

Article

Empirical Likelihood Statistical Inference for the BINAR(1) Process

Yan Liu ¹, Yifei Wang ¹ and Bo Shao ^{2,*}

¹ School of Mathematics and Statistics, Changchun University, Changchun 130022, China; liuyan81@ccu.edu.cn (Y.L.)

² Department of Sports Science and Physical Education, Changchun University of Science and Technology, Changchun 130600, China

* Correspondence: 100477@cstu.edu.cn; Tel.: +86-132-3430-0430

Abstract

In this paper, we apply the empirical likelihood (EL) approach to the bivariate first-order integer-valued autoregressive (BINAR(1)) model. The maximum empirical likelihood (MEL) estimator for this process is constructed by the empirical likelihood ratio (ELR) statistic derived from the EL method. We then discuss the asymptotic properties of this estimator. Subsequently, the model parameters are estimated using the MEL method, the conditional least squares method, and the conditional maximum likelihood method. The finite-sample performance of the proposed estimation procedures is examined through Monte Carlo simulation experiments. A real data example is further presented.

Keywords: BINAR(1) process; empirical likelihood; ELR statistic; MEL estimator

MSC: 62M10; 62F10

1. Introduction

Integer-valued time series appear in many applied fields, including epidemiology, criminology, insurance, and economics, where the observations are naturally recorded as counts [1,2]. A common way to model such data is to use integer-valued autoregressive (INAR) processes based on the binomial thinning operator “ $\circ$ ” [3,4]. The bivariate INAR(1) (BINAR(1)) process extends this idea to two dependent count series and is therefore useful when cross-series dependence is present [5]. Existing studies have considered, for example, BINAR(1) models with bivariate Poisson innovations [5,6], zero-truncated Poisson innovations [7], threshold structures, random coefficients, and nonparametric Bayesian formulations [8–10]. Parameter estimation for the BINAR(1) process, however, still deserves further attention, especially when robustness and efficiency are both required.

Conditional least squares (CLS) and conditional maximum likelihood (CML) are two standard estimation methods for the BINAR(1) model [5,11]. CLS is straightforward to compute, but it uses only limited moment information and may lose efficiency. CML can be more efficient under a correctly specified innovation distribution, yet this advantage depends strongly on the validity of the assumed distributional form. In bivariate count models, misspecification, overdispersion, and complicated dependence structures may also make likelihood-based estimation less stable [5,12]. These issues motivate estimation methods that use estimating equations while avoiding a fully specified likelihood.

Empirical likelihood (EL), introduced by Owen [13–15], provides a likelihood-type inference tool without requiring a full parametric likelihood. It has been applied to several



Academic Editor: Paolo Pagnottoni

Received: 7 May 2026

Revised: 17 June 2026

Accepted: 24 June 2026

Published: 27 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

classes of time series models, such as linear autoregressive models, GARCH models, INGARCH models, and INAR(1) models with explanatory variables [16–19]. For integer-valued models, an EL-based unified test for INAR(1) processes was developed in [20], and EL confidence regions for the parameters of the BINAR(1) process were constructed in [21]. These studies show that EL methods can be useful for count time series, but the estimation theory for the BINAR(1) setting remains incomplete.

The construction of an empirical likelihood statistic is only one part of the inference problem. In [21], empirical likelihood confidence regions were obtained for the parameters of the BINAR(1) process, and a maximum empirical likelihood statistic was also introduced. The large-sample properties of the corresponding MEL estimator, however, were not fully developed. Its finite-sample performance and its comparison with classical estimators also require a more detailed investigation. This leaves room for a systematic study of MEL estimation for the BINAR(1) process.

This paper continues the empirical likelihood framework of [21] and studies the maximum empirical likelihood (MEL) estimator for the BINAR(1) process. Starting from a set of unbiased estimating equations, we define the MEL estimator through the associated empirical log-likelihood objective function. We prove the consistency and asymptotic normality of the estimator under suitable regularity conditions. Its finite-sample behavior is then examined by Monte Carlo simulations under bivariate Poisson and bivariate negative binomial innovations, with CLS and CML used as benchmarks. We also apply the method to monthly burglary and robbery counts from Pittsburgh [22].

The contribution of the paper is threefold. First, we give an asymptotic theory for the MEL estimator of the BINAR(1) process, which complements the confidence-region results in [21]. Second, we compare MEL with CLS and CML in simulations, using bias, standard errors, mean squared errors, and prediction accuracy as evaluation criteria. Third, the real data analysis shows how the proposed method can be used in practice and suggests that MEL is a competitive option when the innovation distribution is not easy to specify.

The remainder of this paper is organized as follows. Section 2 introduces the BINAR(1) model and constructs the MEL estimator. Section 3 reports simulation results. Section 4 provides a real data example. Section 5 concludes the paper with a discussion of limitations and future work.

2. MEL Estimation for the BINAR(1) Process

This section first presents the theoretical construction of the MEL estimator and its asymptotic properties, and then describes the numerical implementation.

2.1. Theoretical Derivation

The BINAR(1) model [3] is defined as follows:

$$\mathbf{X}_t = \mathbf{A} \circ \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad t \in \mathbb{Z}, \quad (1)$$

where $\mathbf{A}$ is a diagonal 2×2 matrix with diagonal entries $\alpha_1 \in [0, 1)$ and $\alpha_2 \in [0, 1)$. The operation “ $\mathbf{A} \circ$ ” is understood as matrix multiplication with the binomial thinning operator applied componentwise. The innovation sequence $\{\boldsymbol{\varepsilon}_t\}$ is i.i.d. and nonnegative integer-valued. Its components are independent of the counting sequences used in the thinning operation and are also independent of $\{\mathbf{X}_{t-l}\}$ for $l \geq 1$. Under appropriate distributional conditions, process (1) is strictly stationary [16].

Let $E(\varepsilon_{j,t}) = \lambda_j$, $j = 1, 2$, and $\text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) = \phi$, all of which are finite. Then,

$$E(X_{j,t} | X_{j,t-1}) = \alpha_j X_{j,t-1} + \lambda_j, \quad j = 1, 2,$$

and

$$\text{Cov}(X_{1,t}, X_{2,t} \mid X_{1,t-1}, X_{2,t-1}) = \phi.$$

Assume that the process satisfies the following conditions:

- (a) $\{X_t\}$ is strictly stationary and ergodic;
- (b) $\|X_t\|^6 < \infty$.

Condition (a) is used for the consistency and asymptotic normality arguments, and condition (b) provides the moment control needed in the proof of the following Lemma 5, where the Borel–Cantelli lemma is applied.

The MEL estimation procedure for model (1) follows the construction in [21] and is described below:

According to the CLS method, let

$$Q(\theta) = \sum_{t=1}^n (X_t - E(X_t \mid X_{t-1})) (X_t - E(X_t \mid X_{t-1}))^T,$$

where $\theta = (\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi)^T$. Expanding $Q(\theta)$ yields

$$\begin{cases} Q(\alpha_j, \lambda_j) = \sum_{t=1}^n [X_{j,t} - E(X_{j,t} \mid X_{j,t-1})]^2, & j = 1, 2, \\ Q(\phi) = \sum_{t=1}^n \{[X_{1,t} - \alpha_1 X_{1,t-1} - \lambda_1][X_{2,t} - \alpha_2 X_{2,t-1} - \lambda_2] - \phi\}^2. \end{cases} \quad (2)$$

It follows that

$$-\frac{1}{2} \frac{\partial Q(\theta)}{\partial \theta} \triangleq \sum_{t=1}^n D_t(\theta) = \mathbf{0},$$

where $D_t(\theta) = (D_{1t}(\theta), \dots, D_{5t}(\theta))^T$, and

$$\begin{cases} D_{1t}(\theta) = (X_{1,t} - \alpha_1 X_{1,t-1} - \lambda_1) X_{1,t-1}, \\ D_{2t}(\theta) = X_{1,t} - \alpha_1 X_{1,t-1} - \lambda_1, \\ D_{3t}(\theta) = (X_{2,t} - \alpha_2 X_{2,t-1} - \lambda_2) X_{2,t-1}, \\ D_{4t}(\theta) = X_{2,t} - \alpha_2 X_{2,t-1} - \lambda_2, \\ D_{5t}(\theta) = (X_{1,t} - \alpha_1 X_{1,t-1} - \lambda_1)(X_{2,t} - \alpha_2 X_{2,t-1} - \lambda_2) - \phi. \end{cases}$$

The empirical likelihood ratio (ELR) function is

$$l(\theta) = \max \left\{ \prod_{t=1}^n (np_t) \mid p_t > 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t D_t(\theta) = \mathbf{0} \right\}. \quad (3)$$

Using the method of Lagrange multipliers, let

$$S = \sum_{t=1}^n \log(np_t) - n \mathbf{b}^T \left(\sum_{t=1}^n p_t D_t(\theta) \right) + \gamma \left(\sum_{t=1}^n p_t - 1 \right),$$

where γ and $\mathbf{b} \in \mathbb{R}^5$ are Lagrange multipliers. From

$$\frac{\partial S}{\partial p_t} = \frac{1}{p_t} - n \mathbf{b}^T(\theta) D_t(\theta) + \gamma = 0,$$

we obtain

$$\sum_{t=1}^n p_t \frac{\partial S}{\partial p_t} = n + n \mathbf{b}^T(\theta) \sum_{t=1}^n p_t D_t(\theta) + \gamma \sum_{t=1}^n p_t = n + \gamma.$$

Hence, $n = -\gamma$. Substituting $n = -\gamma$ into $\frac{\partial S}{\partial p_t} = 0$ gives

$$p_t = \frac{1}{n} \frac{1}{1 + \mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta})}, \quad t = 1, \dots, n,$$

where $\mathbf{b}(\boldsymbol{\theta})$ is determined by

$$\frac{1}{n} \sum_{t=1}^n \frac{\mathbf{D}_t(\boldsymbol{\theta})}{1 + \mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta})} = \mathbf{0}.$$

Therefore, the log-ELR statistic is

$$L(\boldsymbol{\theta}) = -2 \log l(\boldsymbol{\theta}) = 2 \sum_{t=1}^n \log(1 + \mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta})). \quad (4)$$

The log empirical likelihood function is

$$L_E(\boldsymbol{\theta}) = \sum_{t=1}^n \log(1 + \mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta})). \quad (5)$$

Finally, the MEL estimator is defined by

$$\hat{\boldsymbol{\theta}}^{MEL} \triangleq \arg \inf_{\boldsymbol{\theta}} L_E(\boldsymbol{\theta}), \quad (6)$$

where ‘arg inf’ denotes the value of $\boldsymbol{\theta}$ that minimizes the function $L_E(\boldsymbol{\theta})$.

Let $\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}_0, n^{-1/3}) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < n^{-1/3}\}$, where $\boldsymbol{\theta}_0 = (\alpha_{10}, \alpha_{20}, \lambda_{10}, \lambda_{20}, \phi_0)^T$ is the true value of $\boldsymbol{\theta}$.

According to Lemma 1 in [21], suppose that $\{\mathbf{X}_t\}$ satisfies conditions (a) and (b). We have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{D}_t(\boldsymbol{\theta}) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega}(\boldsymbol{\theta})), n \rightarrow \infty,$$

where $\boldsymbol{\Omega}(\boldsymbol{\theta}) = (\sigma_{ij})_{5 \times 5}$ with

$$\sigma_{ij} = \begin{cases} \sigma_j^2, & i = j, \\ \sigma_{ij}, & i \neq j. \end{cases}$$

To establish the limiting properties of $\hat{\boldsymbol{\theta}}^{MEL}$, we first introduce the following lemmas.

