	0	0	0.0303
16	0.0281	0.0049	0.0060	0.0017	0.0056	0.0534
17	-0.0249	0	0	0	0	0.0326
18	-0.0325	0	0.0217	0.0221	0.0208	0.0806
$\hat{\mu}$	0.6667	0.8020	0.9240	0.8637	0.8996	1.3224

Table 13.

The PACF of the fitted GINAR model.

Lag	observation	ALASSO	SCAD	MCP	SELO	CLS
1	0.0184	0	0	0	0	0.0395
2	0.0397	0.0181	0.0209	0.0110	0.0198	0.0715
3	0.0148	0	0	0	0	0.0354
4	-0.0162	0.0048	0.0057	0.0030	0.0054	0.0484
5	0.0199	0	0	0	0	0.0283
6	0.2965	0.2668	0.2784	0.2807	0.2744	0.3002
7	-0.0513	0	0	0	0	0.0075
8	0.0926	0.0629	0.0689	0.0362	0.0666	0.0986
9	-0.0852	0	0	0	0	0.0047
10	0.0376	0	0	0	0	0.0344
11	0.0029	0	0	0	0	0.0121
12	0.0387	0	0	0	0	0.0645
13	-0.0063	0	0	0	0	0.0137
14	0.2157	0.2369	0.2487	0.1292	0.2439	0.2495
15	0.1123	0	0	0	0	0
16	-0.0097	0	0	0	0	0
17	-0.0388	0	0	0	0	0
18	-0.0877	0	0	0	0	0

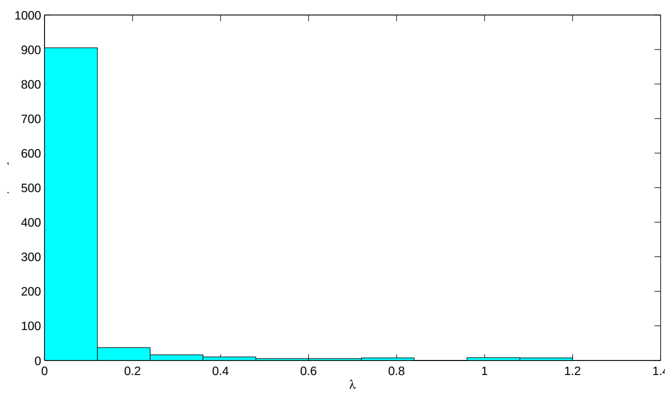

Figure 1. The optimal value of λ in the SELO procedure with $n = 2000$ (Model I).

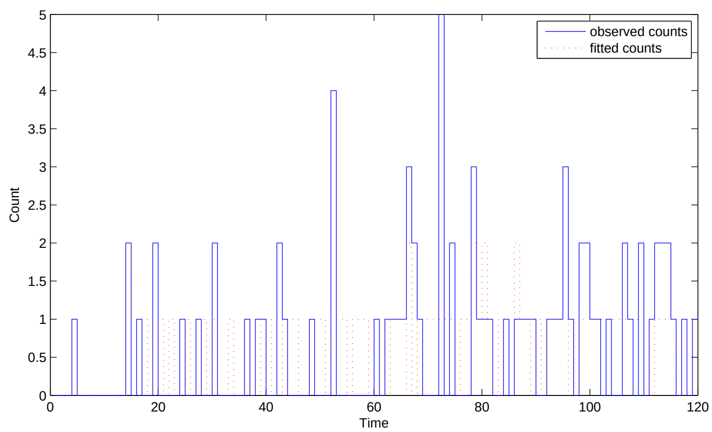

Figure 2. Daily epileptic seizure counts of the patient before treatment.

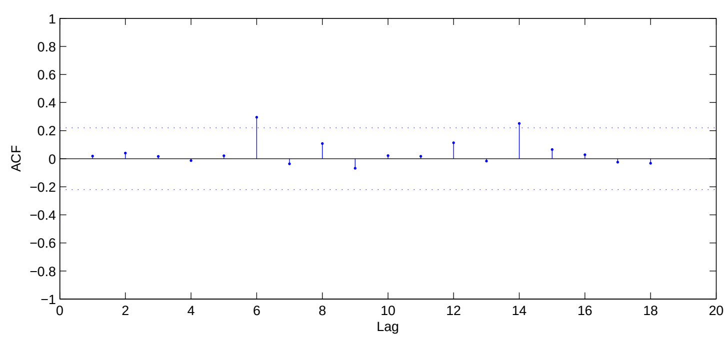

Figure 3. ACF plot of the daily epileptic seizure counts.

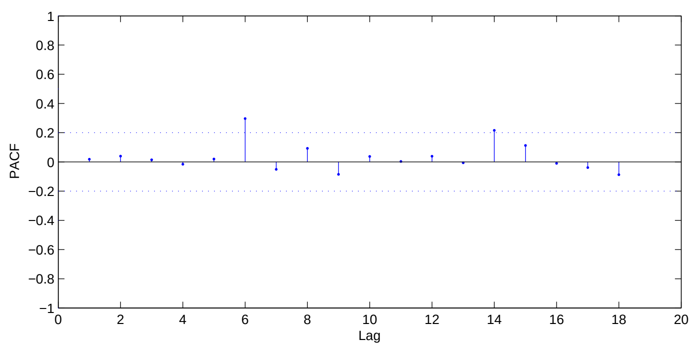

Figure 4. PACF plot of the daily epileptic seizure counts.

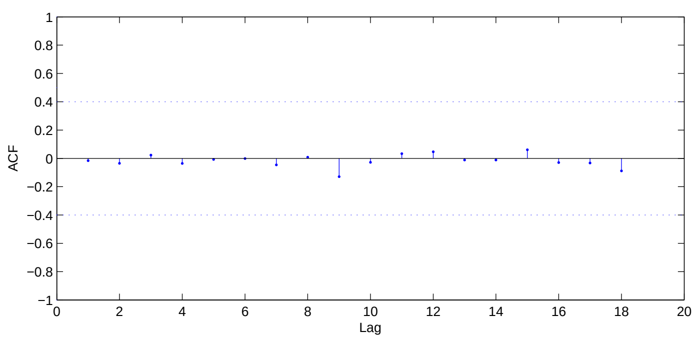

Figure 5. ACF plot of the residuals with the fitted model (SELO).

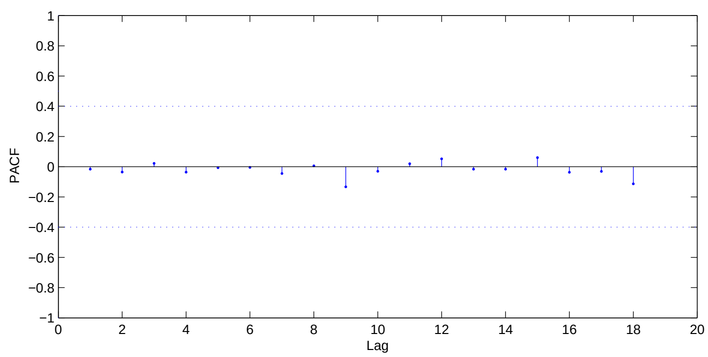

Figure 6. PACF plot of the residuals with the fitted model (SELO).