Lemma 1. Let $\mathbf{D}_n^*(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \mathbf{D}_t(\boldsymbol{\theta})$. Then, as $n \rightarrow \infty$, $\sqrt{n} \mathbf{D}_n^*(\boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega}(\boldsymbol{\theta}_0))$, where $\mathbf{D}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$ are given in Lemma 1 with [21].

Lemma 2. With the conditions of (a) and (b), $\mathbf{D}_n^*(\boldsymbol{\theta}_0) = O(n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}})$, a.s.

Lemma 3. If $\boldsymbol{\theta} \in \mathbf{G}$, then $\sup_{\boldsymbol{\theta} \in \mathbf{G}} \|\mathbf{D}_n^*(\boldsymbol{\theta})\| = O_p(n^{-\frac{1}{3}})$, where $\mathbf{D}_n^*(\boldsymbol{\theta})$ is given in Lemma 1.

Lemma 4. If $\boldsymbol{\theta} \in \mathbf{G}$, then $\sup_{\boldsymbol{\theta} \in \mathbf{G}} \|\boldsymbol{\Omega}_n(\boldsymbol{\theta}) - \boldsymbol{\Omega}(\boldsymbol{\theta}_0)\| = o_p(1)$, $\sup_{\boldsymbol{\theta} \in \mathbf{G}} \|\boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}_0)\| = o_p(1)$, where $\boldsymbol{\Omega}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$ are given in Lemma 1 [21].

Lemma 5. If $\boldsymbol{\theta} \in \mathbf{G}$, then (i) $\max_{1 \leq t \leq n} \sup_{\boldsymbol{\theta} \in \mathbf{G}} \|\mathbf{D}_t(\boldsymbol{\theta})\| = o(n^{\frac{1}{3}})$, a.s.; (ii) $\max_{1 \leq t \leq n} \|\mathbf{D}_t(\boldsymbol{\theta}_0)\| = o(n^{\frac{1}{2}})$, a.s.

Lemma 6. (i) If $\theta \in G$, then $\sup_{\theta \in G} \|\mathbf{b}(\theta)\| = O_p(n^{-\frac{1}{3}})$, where $\mathbf{b}(\theta)$ is given in (3); (ii) $\|\mathbf{b}(\theta_0)\| = O_p(n^{-\frac{1}{2}})$.

Lemma 7. If $\theta \in G$, then $\sup_{\theta \in G} \|\mathbf{b}(\theta) - \Omega_n^{-1}(\theta) D_n^*(\theta)\| = o_p(n^{-\frac{1}{3}})$.

The following theorem gives the consistency and asymptotic normality of the MEL estimator.

Theorem 1. With the conditions (a) and (b), as $n \rightarrow \infty$, $\hat{\theta}^{MEL}$ is a consistent estimator of θ_0 , and

$$\sqrt{n}(\hat{\theta}^{MEL} - \theta_0) \xrightarrow{L} N(\mathbf{0}, V^{-1} \Omega(\theta_0) V^{-1}),$$

where $\Omega(\theta_0)$ is given in Lemma 1 in [21], and V is a nonsingular matrix (see [23]) with $V^{-1} = -\frac{1}{n} \sum_{t=1}^n \left(\frac{\partial D_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right)$ for model (1). (The calculation result of V is shown in the proof.)

2.2. Numerical Implementation

To compute the MEL estimator $\hat{\theta}^{MEL}$, we use a numerical optimization procedure. The implementation involves the choice of initial values, the treatment of parameter constraints, the optimization algorithm, and the solution of the Lagrange multiplier vector.

Initial values: The CLS estimates of θ are used as starting values because they are easy to compute and have standard consistency properties.

Parameter constraints: The autoregressive parameters satisfy $\alpha_1, \alpha_2 \in [0, 1]$; the innovation means satisfy $\lambda_1, \lambda_2 > 0$; and the covariance ϕ satisfies $0 < \phi < \min(\lambda_1, \lambda_2)$ for both the bivariate Poisson and bivariate negative binomial models. All constraints are enforced via reparameterization to an unconstrained scale: $\alpha_i = \text{logit}^{-1}(\tilde{\alpha}_i)$, $\lambda_i = \exp(\tilde{\lambda}_i)$, and $\phi = \tanh(\tilde{\phi})$.

Optimization algorithm: The objective function is minimized by the quasi-Newton BFGS algorithm with finite-difference gradients, implemented through the R function `optim`. Convergence is declared when the relative change in the objective function is below 10^{-8} or when the gradient norm is below 10^{-6} .

Lagrange multiplier: For a given θ , the vector $\mathbf{b}(\theta)$ is obtained by solving Equation (3) using the Newton–Raphson method, with starting value $\mathbf{b} = \mathbf{0}$.

3. Simulation Study

Monte Carlo simulations are used to examine the finite-sample performance of the MEL estimator and to compare it with CLS and CML. Section 3.1 gives the simulation design, including innovation generation, parameter settings, sample sizes, and reproducibility settings. Section 3.2 reports the results for parameter estimation and prediction accuracy.

3.1. Simulation Design

We conduct Monte Carlo experiments to assess the finite-sample behavior of the MEL estimator and to compare it with conditional least squares (CLS) and conditional maximum likelihood (CML).

For model (1), the Yule–Walker estimator and the CLS estimator are equivalent [19]. We therefore use CLS and CML as benchmark methods for comparison with $\hat{\theta}^{MEL}$. In (2),

letting $\frac{1}{2} \frac{\partial Q(\theta)}{\partial \theta} \triangleq \sum_{t=1}^n D_t(\theta) = \mathbf{0}$, we can obtain the CLS estimator $\hat{\theta}^{CLS}$ of θ ; $\hat{\theta}^{CML}$ is obtained by maximizing the following conditional likelihood function of model (1),

$$L(\theta|x) = \prod_{t=1}^n f(x_t|x_{t-1}, \theta) = \sum_{k=0}^{\infty} \sum_{s=0}^{\infty} f_1(x_{1t} - k) f_2(x_{2t} - s) f_3(k, s),$$

where f_1 , f_2 , and f_3 are the conditional distributions of x_{1t} , x_{2t} , and ε_t , respectively.

Innovation distributions: We consider two bivariate innovation distributions:

Model I: bivariate Poisson (BP) with parameters $(\lambda_1, \lambda_2, \phi)$ [24];

Model II: bivariate negative binomial (BNB) with mean parameters $(\lambda_1, \lambda_2, \beta)$ [25].

For Model II, $\beta > 0$ is the dispersion parameter, and the covariance is implicitly given by $\phi = \lambda_1 \lambda_2 / \beta$.

Parameter settings: The true parameter vector is $\theta = (\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi)^T$ for Model I and $\theta = (\alpha_1, \alpha_2, \lambda_1, \lambda_2, \beta)^T$ for Model II. For Model I, six parameter settings are considered: $(0.1, 0.1, 2, 2, 1)$, $(0.1, 0.1, 2, 4, \frac{1}{8})$, $(0.3, 0.3, 2, 2, 1)$, $(0.3, 0.3, 2, 4, 1)$, $(0.1, 0.3, 2, 2, 1)$, and $(0.1, 0.3, 2, 4, 1)$. For Model II, the corresponding settings are $(0.1, 0.1, 2, 2, \frac{1}{4})$, $(0.1, 0.1, 2, 4, \frac{1}{8})$, $(0.3, 0.3, 2, 2, \frac{1}{4})$, $(0.3, 0.3, 2, 4, \frac{1}{8})$, $(0.1, 0.3, 2, 2, \frac{1}{4})$, and $(0.1, 0.3, 2, 4, \frac{1}{8})$. Sample paths for $n = 100$ are displayed in Figure A1.

Simulation settings: For each parameter configuration, we generate $N = 2000$ independent replications. The sample sizes are $n = 100, 300, 500$.

Data generation: The bivariate Poisson innovations are generated as described in [24]. The bivariate negative binomial innovations follow the approach of [25].

Burn-in and initial values: To reduce the effect of initial values, we generate $n + 500$ observations for each replication and discard the first 500 as burn-in. The initial values (X_{10}, X_{20}) are set to the stationary mean $\left(\frac{\lambda_1}{1-\alpha_1}, \frac{\lambda_2}{1-\alpha_2}\right)$.

Random seeds and reproducibility: Within each replication, the same simulated sample is used for CLS, CML, and MEL, so that differences among the methods are not driven by Monte Carlo variation.

Evaluation criteria: For each replication and each estimation method, we compute the parameter estimates and the one-step-ahead predictions. The performance of the parameter estimators is evaluated by the bias, standard error (SE), and mean squared error (MSE) over all replications. To assess prediction accuracy, we compute the mean squared prediction error (MSPE) and mean absolute error (MAE).

For model (1), the conditional expectation satisfies $E(\alpha_i \circ X_{i,t} | X_{i,t}) = \alpha_i X_{i,t}$, $i = 1, 2$. Therefore, for the r th replication and estimation method $m \in \{\text{CLS}, \text{CML}, \text{MEL}\}$, the one-step-ahead forecast is defined as $\hat{X}_{i,t+1|t}^{(r,m)} = \hat{\alpha}_i^{(r,m)} X_{i,t}^{(r)} + \hat{\lambda}_i^{(r,m)}$, $i = 1, 2$, $t = 1, \dots, n-1$.

The MSPE and MAE for each component are then given by

$$\text{MSPE}_i^{(m)} = \frac{1}{N} \sum_{r=1}^N \left[\frac{1}{n-1} \sum_{t=1}^{n-1} \left(X_{i,t+1}^{(r)} - \hat{X}_{i,t+1|t}^{(r,m)} \right)^2 \right], \quad i = 1, 2,$$

and

$$\text{MAE}_i^{(m)} = \frac{1}{N} \sum_{r=1}^N \left[\frac{1}{n-1} \sum_{t=1}^{n-1} \left| X_{i,t+1}^{(r)} - \hat{X}_{i,t+1|t}^{(r,m)} \right| \right], \quad i = 1, 2.$$

Their average values are defined as

$$\text{AMSPE}^{(m)} = \frac{\text{MSPE}_1^{(m)} + \text{MSPE}_2^{(m)}}{2}, \quad \text{AMAE}^{(m)} = \frac{\text{MAE}_1^{(m)} + \text{MAE}_2^{(m)}}{2}.$$

In addition, Quantile–Quantile (Q–Q) plots are used to check the approximate normality of selected MEL estimates; representative plots are shown in Figure A2.

3.2. Simulation Results

Tables A1–A6 summarize the bias, standard error (SE), and mean squared error (MSE) of the estimators for the two innovation models. The biases are generally small and become smaller as the sample size increases. For Model I with bivariate Poisson innovations, the results for $n = 100$ in Table A1 show that MEL has slightly smaller MSE than CML for most parameters. For instance, under $(\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi) = (0.1, 0.1, 2, 2, 1)$, the MSE values of $\hat{\alpha}_1$ are 0.0099, 0.0078, and 0.0076 for CLS, CML, and MEL, respectively. For $\hat{\phi}$ under the same setting, the corresponding MSE values are 0.1350, 0.0882, and 0.0845. Thus, MEL improves clearly over CLS and only slightly over CML in this small-sample case.

The difference between MEL and CML becomes less visible for larger samples. When $n = 500$ (Table A3), the MSE values of the two methods are almost the same for most parameters. For example, for $\hat{\alpha}_1$ under $(0.1, 0.1, 2, 2, 1)$, both CML and MEL have an MSE close to 0.0010. The results for Model II with bivariate negative binomial innovations (Tables A4–A6) show the same pattern. MEL often has a small MSE advantage over CML when the sample size is small or moderate, but this advantage decreases as n grows.

Table A7 reports the one-step-ahead average mean squared prediction error (AMSPE) and average mean absolute error (AMAE) for Model I and Model II when $n = 500$. For Model I, the AMSPE values are close to each other: 2.5351 for CLS, 2.5228 for CML, and 2.5233 for MEL. The corresponding AMAE values are 1.2427, 1.2338, and 1.2329. For Model II, the AMSPE values are 3.4220, 3.4055, and 3.4069 for CLS, CML, and MEL, respectively, while the AMAE values are 1.4267, 1.4134, and 1.4125. These results indicate that the three methods lead to almost the same prediction errors. MEL has only a negligible advantage over CML and a small advantage over CLS. This is consistent with the estimation results, since the main gains of MEL occur for the dependence parameter ϕ or β , which does not enter the point forecast directly, whereas the improvements for α_i and λ_i are quite small.

Taken together, the simulation results suggest that MEL is a competitive alternative to CLS and CML. Relative to CLS, it usually gives smaller SE and MSE, particularly in small-to-moderate samples. Relative to CML, its estimation accuracy is comparable and sometimes slightly better, mainly for dependence-related parameters and smaller sample sizes. The advantage becomes weaker as the sample size increases. For one-step-ahead point prediction, the three methods are almost indistinguishable at $n = 500$. The main appeal of MEL is that it avoids a fully specified parametric likelihood while still performing close to CML when the assumed innovation distribution is correct.

4. Real Data Example

In this section, process (1) and the MEL method are applied to crime count data. The data come from the study “Development of Crime Forecasting and Mapping Systems for Use by Police in Pittsburgh, Pennsylvania, and Rochester, New York, 1990–2001” (ICPSR 4545), collected by Jacqueline Cohen and Wilpen L. Gorr and distributed by ICPSR [22]. We use the monthly counts of burglary (BURGLARY) and robbery (ROBBERY) in tract 502 of Pittsburgh, USA, from January 1990 to December 2001, giving 144 observations for each series. The sample means are 1.4653 and 1.1181, and the sample variances are 2.2925 and 1.4055, respectively. Figures A3–A5 show the sample paths, autocorrelation functions (ACFs), partial autocorrelation functions (PACFs), and cross-correlation functions (CCFs). The plots suggest that both series are approximately stationary and have AR(1)-type serial dependence. The CCF plots also show dependence between the two series, which is plausible because burglary and robbery in the same tract may be related in practice.

We fit this real data set to the following four models:

- (i) BINAR(1)-I: the BINAR(1) process with bivariate Poisson innovations.
- (ii) BINAR(1)-II: the BINAR(1) process with bivariate negative binomial innovations.

- (iii) IINAR(1)-I: two INAR(1) processes with two independent Poisson innovations.
- (iv) IINAR(1)-II: two INAR(1) processes with two independent negative binomial innovations.

CML and MEL are used to estimate the model parameters and their standard errors. Model fit is compared by the Akaike information criterion (AIC) and Bayesian information criterion (BIC), and the results are given in Table A8. Among the four candidate models, the BINAR(1)-II model estimated by MEL has the smallest AIC and BIC values. This suggests a better fit for the crime data, although the differences are not large. MEL also gives slightly smaller standard errors than CML for most parameters. Figure A7 compares the observed series with the one-step-ahead fitted values, and the fitted values follow the main movements of the data reasonably well.

The adequacy of the selected BINAR(1)-II model with MEL estimation is further checked using residual diagnostics. Figure A6 gives the standardized residuals against fitted values and the residual ACFs for the two series. No clear systematic pattern is visible in the residual plots. In addition, all residual autocorrelations fall within the 95% confidence bands. The Ljung–Box test at lag 10 gives p -values of 0.47 for burglary and 0.71 for robbery, indicating no significant remaining autocorrelation. These diagnostics support the use of the BINAR(1)-II model with MEL for these two crime count series.

5. Conclusions and Discussion

This paper considered empirical likelihood (EL) estimation for the BINAR(1) process. Based on unbiased estimating equations, the maximum empirical likelihood (MEL) estimator was obtained from the empirical likelihood ratio statistic. Under mild regularity conditions, we proved the consistency and asymptotic normality of the estimator. Monte Carlo simulations under bivariate Poisson and bivariate negative binomial innovations were used to study its finite-sample behavior. The results show that MEL is competitive with CLS and CML, and it often gives smaller standard errors and mean squared errors. The one-step-ahead prediction results also show comparable performance in terms of average mean squared prediction error (AMSPE) and average mean absolute error (AMAE).

The empirical application to monthly burglary and robbery counts from Pittsburgh illustrates the practical use of the method. According to AIC and BIC, the BINAR(1)-II model estimated by MEL gives the best fit among the candidate models considered. The residual plots, residual autocorrelation checks, and observed-versus-fitted comparisons also support the adequacy of the selected model. The standardized residuals do not show a clear pattern, and the Ljung–Box test does not indicate remaining serial correlation. The fitted conditional means reproduce the main level and movement of the two crime series, although some extreme observations are still underestimated.

Some limitations remain. First, the simulations used only two innovation distributions: bivariate Poisson and bivariate negative binomial. Other innovation structures, including zero-inflated, generalized Poisson, and more strongly overdispersed distributions, were not considered. Second, the study was restricted to the first-order bivariate INAR model. Higher-order BINAR(p) models, threshold versions, and models with structural changes are left for future work. Third, the empirical analysis used only one crime dataset. Applications to insurance claims, hospital admissions, financial transaction counts, or other bivariate count data would help assess the generality of the method. Fourth, although residual diagnostics and fitted-versus-observed plots were included, further out-of-sample forecasting and cross-validation studies would provide additional evidence. Finally, the fitted BINAR(1)-II model tends to underestimate some large crime counts, suggesting that occasional spikes or stronger overdispersion may require richer model structures.

Several extensions are possible. The empirical likelihood approach may be developed for BINAR(p) processes with $p > 1$, or for BINAR(1) models with covariates through a profile empirical likelihood construction. It could also be adapted to bivariate INAR models with time-varying parameters, structural breaks, or threshold effects. In view of the underestimation of some large counts in the real data example, zero-inflated, hurdle, or threshold bivariate INAR models are also worth considering. Another possible direction is to extend MEL estimation to other multivariate count time series models, such as multivariate INGARCH models.

Author Contributions: Conceptualization, Y.L.; Methodology, Y.L.; Software, Y.L.; Validation, Y.L.; Formal analysis, Y.L.; Investigation, Y.L. and B.S.; Resources, B.S.; Data curation, Y.W.; Writing—original draft, Y.L. and Y.W.; Writing—review and editing, Y.L. and Y.W.; Visualization, Y.W.; Supervision, B.S.; Project administration, B.S.; Funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported and funded by the Scientific Research Project of Jilin Provincial Department of Education (No. JJKH20251088KJ).

Data Availability Statement: The datasets used and analyzed during the current study are publicly available on ICPSR [22].

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

Appendix A

Proof of Lemma 1. The proof follows directly from Lemma 1 of [21] by applying the properties of the multivariate normal distribution. No further argument is required. $\square$

Proof of Lemma 2. By Lemma 1 in [21], it follows that $E(D_{1t}|\mathcal{F}_{t-1}) = 0$. Hence, $\{D_{1t}\}$ is a strictly stationary and ergodic martingale difference sequence, and $E(D_{1t}^T D_{1t}) < \infty$. Therefore, by the law of the iterated logarithm,

$$\limsup_{n \rightarrow \infty} n^{-\frac{1}{2}} \sum_{t=1}^n \frac{D_{1t}(\theta_0)}{\sqrt{E(D_{1t}^2) \log_2^{(nE(D_{1t}^2))}}} = \sqrt{2}, \text{ a.s.}$$

Consequently, $\frac{1}{n} \sum_{t=1}^n D_{1t}(\theta_0) = O(n^{-\frac{1}{2}}(\log_2^n)^{\frac{1}{2}}), \text{ a.s.}$

The same argument applies to $\frac{1}{n} \sum_{t=1}^n D_{jt}(\theta_0) = O(n^{-\frac{1}{2}}(\log_2^n)^{\frac{1}{2}}), \text{ a.s.}, j = 2, 3, 4, 5$, yielding the same order of magnitude almost surely. Combining these results, we obtain $D_n^*(\theta_0) = O(n^{-\frac{1}{2}}(\log_2^n)^{\frac{1}{2}}), \text{ a.s.}$, which completes the proof. $\square$

Proof of Lemma 3. We prove the result componentwise.

Recall that $D_{jt}(\theta) = (X_{i,t} - \alpha_i X_{i,t-1} - \lambda_i) X_{i,t-1}, i = 1, 2, j = 1, 3$, then

$$\left| \frac{1}{n} \sum_{t=1}^n D_{jt}(\theta) - \frac{1}{n} \sum_{t=1}^n D_{jt}(\theta_0) \right| \leq \left| \frac{1}{n} \sum_{t=1}^n X_{j,t-1}^2 \right| |\alpha_j - \alpha_{j0}| + \left| \frac{1}{n} \sum_{t=1}^n X_{j,t-1} \right| |\lambda_j - \lambda_{j0}|.$$

By the ergodic theorem for strictly stationary sequences [26], we have

$$\left| \frac{1}{n} \sum_{t=1}^n D_{jt}(\theta) - \frac{1}{n} \sum_{t=1}^n D_{jt}(\theta_0) \right| \leq B(|\alpha_i - \alpha_{i0}| + |\lambda_i - \lambda_{i0}|) = O(n^{-\frac{1}{3}}),$$

where $i = 1, 2, j = 1, 3$, and $B > 0$ is a constant.

Therefore,

$$\begin{aligned} \left| \frac{1}{n} \sum_{t=1}^n D_{jt}(\boldsymbol{\theta}) \right| &\leq \left| \frac{1}{n} \sum_{t=1}^n D_{jt}(\boldsymbol{\theta}) - \frac{1}{n} \sum_{t=1}^n D_{jt}(\boldsymbol{\theta}_0) \right| + \frac{1}{n} \sum_{t=1}^n D_{jt}(\boldsymbol{\theta}_0) \\ &= O(n^{-\frac{1}{3}}) + O_p(n^{-\frac{1}{2}}) \\ &= O_p(n^{-\frac{1}{3}}). \end{aligned}$$

$$\text{Thus, } \sup_{\boldsymbol{\theta} \in G} \left| \frac{1}{n} \sum_{t=1}^n D_{jt}(\boldsymbol{\theta}) \right| = O_p(n^{-\frac{1}{3}}), \quad j = 1, 3.$$

$$\text{Similarly, by the preceding argument, } \sup_{\boldsymbol{\theta} \in G} \left| \frac{1}{n} \sum_{t=1}^n D_{jt}(\boldsymbol{\theta}) \right| = O_p(n^{-\frac{1}{3}}), \quad j = 2, 4.$$

Note that

$$D_{5t}(\boldsymbol{\theta}) = (X_{1,t} - \alpha_1 X_{1,t-1} - \lambda_1)(X_{2,t} - \alpha_2 X_{2,t-1} - \lambda_2) - \phi.$$

We obtain

$$\begin{aligned} &\left| \frac{1}{n} \sum_{t=1}^n D_{5t}(\boldsymbol{\theta}) - \frac{1}{n} \sum_{t=1}^n D_{5t}(\boldsymbol{\theta}_0) \right| \\ &\leq \left| \frac{1}{n} \sum_{t=1}^n X_{1,t-1} X_{2t} \right| |\alpha_1 - \alpha_{10}| + \left| \frac{1}{n} \sum_{t=1}^n X_{1t} X_{2,t-1} \right| |\alpha_2 - \alpha_{20}| \\ &+ \left| \frac{1}{n} \sum_{t=1}^n X_{2t} \right| |\lambda_1 - \lambda_{10}| + \left| \frac{1}{n} \sum_{t=1}^n X_{1t} \right| |\lambda_2 - \lambda_{20}| + |\phi - \phi_0| \\ &+ \left| \frac{1}{n} \sum_{t=1}^n X_{1,t-1} X_{2,t-1} \right| |\alpha_1 \alpha_2 - \alpha_{10} \alpha_{20}| + \left| \frac{1}{n} \sum_{t=1}^n X_{1,t-1} \right| |\alpha_1 \lambda_2 - \alpha_{10} \lambda_{20}| \\ &+ \left| \frac{1}{n} \sum_{t=1}^n X_{2,t-1} \right| |\alpha_2 \lambda_1 - \alpha_{20} \lambda_{10}| + |\lambda_1 \lambda_2 - \lambda_{10} \lambda_{20}|. \end{aligned}$$

By the ergodic theorem for strictly stationary sequences [26], it follows that

$$\begin{aligned} &\left| \frac{1}{n} \sum_{t=1}^n D_{5t}(\boldsymbol{\theta}) - \frac{1}{n} \sum_{t=1}^n D_{5t}(\boldsymbol{\theta}_0) \right| \\ &\leq C(|\alpha_1 - \alpha_{10}| + |\alpha_2 - \alpha_{20}| + |\lambda_1 - \lambda_{10}| + |\lambda_2 - \lambda_{20}| + |\phi - \phi_0| \\ &+ |\alpha_1 \alpha_2 - \alpha_{10} \alpha_{20}| + |\alpha_1 \lambda_2 - \alpha_{10} \lambda_{20}| + |\alpha_2 \lambda_1 - \alpha_{20} \lambda_{10}| + |\lambda_1 \lambda_2 - \lambda_{10} \lambda_{20}|) \\ &= O(n^{-\frac{1}{3}}), \end{aligned}$$

where $C > 0$ is a constant.

Therefore,

$$\begin{aligned} \left| \frac{1}{n} \sum_{t=1}^n D_{5t}(\boldsymbol{\theta}) \right| &\leq \left| \frac{1}{n} \sum_{t=1}^n D_{5t}(\boldsymbol{\theta}) - \frac{1}{n} \sum_{t=1}^n D_{5t}(\boldsymbol{\theta}_0) \right| + \left| \frac{1}{n} \sum_{t=1}^n D_{5t}(\boldsymbol{\theta}_0) \right| \\ &\leq O(n^{-\frac{1}{3}}) + O_p(n^{-\frac{1}{2}}) \\ &= O_p(n^{-\frac{1}{3}}). \end{aligned}$$

$$\text{Therefore, } \sup_{\boldsymbol{\theta} \in G} \left| \frac{1}{n} \sum_{t=1}^n D_{5t}(\boldsymbol{\theta}) \right| = O_p(n^{-\frac{1}{3}}).$$

Combining the results above yields

$$\sup_{\theta \in G} \|D_n^*(\theta)\| = O_p(n^{-\frac{1}{3}}).$$

□

Proof of Lemma 4. Step 1. We first show that $\sup_{\theta \in G} \|\Omega_n(\theta) - \Omega(\theta_0)\| = o_p(1)$.

By Lemma 1 in [21],

$$\begin{aligned} \|\Omega_n(\theta) - \Omega(\theta_0)\| &= \|\Omega_n(\theta) - \Omega_n(\theta_0)\| + \|\Omega_n(\theta_0) - \Omega(\theta_0)\| \\ &= O(n^{-\frac{1}{3}}) + o_p(1) \\ &= o_p(1). \end{aligned}$$

Therefore, $\sup_{\theta \in G} \|\Omega_n(\theta) - \Omega(\theta_0)\| = o_p(1)$.

Step 2. We next show that $\sup_{\theta \in G} \|\Omega_n^{-1}(\theta) - \Omega^{-1}(\theta_0)\| = o_p(1)$.

By Step 1, $\Omega_n(\theta) \rightarrow \Omega(\theta_0)$, *a.s.* Moreover, by Slutsky's theorem, as $n \rightarrow \infty$, $|\Omega_n(\theta)| \rightarrow |\Omega(\theta_0)|$ and $\Omega_n^*(\theta) \rightarrow \Omega^*(\theta_0)$, where Ω_n^* and Ω^* denote the adjugate matrices of Ω_n and Ω , respectively. Hence, $\Omega_n^*(\theta)|\Omega_n(\theta)|^{-1} \rightarrow \Omega^*(\theta_0)|\Omega(\theta_0)|^{-1}$, *a.s.* Therefore,

$$\Omega_n^{-1}(\theta) \rightarrow \Omega^{-1}(\theta_0), \text{ a.s.}$$

The uniform convergence over the compact neighborhood G follows directly. Hence,

$$\sup_{\theta \in G} \|\Omega_n^{-1}(\theta) - \Omega^{-1}(\theta_0)\| = o_p(1).$$

□

Proof of Lemma 5. (i) We first show that $\max_{1 \leq t \leq n} \sup_{\theta \in G} \|D_t(\theta)\| = o(n^{\frac{1}{3}})$, *a.s.*

From $|D_{1t}(\theta)|^3 = |(X_{1t} - \alpha_1 X_{1,t-1} - \lambda_1 X_{1,t-1})|^3$, under condition (b), we have $E \left[\sup_{\theta \in G} |D_{1t}(\theta)| \right]^3 < \infty$. Similarly, we can prove that $E \left[\sup_{\theta \in G} |D_{jt}(\theta)| \right]^3 < \infty$, $j = 2, 3, 4, 5$.

Therefore, $E \left[\sup_{\theta \in G} \|D_t(\theta)\| \right]^3 < \infty$.

For any constant $A > 0$, it follows that

$$\begin{aligned} & E \left[\sup_{\theta \in G} \|D_t(\theta)\| \right]^3 \\ &= \int_0^\infty P \left\{ \left[\sup_{\theta \in G} |D_{jt}(\theta)| \right]^3 > s \right\} ds \\ &\geq \sum_{n=1}^\infty \int_{(n-1)A^3}^{nA^3} P \left\{ \left[\sup_{\theta \in G} |D_{jt}(\theta)| \right]^3 > nA^3 \right\} ds \\ &= \sum_{n=1}^\infty P \left\{ \left[\sup_{\theta \in G} |D_{jt}(\theta)| \right]^3 > nA^3 \right\} A^3. \end{aligned}$$

By the Borel–Cantelli lemma, $P(\sup_{\theta \in G} \|D_n(\theta)\| > n^{\frac{1}{3}}A, i.o.) = 0$. Therefore,

$$\sup_{\theta \in G} \|D_n(\theta)\| \leq n^{\frac{1}{3}}A, a.s. \quad (A1)$$

Let Q_r be a null set, where r is arbitrary. From (A1), when $\omega \in Q_r^c$, we have $\sup_{\theta \in G} \|D_n(\theta)\| < n^{\frac{1}{3}}r^{-1}$. Let $Q = \bigcup_{r=1}^{\infty} Q_r$. Then, as $n \rightarrow \infty$ and $\omega \in Q^c$, $\sup_{\theta \in G} \|D_n(\theta)\| n^{-\frac{1}{3}} \rightarrow 0$. That is, there exists $N_1(\omega)$ such that, when $n > N_1(\omega)$, $\sup_{\theta \in G} \|D_n(\theta)\| n^{-\frac{1}{3}} < \frac{\varepsilon}{2}$ holds. Meanwhile, there exists $N_2(\omega) > N_1(\omega)$ such that, when $n > N_2(\omega)$, $\max_{1 \leq t \leq N_1} \sup_{\theta \in G} \|D_t(\theta)\| < \frac{\varepsilon}{2}$ holds. Therefore,

$$\begin{aligned} & \frac{\max_{1 \leq t \leq n} \sup_{\theta \in G} \|D_t(\theta)\|}{n^{\frac{1}{3}}} \\ & \leq \frac{\max_{1 \leq t \leq N_1} \sup_{\theta \in G} \|D_t(\theta)\|}{n^{\frac{1}{3}}} + \frac{\max_{N_1+1 \leq t \leq n} \sup_{\theta \in G} \|D_t(\theta)\|}{n^{\frac{1}{3}}} \\ & \leq \frac{\varepsilon}{2} + \max_{N_1+1 \leq t \leq n} \left\{ \frac{\sup_{\theta \in G} \|D_t(\theta)\|}{t^{\frac{1}{3}}} \right\} \\ & \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Hence, $\max_{1 \leq t \leq n} \sup_{\theta \in G} \|D_t(\theta)\| = o(n^{\frac{1}{3}})$, *a.s.*

(ii) The proof is analogous to that of (i). $\square$

Proof of Lemma 6. (i) We first show that $\sup_{\theta \in G} \|b(\theta)\| = O_p(n^{-\frac{1}{3}})$.

Let $\beta = \frac{b(\theta)}{\|b(\theta)\|}$, so that β is a unit vector. Define

$$\begin{aligned} S_{1n}(\theta, b(\theta)) &= \frac{1}{n} \sum_{t=1}^n D_t(\theta) \frac{b(\theta)}{1 + b^T(\theta) D_t(\theta)} \\ &= \frac{1}{n} \sum_{t=1}^n D_t(\theta) \left(1 - \frac{b^T(\theta) D_t(\theta)}{1 + b^T(\theta) D_t(\theta)} \right) \\ &= 0. \end{aligned} \quad (A2)$$

Thus,

$$\begin{aligned} 0 &= \|S_{1n}(\theta, b(\theta))\| \\ &\geq \left\| \beta^T(\theta) S_{1n}(\theta, b(\theta)) \right\| \\ &\geq \left| \frac{1}{n} \|b(\theta)\| \sum_{t=1}^n \frac{\beta^T(\theta) D_t(\theta) D_t^T(\theta) \beta(\theta)}{1 + \|b(\theta)\| \beta^T(\theta) D_t(\theta)} \right| - \left| \frac{1}{n} \sum_{t=1}^n \beta^T(\theta) D_t(\theta) \right| \\ &\geq \frac{\|b(\theta)\| \beta^T(\theta) \Omega_n(\theta) \beta(\theta)}{\max_{1 \leq t \leq n} [1 + \|b(\theta)\| \beta^T(\theta) D_t(\theta)]} - \left| \beta^T(\theta) D_n^*(\theta) \right|. \end{aligned}$$

Consequently,

$$\|D_n^*(\theta)\| \geq \left\| \beta^T(\theta) D_n^*(\theta) \right\| \geq \frac{\|b(\theta)\| \beta^T(\theta) \Omega_n(\theta) \beta(\theta)}{1 + \|b(\theta)\| \max_{1 \leq t \leq n} \sup_{\theta \in G} \|D_t(\theta)\|}.$$

Since $\Omega_n(\theta) \xrightarrow{P} \Omega(\theta_0)$, we have $\Omega_n(\theta) = \Omega(\theta_0) + o_p(1)$. Therefore,

$$\sigma_{\min} + o_p(1) \leq \beta^T(\theta) \Omega_n(\theta) \beta(\theta) = \beta^T(\theta) (\Omega(\theta_0) + o_p(1)) \beta(\theta) \leq \sigma_{\max} + o_p(1),$$

where $\sigma_{\min}$ and $\sigma_{\max}$ are the smallest and largest eigenvalues of $\Omega(\theta_0)$, respectively. Hence,

$$\begin{aligned} \sup_{\theta \in G} \|D_n^*(\theta)\| &\geq \left| \beta^T(\theta) D_n^*(\theta) \right| \\ &\geq \frac{\|b(\theta)\| \beta^T(\theta) \Omega_n(\theta) \beta(\theta)}{1 + \|b(\theta)\| \max_{1 \leq t \leq n} \sup_{\theta \in G} \|D_t(\theta)\|} \\ &\geq \frac{\|b(\theta)\| (\sigma_{\min} + o_p(1))}{1 + \|b(\theta)\| \max_{1 \leq t \leq n} \sup_{\theta \in G} \|D_t(\theta)\|}. \end{aligned} \quad (A3)$$

Rearranging (A3), we obtain

$$\sup_{\theta \in G} \|D_n^*(\theta)\| \geq \sup_{\theta \in G} \|b(\theta)\| \left[\sigma_{\min} + o_p(1) - \max_{1 \leq t \leq n} \sup_{\theta \in G} \|D_t(\theta)\| \sup_{\theta \in G} \|D_n^*(\theta)\| \right].$$

Moreover, since $\sup_{\theta \in G} \|D_n^*(\theta)\| = O_p(n^{-\frac{1}{3}})$,

$$\begin{aligned} \sup_{\theta \in G} \|b(\theta)\| &\leq \frac{\sup_{\theta \in G} \|D_n^*(\theta)\|}{\sigma_{\min} + o_p(1) - \max_{1 \leq t \leq n} \sup_{\theta \in G} \|D_n^*(\theta)\|} \\ &= \frac{O_p(n^{-\frac{1}{3}})}{\sigma_{\min} + o_p(1) - o_p(n^{\frac{1}{3}}) O_p(n^{-\frac{1}{3}})} \\ &= \frac{1}{\sigma_{\min}} O_p(n^{-\frac{1}{3}}) = O_p(n^{-\frac{1}{3}}). \end{aligned}$$

(ii) We next show that $\|b(\theta_0)\| = O_p(n^{-\frac{1}{2}})$.

By Lemma 1 in [21], $D_n^*(\theta_0) = O_p(n^{\frac{1}{2}})$. Then, by the same argument as in (i), we obtain $\|b(\theta_0)\| = O_p(n^{-\frac{1}{2}})$. $\square$

Proof of Lemma 7. Let $Z_t(\theta) = b^T(\theta) D_t(\theta)$. From (A2),

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{t=1}^n \frac{D_t(\theta)}{1 + b^T(\theta) D_t(\theta)} \\ &= \frac{1}{n} \sum_{t=1}^n D_t(\theta) \left(1 - b^T(\theta) D_t(\theta) + \frac{Z_t^2(\theta)}{1 + Z_t(\theta)} \right) \\ &= D_n^*(\theta) - \Omega_n(\theta) b(\theta) + \frac{1}{n} \sum_{t=1}^n D_t(\theta) \frac{Z_t^2(\theta)}{1 + Z_t(\theta)}, \end{aligned}$$

Hence,

$$\mathbf{b}(\boldsymbol{\theta}) = \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) \mathbf{D}_n^*(\boldsymbol{\theta}) + \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) \frac{1}{n} \sum_{t=1}^n \mathbf{D}_t(\boldsymbol{\theta}) \frac{Z_t^2(\boldsymbol{\theta})}{1 + Z_t(\boldsymbol{\theta})}. \quad (\text{A4})$$

Since $Z_t(\boldsymbol{\theta}) = \mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta})$, we have

$$\max_{1 \leq t \leq n} \sup_{\boldsymbol{\theta} \in G} |Z_t(\boldsymbol{\theta})| \leq \sup_{\boldsymbol{\theta} \in G} \|\mathbf{b}^T(\boldsymbol{\theta})\| \max_{1 \leq t \leq n} \sup_{\boldsymbol{\theta} \in G} \|\mathbf{D}_t(\boldsymbol{\theta})\| = O_p(n^{-\frac{1}{3}}) o(n^{\frac{1}{3}}) = o_p(1),$$

Therefore, as $n \rightarrow \infty$, $1 + Z_t \geq 1 - |Z_t| \geq 1 - \max_{1 \leq t \leq n} \sup_{\boldsymbol{\theta} \in G} |Z_t|$. Moreover,

$$\begin{aligned} & \left\| \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) \frac{1}{n} \sum_{t=1}^n \mathbf{D}_t(\boldsymbol{\theta}) \frac{Z_t^2(\boldsymbol{\theta})}{1 + Z_t(\boldsymbol{\theta})} \right\| \\ & \leq \left\| \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}_0) + o_p(1) \right\| \left\| \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta}) \mathbf{D}_t^T(\boldsymbol{\theta}) \mathbf{b}(\boldsymbol{\theta})}{1 + Z_t(\boldsymbol{\theta})} \right\| \max_{1 \leq t \leq n} \sup_{\boldsymbol{\theta} \in G} \|\mathbf{D}_t(\boldsymbol{\theta})\| \\ & \leq O_p(1) \frac{\max_{1 \leq t \leq n} \sup_{\boldsymbol{\theta} \in G} \|\mathbf{D}_t(\boldsymbol{\theta})\|}{1 - \max_{1 \leq t \leq n} \sup_{\boldsymbol{\theta} \in G} |Z_t|} \mathbf{b}^T(\boldsymbol{\theta}) \boldsymbol{\Omega}_n(\boldsymbol{\theta}) \mathbf{b}(\boldsymbol{\theta}) \\ & \leq O_p(1) O_p(1) o(n^{\frac{1}{3}}) \left(\sup_{\boldsymbol{\theta} \in G} \|\mathbf{b}(\boldsymbol{\theta})\| \right)^2 \boldsymbol{\beta}^T(\boldsymbol{\theta}) (\boldsymbol{\Omega}(\boldsymbol{\theta}_0) + o_p(1)) \boldsymbol{\beta}(\boldsymbol{\theta}) \\ & \leq o_p(n^{\frac{1}{3}}) O_p(n^{-\frac{2}{3}}) (\sigma_{\max} + o_p(1)) = o_p(n^{-\frac{1}{3}}). \end{aligned}$$

Consequently, $\sup_{\boldsymbol{\theta} \in G} \|\mathbf{b}(\boldsymbol{\theta}) - \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) \mathbf{D}_n^*(\boldsymbol{\theta})\| = o_p(n^{-\frac{1}{3}})$. $\square$

Proof of Theorem 1. Step 1. Proof of consistency.

First, consider the Taylor expansion of $L_E(\boldsymbol{\theta})$,

$$L_E(\boldsymbol{\theta}) = \sum_{t=1}^n \mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta}) - \frac{1}{2} \sum_{t=1}^n [\mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta})]^2 + o \left\{ \sum_{t=1}^n [\mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta})]^2 \right\}. \quad (\text{A5})$$

By Lemma 7, it follows that $\mathbf{b}(\boldsymbol{\theta}) = \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) \mathbf{D}_n^*(\boldsymbol{\theta}) + o_p(n^{-\frac{1}{3}})$.

Now, consider the first term on the right side of (A5),

$$\begin{aligned} \sum_{t=1}^n \mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta}) &= \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) \mathbf{D}_n^*(\boldsymbol{\theta}) + o_p(n^{-\frac{1}{3}}) \sum_{t=1}^n \mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta}) \\ &= n(\mathbf{D}_n^*(\boldsymbol{\theta}))^T \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) \mathbf{D}_n^*(\boldsymbol{\theta}) + n o_p(n^{-\frac{1}{3}}) \mathbf{D}_n^*(\boldsymbol{\theta}), \end{aligned} \quad (\text{A6})$$

where $n |o_p(n^{-\frac{1}{3}}) \mathbf{D}_n^*(\boldsymbol{\theta})| \leq n o_p(n^{-\frac{1}{3}}) \sup_{\boldsymbol{\theta} \in G} \|\mathbf{D}_n^*(\boldsymbol{\theta})\| = n o_p(n^{-\frac{1}{3}}) O_p(n^{-\frac{1}{3}}) = o_p(n^{\frac{1}{3}})$.

Consider the second term on the right side of (A5),

$$\begin{aligned} \sum_{t=1}^n [\mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta})]^2 &= n(\mathbf{D}_n^*(\boldsymbol{\theta}))^T \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) \mathbf{D}_n^*(\boldsymbol{\theta}) + 2n o_p(n^{-\frac{1}{3}}) \mathbf{D}_n^*(\boldsymbol{\theta}) \\ &\quad + n o_p(n^{-\frac{1}{3}}) \boldsymbol{\Omega}_n(\boldsymbol{\theta}) o_p(n^{-\frac{1}{3}}), \end{aligned} \quad (\text{A7})$$

where $|n o_p(n^{-\frac{1}{3}}) \boldsymbol{\Omega}_n(\boldsymbol{\theta}) o_p(n^{-\frac{1}{3}})| \leq n o_p(n^{-\frac{2}{3}}) (\sigma_{\max} + o_p(1)) = o_p(n^{\frac{1}{3}})$.

Consider the third term in (A5). We have

$$\begin{aligned}\sum_{t=1}^n \left[\mathbf{b}^T(\boldsymbol{\theta}) \mathbf{D}_t(\boldsymbol{\theta}) \right]^2 &= n \mathbf{b}^T(\boldsymbol{\theta}) \boldsymbol{\Omega}_n(\boldsymbol{\theta}) \mathbf{b}(\boldsymbol{\theta}) \\ &\leq n O_p(n^{-\frac{2}{3}}) (\sigma_{\max} + o_p(1)) \\ &= O_p(n^{\frac{1}{3}}).\end{aligned}\quad (\text{A8})$$

Therefore, from (A6)–(A8), we obtain

$$L_E(\boldsymbol{\theta}) = \frac{n}{2} (\mathbf{D}_n^*(\boldsymbol{\theta}))^T \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}) \mathbf{D}_n^*(\boldsymbol{\theta}) + o_p(n^{\frac{1}{3}}), \quad (\text{A9})$$

Second, taking the Taylor expansion of $\mathbf{D}_n^*(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$\mathbf{D}_n^*(\boldsymbol{\theta}) = \mathbf{D}_n^*(\boldsymbol{\theta}_0) + \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial \mathbf{D}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \quad (\text{A10})$$

Substituting (A10) into (A9), we have

$$L_E(\boldsymbol{\theta}) = \frac{n}{2} \left(O_p(n^{-\frac{5}{6}}) + O(n^{-\frac{2}{3}}) \boldsymbol{\delta}^T \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}_0) \boldsymbol{\delta} \right) + o_p(n^{\frac{1}{3}}).$$

Thus,

$$\inf_{\boldsymbol{\theta} \in \mathcal{B}} L_E(\boldsymbol{\theta}) \geq \frac{1}{2} O_p(n^{\frac{1}{6}}) + \frac{1}{2} O(n^{\frac{1}{3}}) \sigma_1 + o_p(n^{\frac{1}{3}}), \quad (\text{A11})$$

where σ_1 is the smallest eigenvalue of $\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}_0)$.

On the other hand,

$$L_E(\boldsymbol{\theta}_0) = \sum_{t=1}^n \mathbf{b}^T(\boldsymbol{\theta}_0) \mathbf{D}_t(\boldsymbol{\theta}_0) - \frac{1}{2} \sum_{t=1}^n \left[\mathbf{b}^T(\boldsymbol{\theta}_0) \mathbf{D}_t(\boldsymbol{\theta}_0) \right]^2 + o \left\{ \sum_{t=1}^n \left[\mathbf{b}^T(\boldsymbol{\theta}_0) \mathbf{D}_t(\boldsymbol{\theta}_0) \right]^2 \right\}.$$

Since $\sum_{t=1}^n \left[\mathbf{b}^T(\boldsymbol{\theta}_0) \mathbf{D}_t(\boldsymbol{\theta}_0) \right]^2 = n \mathbf{b}^T(\boldsymbol{\theta}_0) \boldsymbol{\Omega}_n(\boldsymbol{\theta}_0) \mathbf{b}(\boldsymbol{\theta}_0)$, and

$$\left| n \mathbf{b}^T(\boldsymbol{\theta}_0) \boldsymbol{\Omega}_n(\boldsymbol{\theta}_0) \mathbf{b}(\boldsymbol{\theta}_0) \right| \leq n \|\mathbf{b}(\boldsymbol{\theta}_0)\|^2 (\sigma_{\max} + o_p(1)) = n O_p(n^{-\frac{1}{2}}) O_p(n^{-\frac{1}{2}}) = O_p(1).$$

therefore,

$$L_E(\boldsymbol{\theta}_0) = \frac{n}{2} (\mathbf{D}_n^*(\boldsymbol{\theta}_0))^T \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}_0) \mathbf{D}_n^*(\boldsymbol{\theta}_0) + o_p(1) = O_p(\log_2^n). \quad (\text{A12})$$

Since $\boldsymbol{\theta} \in \mathcal{G} = \left\{ \boldsymbol{\theta} \mid \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < n^{-\frac{1}{3}} \right\}$ and $L_E(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, from (A11) and (A12), as $n \rightarrow \infty$,

$$\begin{aligned}P(\hat{\boldsymbol{\theta}}^{MEL} \in \mathcal{G}) &\geq P \left(\frac{\inf_{\boldsymbol{\theta} \in \mathcal{B}} L_E(\boldsymbol{\theta}) - L_E(\boldsymbol{\theta}_0)}{n^{\frac{1}{3}}} \geq \frac{O(n^{\frac{1}{3}}) \sigma_1}{3n^{\frac{1}{3}}} \right) \\ &\geq P \left(\frac{\frac{1}{2} O_p(n^{\frac{1}{6}}) + \frac{1}{2} O(n^{\frac{1}{3}}) \sigma_1 + o_p(n^{\frac{1}{3}}) - O_p(\log_2^n)}{n^{\frac{1}{3}}} \geq \frac{O(n^{\frac{1}{3}}) \sigma_1}{3n^{\frac{1}{3}}} \right) \\ &\rightarrow 1.\end{aligned}$$

Therefore, $\lim_{n \rightarrow \infty} P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > n^{-\frac{1}{3}}) = 0$, that is, $\hat{\boldsymbol{\theta}}^{MEL} \xrightarrow{P} \boldsymbol{\theta}_0$. Hence, $\hat{\boldsymbol{\theta}}^{MEL}$ is a consistent estimator of $\boldsymbol{\theta}_0$.

Step 2. Proof of asymptotic normality.

Let

$$S_{1n}(\theta, \mathbf{b}(\theta)) = \frac{1}{n} \sum_{t=1}^n D_t(\theta) \left(1 - \frac{\mathbf{b}^T(\theta) D_t(\theta)}{1 + \mathbf{b}^T(\theta) D_t(\theta)} \right),$$

$$S_{2n}(\theta, \mathbf{b}(\theta)) = \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial D_t(\theta)}{\partial \theta} \right) \frac{\mathbf{b}(\theta)}{1 + \mathbf{b}^T(\theta) D_t(\theta)},$$

Since step 1 in this theorem, $L_E(\theta)$, achieves the minimum in G , $\hat{\theta}^{MEL}$ satisfies

$$\mathbf{0} = \frac{1}{n} \frac{\partial L_E(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}^{MEL}} = \left(\frac{\partial \mathbf{b}(\theta)}{\partial \theta} \right)^T S_{1n}(\theta, \mathbf{b}(\theta)) \Big|_{\theta=\hat{\theta}^{MEL}} + S_{2n}(\theta, \mathbf{b}(\theta)) \Big|_{\theta=\hat{\theta}^{MEL}}.$$

Let $\hat{\mathbf{b}} = \mathbf{b}(\hat{\theta}^{MEL})$. Since $S_{1n}(\hat{\theta}^{MEL}, \hat{\mathbf{b}}) = \mathbf{0}$, it follows that $S_{2n}(\hat{\theta}^{MEL}, \hat{\mathbf{b}}) = \mathbf{0}$.

We now take the Taylor expansions of $S_{1n}(\hat{\theta}^{MEL}, \hat{\mathbf{b}})$ and $S_{2n}(\hat{\theta}^{MEL}, \hat{\mathbf{b}})$ at $(\theta_0, \mathbf{0})$; therefore,

$$\begin{aligned} \mathbf{0} &= S_{1n}(\theta_0, \mathbf{0}) + \frac{\partial S_{1n}(\theta, \mathbf{b})}{\partial \theta^T} \Big|_{\substack{\theta=\theta_0 \\ \mathbf{b}=\mathbf{0}}} (\hat{\theta}^{MEL} - \theta_0) + \frac{\partial S_{1n}(\theta, \mathbf{b})}{\partial \mathbf{b}^T} \Big|_{\substack{\theta=\theta_0 \\ \mathbf{b}=\mathbf{0}}} (\hat{\mathbf{b}} - \mathbf{0}) \\ &\quad + o_p(\rho_n), \end{aligned} \quad (\text{A13})$$

$$\begin{aligned} \mathbf{0} &= S_{2n}(\theta_0, \mathbf{0}) + \frac{\partial S_{1n}(\theta, \mathbf{b})}{\partial \theta^T} \Big|_{\substack{\theta=\theta_0 \\ \mathbf{b}=\mathbf{0}}} (\hat{\theta}^{MEL} - \theta_0) + \frac{\partial S_{2n}(\theta, \mathbf{b})}{\partial \mathbf{b}^T} \Big|_{\substack{\theta=\theta_0 \\ \mathbf{b}=\mathbf{0}}} (\hat{\mathbf{b}} - \mathbf{0}) \\ &\quad + o_p(\rho_n), \end{aligned} \quad (\text{A14})$$

where $\rho_n = \|\hat{\theta}^{MEL} - \theta_0\|^2 + \|\hat{\mathbf{b}} - \mathbf{0}\|^2 = O(n^{-\frac{2}{3}})$.

Note that $S_{1n}(\theta_0, \mathbf{0}) = D_n^*(\theta_0)$, then $S_{2n}(\theta_0, \mathbf{0}) = \mathbf{0}$, and

$$\frac{\partial S_{1n}(\theta, \mathbf{b})}{\partial \theta^T} \Big|_{\substack{\theta=\theta_0 \\ \mathbf{b}=\mathbf{0}}} = \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial D_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right), \quad \frac{\partial S_{1n}(\theta, \mathbf{b})}{\partial \mathbf{b}^T} \Big|_{\substack{\theta=\theta_0 \\ \mathbf{b}=\mathbf{0}}} = -\Omega_n(\theta_0),$$

$$\frac{\partial S_{2n}(\theta, \mathbf{b})}{\partial \theta^T} \Big|_{\substack{\theta=\theta_0 \\ \mathbf{b}=\mathbf{0}}} = \mathbf{0}, \quad \frac{\partial S_{2n}(\theta, \mathbf{b})}{\partial \mathbf{b}^T} \Big|_{\substack{\theta=\theta_0 \\ \mathbf{b}=\mathbf{0}}} = \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial D_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right).$$

Hence, (A13) and (A14) can be combined and rearranged as

$$\begin{bmatrix} \hat{\theta}^{MEL} - \theta_0 \\ \hat{\mathbf{b}} - \mathbf{0} \end{bmatrix} = \mathbf{M}_n^{-1} \begin{bmatrix} D_n^*(\theta_0) + o_p(\rho_n) \\ o_p(\rho_n) \end{bmatrix},$$

$$\text{where } \mathbf{M}_n = \begin{bmatrix} \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial D_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right) & -\Omega_n(\theta_0) \\ \mathbf{0} & \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial D_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right) \end{bmatrix}.$$

By the ergodicity of model (1) and [23], we have $\mathbf{M}_n^{-1} \xrightarrow{a.s.} \begin{bmatrix} -\mathbf{V}^{-1} & -\Phi \\ \mathbf{0} & -\mathbf{V}^{-1} \end{bmatrix}$.

Therefore, $\sqrt{n}(\hat{\theta}^{MEL} - \theta_0) = \sqrt{n}\mathbf{V}^{-1}D_n^*(\theta_0) + o_p(1)$. By Lemma 1,

$$\sqrt{n}(\hat{\theta}^{MEL} - \theta_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{V}^{-1}\Omega(\theta_0)\mathbf{V}^{-1}).$$

Moreover, for model (1), we have

$$\mathbf{V}^{-1} = (v_{ij})_{5 \times 5} = -\frac{1}{n} \sum_{t=1}^n \left(\frac{\partial D_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right),$$

where the nonzero entries are given by

$$\begin{aligned}
 v_{11} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n X_{1,t-1}^2, & v_{13} = v_{31} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n X_{1,t-1}, \\
 v_{22} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n X_{2,t-1}^2, & v_{24} = v_{42} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n X_{2,t-1}, \\
 v_{33} &= 1, & v_{44} &= 1, & v_{55} &= 1, \\
 v_{51} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n (X_{2,t} - \alpha_{20} X_{2,t-1} - \lambda_{20}) X_{1,t-1}, \\
 v_{52} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n (X_{1,t} - \alpha_{10} X_{1,t-1} - \lambda_{10}) X_{2,t-1}, \\
 v_{53} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n (X_{2,t} - \alpha_{20} X_{2,t-1} - \lambda_{20}), \\
 v_{54} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n (X_{1,t} - \alpha_{10} X_{1,t-1} - \lambda_{10}).
 \end{aligned}$$

All other entries are zeros. $\square$

Appendix B

Table A1. Bias, SE and MSE of the parameter estimators in Model I ($n = 100$).

True Value ($\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi$)		Bias			SE			MSE		
		CLS	CML	MEL	CLS	CML	MEL	CLS	CML	MEL
(0.1, 0.1, 2, 2, 1)	$\hat{\alpha}_1$	−0.0113	0.0108	0.0106	0.0990	0.0875	0.0863	0.0099	0.0078	0.0076
	$\hat{\alpha}_2$	−0.0177	−0.0124	−0.0128	0.1027	0.0954	0.0947	0.0109	0.0093	0.0091
	$\hat{\lambda}_1$	0.0202	0.0186	0.0179	0.2584	0.2136	0.2120	0.0672	0.0460	0.0453
	$\hat{\lambda}_2$	0.0391	−0.0362	−0.0354	0.2704	0.2480	0.2425	0.0746	0.0628	0.0601
	$\hat{\phi}$	0.0678	0.0651	0.0647	0.3611	0.2897	0.2834	0.1350	0.0882	0.0845
(0.1, 0.1, 2, 4, 1)	$\hat{\alpha}_1$	−0.0132	0.0127	0.0132	0.0993	0.0864	0.0851	0.0100	0.0076	0.0074
	$\hat{\alpha}_2$	−0.0189	−0.0163	−0.0160	0.0991	0.0852	0.0849	0.0102	0.0075	0.0075
	$\hat{\lambda}_1$	0.0288	0.0204	0.0207	0.2617	0.2243	0.2206	0.0693	0.0507	0.0491
	$\hat{\lambda}_2$	0.0622	0.0675	0.0668	0.2857	0.2471	0.2408	0.0855	0.0656	0.0624
	$\hat{\phi}$	0.0488	−0.0479	−0.0473	0.3467	0.2526	0.2517	0.1226	0.0661	0.0656
(0.3, 0.3, 2, 2, 1)	$\hat{\alpha}_1$	−0.0237	−0.0221	−0.0216	0.0984	0.0821	0.0814	0.0102	0.0072	0.0071
	$\hat{\alpha}_2$	−0.0227	−0.0218	−0.0215	0.0990	0.0833	0.0827	0.0103	0.0074	0.0073
	$\hat{\lambda}_1$	0.0586	0.0476	0.0477	0.3039	0.2549	0.2526	0.0958	0.0672	0.0661
	$\hat{\lambda}_2$	0.0649	−0.0523	−0.0525	0.3099	0.2507	0.2491	0.1003	0.0656	0.0648
	$\hat{\phi}$	0.0432	0.0387	0.0384	0.4089	0.3230	0.3213	0.1691	0.1058	0.1047
(0.3, 0.3, 2, 4, 1)	$\hat{\alpha}_1$	−0.0217	0.0202	0.0210	0.0995	0.0826	0.0815	0.0104	0.0072	0.0071
	$\hat{\alpha}_2$	−0.0178	0.0157	0.0143	0.0958	0.0834	0.0821	0.0095	0.0072	0.0069
	$\hat{\lambda}_1$	0.0647	−0.0569	−0.0545	0.3152	0.2637	0.2633	0.1035	0.0728	0.0723
	$\hat{\lambda}_2$	0.0624	−0.0576	−0.0569	0.3795	0.2864	0.2853	0.1479	0.0853	0.0846
	$\hat{\phi}$	0.0370	−0.0324	−0.0331	0.4530	0.3001	0.2988	0.2066	0.0911	0.0904
(0.1, 0.3, 2, 2, 1)	$\hat{\alpha}_1$	−0.0116	−0.0104	−0.0102	0.1005	0.0908	0.0881	0.0102	0.0084	0.0079
	$\hat{\alpha}_2$	−0.0256	−0.0227	−0.0230	0.0992	0.0856	0.0837	0.0105	0.0078	0.0075
	$\hat{\lambda}_1$	0.0234	0.0219	0.0224	0.2637	0.2249	0.2216	0.0701	0.0511	0.0496
	$\hat{\lambda}_2$	0.0583	0.0468	0.0457	0.3024	0.2463	0.2403	0.0948	0.0629	0.0598
	$\hat{\phi}$	0.0426	0.0376	0.0372	0.3562	0.3168	0.3114	0.1287	0.1018	0.0984

Table A1. Cont.

True Value ($\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi$)		Bias			SE			MSE		
		CLS	CML	MEL	CLS	CML	MEL	CLS	CML	MEL
(0.1, 0.3, 2, 4, 1)	$\hat{\alpha}_1$	−0.0101	0.0114	0.0109	0.0972	0.0826	0.0804	0.0095	0.0070	0.0066
	$\hat{\alpha}_2$	−0.0217	−0.0167	−0.0153	0.0957	0.0814	0.0819	0.0096	0.0069	0.0069
	$\hat{\lambda}_1$	0.0319	0.0303	0.0304	0.2614	0.2289	0.2263	0.0693	0.0533	0.0521
	$\hat{\lambda}_2$	0.0497	0.0453	0.0476	0.2681	0.2304	0.2298	0.0743	0.0551	0.0551
	$\hat{\phi}$	0.0517	−0.0471	−0.0462	0.3538	0.2906	0.2876	0.1278	0.0867	0.0848

Table A2. Bias, SE and MSE of the parameter estimators in Model I ($n = 300$).

True Value ($\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi$)		Bias			SE			MSE		
		CLS	CML	MEL	CLS	CML	MEL	CLS	CML	MEL
(0.1, 0.1, 2, 2, 1)	$\hat{\alpha}_1$	−0.0044	0.0041	0.0043	0.0568	0.0456	0.0452	0.0032	0.0021	0.0021
	$\hat{\alpha}_2$	−0.0036	−0.0032	−0.0030	0.0597	0.0473	0.0471	0.0036	0.0022	0.0022
	$\hat{\lambda}_1$	0.0080	0.0073	0.0068	0.1496	0.1089	0.1083	0.0224	0.0119	0.0118
	$\hat{\lambda}_2$	0.0039	−0.0044	−0.0037	0.1524	0.1123	0.1121	0.0232	0.0126	0.0126
	$\hat{\phi}$	0.0164	0.0158	0.0154	0.1940	0.1217	0.1215	0.0379	0.0151	0.0150
(0.1, 0.1, 2, 4, 1)	$\hat{\alpha}_1$	−0.0058	0.0049	0.0045	0.0574	0.0453	0.0454	0.0033	0.0021	0.0021
	$\hat{\alpha}_2$	−0.0102	−0.0067	−0.0062	0.0579	0.0428	0.0426	0.0035	0.0019	0.0019
	$\hat{\lambda}_1$	0.0124	0.0103	0.0098	0.1515	0.1142	0.1143	0.0231	0.0131	0.0132
	$\hat{\lambda}_2$	0.0259	0.0128	0.0117	0.1738	0.1209	0.1217	0.0309	0.0148	0.0149
	$\hat{\phi}$	0.0110	−0.0115	−0.0109	0.1998	0.1155	0.1150	0.0400	0.0135	0.0133
(0.3, 0.3, 2, 2, 1)	$\hat{\alpha}_1$	−0.0041	−0.0033	−0.0031	0.0570	0.0469	0.0465	0.0033	0.0022	0.0022
	$\hat{\alpha}_2$	−0.0048	−0.0052	−0.0054	0.0573	0.0471	0.0468	0.0033	0.0022	0.0022
	$\hat{\lambda}_1$	0.0081	0.0079	0.0082	0.1780	0.1205	0.1209	0.0317	0.0146	0.0147
	$\hat{\lambda}_2$	0.0119	−0.0114	−0.0106	0.1780	0.1193	0.1186	0.0318	0.0144	0.0142
	$\hat{\phi}$	0.0143	0.0135	0.0131	0.2376	0.1827	0.1821	0.0567	0.0336	0.0333
(0.3, 0.3, 2, 4, 1)	$\hat{\alpha}_1$	−0.0037	0.0040	0.0046	0.0574	0.0428	0.0425	0.0033	0.0018	0.0018
	$\hat{\alpha}_2$	−0.0067	0.0062	0.0061	0.0580	0.0457	0.0451	0.0034	0.0021	0.0021
	$\hat{\lambda}_1$	0.0056	−0.0057	−0.0059	0.1765	0.1225	0.1228	0.0312	0.0150	0.0151
	$\hat{\lambda}_2$	0.0426	−0.0119	−0.0123	0.1809	0.1237	0.1241	0.0345	0.0154	0.0156
	$\hat{\phi}$	0.0121	−0.0116	−0.0118	0.2479	0.1883	0.1879	0.0616	0.0356	0.0354
(0.1, 0.3, 2, 2, 1)	$\hat{\alpha}_1$	−0.0052	−0.0058	−0.0054	0.0594	0.0449	0.0443	0.0036	0.0020	0.0020
	$\hat{\alpha}_2$	−0.0095	−0.0101	−0.0099	0.0578	0.0461	0.0457	0.0034	0.0022	0.0022
	$\hat{\lambda}_1$	0.0140	0.0133	0.0123	0.1577	0.1131	0.1129	0.0251	0.0130	0.0129
	$\hat{\lambda}_2$	0.0262	0.0259	0.0258	0.1743	0.1276	0.1277	0.0311	0.0170	0.0170
	$\hat{\phi}$	0.0168	0.0153	0.0153	0.2042	0.1108	0.1115	0.0420	0.0125	0.0127
(0.1, 0.3, 2, 4, 1)	$\hat{\alpha}_1$	−0.0047	0.0044	0.0042	0.0603	0.0520	0.0518	0.0037	0.0027	0.0027
	$\hat{\alpha}_2$	−0.0082	−0.0078	−0.0077	0.0571	0.0492	0.0487	0.0033	0.0025	0.0024
	$\hat{\lambda}_1$	0.0155	0.0149	0.0152	0.1598	0.1116	0.1103	0.0258	0.0127	0.0124
	$\hat{\lambda}_2$	0.0152	0.0136	0.0129	0.1521	0.1129	0.1124	0.0234	0.0129	0.0128
	$\hat{\phi}$	0.0135	−0.0124	−0.0121	0.2035	0.1213	0.1209	0.0416	0.0149	0.0148

Table A3. Bias, SE and MSE of the parameter estimators in Model I ($n = 500$).

True Value ($\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi$)		Bias			SE			MSE		
		CLS	CML	MEL	CLS	CML	MEL	CLS	CML	MEL
(0.1, 0.1, 2, 2, 1)	$\hat{\alpha}_1$	−0.0033	0.0031	0.0032	0.0443	0.0316	0.0310	0.0020	0.0010	0.0010
	$\hat{\alpha}_2$	−0.0020	−0.0021	−0.0023	0.0449	0.0322	0.0321	0.0020	0.0010	0.0010
	$\hat{\lambda}_1$	0.0023	0.0016	0.0015	0.1168	0.0897	0.0898	0.0136	0.0080	0.0081
	$\hat{\lambda}_2$	0.0023	−0.0018	−0.0013	0.1140	0.0885	0.0883	0.0130	0.0078	0.0078
	$\hat{\phi}$	0.0094	0.0089	0.0087	0.1583	0.0913	0.0911	0.0251	0.0084	0.0084

Table A3. Cont.

True Value ($\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi$)		Bias			SE			MSE		
		CLS	CML	MEL	CLS	CML	MEL	CLS	CML	MEL
(0.1, 0.1, 2, 4, 1)	$\hat{\alpha}_1$	−0.0044	0.0041	0.0040	0.0466	0.0332	0.0329	0.0022	0.0011	0.0011
	$\hat{\alpha}_2$	−0.0056	−0.0049	−0.0047	0.0461	0.0325	0.0324	0.0022	0.0011	0.0011
	$\hat{\lambda}_1$	0.0120	0.0093	0.0091	0.1202	0.0974	0.0971	0.0146	0.0096	0.0095
	$\hat{\lambda}_2$	0.0211	0.0152	0.0144	0.2173	0.0893	0.0885	0.0477	0.0082	0.0080
	$\hat{\phi}$	0.0098	−0.0087	−0.0085	0.1559	0.0908	0.0905	0.0244	0.0083	0.0083
(0.3, 0.3, 2, 2, 1)	$\hat{\alpha}_1$	−0.0031	−0.0033	−0.0030	0.0465	0.0347	0.0345	0.0022	0.0012	0.0012
	$\hat{\alpha}_2$	−0.0045	−0.0040	−0.0033	0.0443	0.0342	0.0341	0.0020	0.0012	0.0012
	$\hat{\lambda}_1$	0.0073	0.0076	0.0075	0.1401	0.1059	0.1046	0.0197	0.0113	0.0110
	$\hat{\lambda}_2$	0.0104	−0.0099	−0.0092	0.1342	0.1063	0.1062	0.0181	0.0114	0.0114
	$\hat{\phi}$	0.0096	0.0090	0.0091	0.1868	0.1357	0.1355	0.0350	0.0185	0.0184
(0.3, 0.3, 2, 4, 1)	$\hat{\alpha}_1$	−0.0032	0.0031	0.0033	0.0459	0.0336	0.0331	0.0021	0.0011	0.0011
	$\hat{\alpha}_2$	−0.0054	0.0045	0.0046	0.0463	0.0341	0.0339	0.0022	0.0012	0.0012
	$\hat{\lambda}_1$	0.0051	−0.0047	−0.0045	0.1411	0.0995	0.0987	0.0199	0.0099	0.0098
	$\hat{\lambda}_2$	0.0098	−0.0078	−0.0077	0.1382	0.0988	0.0983	0.0192	0.0098	0.0097
	$\hat{\phi}$	0.0086	−0.0082	−0.0081	0.1727	0.1263	0.1259	0.0299	0.0160	0.0159
(0.1, 0.3, 2, 2, 1)	$\hat{\alpha}_1$	−0.0049	−0.0037	−0.0036	0.0434	0.0333	0.0329	0.0019	0.0011	0.0011
	$\hat{\alpha}_2$	−0.0030	−0.0029	−0.0027	0.0452	0.0358	0.0353	0.0021	0.0013	0.0013
	$\hat{\lambda}_1$	0.0082	0.0077	0.0074	0.1157	0.0862	0.0862	0.0135	0.0075	0.0075
	$\hat{\lambda}_2$	0.0064	0.0061	0.0065	0.1384	0.1027	0.1024	0.0192	0.0106	0.0105
	$\hat{\phi}$	0.0089	0.0084	0.0083	0.1578	0.0921	0.0917	0.0250	0.0086	0.0085
(0.1, 0.3, 2, 4, 1)	$\hat{\alpha}_1$	−0.0038	0.0035	0.0032	0.0450	0.0334	0.0333	0.0020	0.0011	0.0011
	$\hat{\alpha}_2$	−0.0054	−0.0046	−0.0047	0.0440	0.0339	0.0335	0.0020	0.0012	0.0011
	$\hat{\lambda}_1$	0.0094	0.0083	0.0084	0.1144	0.0855	0.0851	0.0132	0.0074	0.0073
	$\hat{\lambda}_2$	0.0108	0.0079	0.0076	0.2653	0.1038	0.1037	0.0705	0.0108	0.0108
	$\hat{\phi}$	0.0083	−0.0080	−0.0078	0.1640	0.1154	0.1152	0.0270	0.0134	0.0133

Table A4. Bias, SE and MSE of the parameter estimators in Model II ($n = 100$).

True Value ($\alpha_1, \alpha_2, \lambda_1, \lambda_2, \beta$)		Bias			SE			MSE		
		CLS	CML	MEL	CLS	CML	MEL	CLS	CML	MEL
(0.1, 0.1, 2, 2, $\frac{1}{4}$)	$\hat{\alpha}_1$	−0.0160	0.0142	0.0145	0.1014	0.0926	0.0924	0.0105	0.0088	0.0087
	$\hat{\alpha}_2$	−0.0121	−0.0117	−0.0113	0.1001	0.0917	0.0915	0.0102	0.0085	0.0085
	$\hat{\lambda}_1$	0.0355	0.0335	0.0333	0.2839	0.2469	0.2462	0.0819	0.0621	0.0617
	$\hat{\lambda}_2$	0.0203	−0.0186	−0.0174	0.2879	0.2473	0.2474	0.0833	0.0615	0.0615
	$\hat{\beta}$	0.0583	0.0541	0.0542	0.1943	0.1520	0.1523	0.0412	0.0260	0.0261
(0.1, 0.1, 2, 4, $\frac{1}{8}$)	$\hat{\alpha}_1$	−0.0200	0.0183	0.0180	0.1037	0.0945	0.0943	0.0112	0.0093	0.0092
	$\hat{\alpha}_2$	−0.0230	−0.0216	−0.0213	0.1017	0.0938	0.0941	0.0109	0.0093	0.0093
	$\hat{\lambda}_1$	0.0496	0.0474	0.0476	0.2793	0.2406	0.2403	0.0805	0.0601	0.0600
	$\hat{\lambda}_2$	0.0506	0.0481	0.0478	0.2855	0.2482	0.2487	0.0841	0.0639	0.0641
	$\hat{\beta}$	0.0598	−0.0552	−0.0549	0.0787	0.0691	0.0686	0.0098	0.0078	0.0077
(0.3, 0.3, 2, 2, $\frac{1}{4}$)	$\hat{\alpha}_1$	−0.0239	−0.0218	−0.0218	0.1015	0.0922	0.0911	0.0109	0.0090	0.0088
	$\hat{\alpha}_2$	−0.0147	−0.0123	−0.0121	0.0977	0.0889	0.0884	0.0098	0.0081	0.0080
	$\hat{\lambda}_1$	0.0603	0.0587	0.0583	0.3290	0.2676	0.3677	0.1119	0.0751	0.1386
	$\hat{\lambda}_2$	0.0474	−0.0440	−0.0442	0.3214	0.2664	0.3663	0.1055	0.0729	0.1361
	$\hat{\beta}$	0.0422	0.0411	0.0407	0.2351	0.2012	0.2014	0.0571	0.0422	0.0422
(0.3, 0.3, 2, 4, $\frac{1}{8}$)	$\hat{\alpha}_1$	−0.0275	0.0259	0.0243	0.0990	0.0871	0.0869	0.0106	0.0083	0.0081
	$\hat{\alpha}_2$	−0.0192	0.0166	0.0158	0.0944	0.0832	0.0834	0.0093	0.0072	0.0072
	$\hat{\lambda}_1$	0.0636	−0.0573	−0.0567	0.3106	0.2588	0.2583	0.1005	0.0703	0.0699
	$\hat{\lambda}_2$	0.0647	−0.0582	−0.0581	0.3542	0.2890	0.2886	0.1296	0.0869	0.0867
	$\hat{\beta}$	0.0390	0.0348	0.0349	0.0924	0.0818	0.0815	0.0101	0.0079	0.0079

Table A4. Cont.

True Value ($\alpha_1, \alpha_2, \lambda_1, \lambda_2, \beta$)		Bias			SE			MSE		
		CLS	CML	MEL	CLS	CML	MEL	CLS	CML	MEL
(0.1, 0.3, 2, 2, $\frac{1}{4}$)	$\hat{\alpha}_1$	−0.0153	−0.0137	−0.0132	0.1051	0.0938	0.0934	0.0113	0.0090	0.0089
	$\hat{\alpha}_2$	−0.0198	−0.0172	−0.0173	0.0979	0.0865	0.0851	0.0100	0.0078	0.0075
	$\hat{\lambda}_1$	0.0355	0.0326	0.0313	0.2925	0.2541	0.2543	0.0868	0.0656	0.0656
	$\hat{\lambda}_2$	0.0633	0.0578	0.0574	0.3236	0.2680	0.2674	0.1087	0.0752	0.0748
	$\hat{\beta}$	0.0556	0.0534	0.0531	0.2188	0.1872	0.1868	0.0510	0.0379	0.0377
(0.1, 0.3, 2, 4, $\frac{1}{8}$)	$\hat{\alpha}_1$	−0.0283	0.0259	0.0254	0.1013	0.0947	0.0942	0.0111	0.0096	0.0095
	$\hat{\alpha}_2$	−0.0215	−0.0187	−0.0182	0.0985	0.0859	0.0885	0.0102	0.0077	0.0082
	$\hat{\lambda}_1$	0.0371	0.0334	0.0332	0.2746	0.2408	0.2403	0.0768	0.0591	0.0588
	$\hat{\lambda}_2$	0.0435	0.0398	0.0376	0.2855	0.2454	0.2457	0.0834	0.0618	0.0618
	$\hat{\beta}$	0.0573	0.0536	0.0537	0.0799	0.0683	0.0676	0.0097	0.0075	0.0075

Table A5. Bias, SE and MSE of the parameter estimators in Model II ($n = 300$).

True Value ($\alpha_1, \alpha_2, \lambda_1, \lambda_2, \beta$)		Bias			SE			MSE		
		CLS	CML	MEL	CLS	CML	MEL	CLS	CML	MEL
(0.1, 0.1, 2, 2, $\frac{1}{4}$)	$\hat{\alpha}_1$	−0.0041	0.0046	0.0045	0.0611	0.0573	0.0567	0.0038	0.0033	0.0032
	$\hat{\alpha}_2$	−0.0058	−0.0053	−0.0052	0.0584	0.0476	0.0473	0.0034	0.0023	0.0023
	$\hat{\lambda}_1$	0.0063	0.0058	0.0055	0.1612	0.1237	0.1236	0.0260	0.0153	0.0153
	$\hat{\lambda}_2$	0.0153	−0.0091	−0.0087	0.1598	0.1159	0.1155	0.0258	0.0135	0.0134
	$\hat{\beta}$	0.1405	0.0137	0.0132	0.1116	0.0842	0.0843	0.0322	0.0073	0.0073
(0.1, 0.1, 2, 4, $\frac{1}{8}$)	$\hat{\alpha}_1$	−0.0086	0.0075	0.0073	0.0585	0.0471	0.0469	0.0035	0.0023	0.0023
	$\hat{\alpha}_2$	−0.0037	−0.0034	−0.0035	0.0595	0.0488	0.0485	0.0036	0.0024	0.0024
	$\hat{\lambda}_1$	0.0152	0.0101	0.0101	0.1531	0.1126	0.1019	0.0237	0.0128	0.0105
	$\hat{\lambda}_2$	0.0177	0.0126	0.0118	0.2965	0.1245	0.1243	0.0882	0.0157	0.0156
	$\hat{\beta}$	0.0170	−0.0143	−0.0140	0.0433	0.0382	0.0377	0.0022	0.0017	0.0016
(0.3, 0.3, 2, 2, $\frac{1}{4}$)	$\hat{\alpha}_1$	−0.0031	−0.0035	−0.0032	0.0587	0.0470	0.0466	0.0035	0.0022	0.0022
	$\hat{\alpha}_2$	−0.0070	−0.0068	−0.0065	0.0577	0.0481	0.0478	0.0034	0.0024	0.0023
	$\hat{\lambda}_1$	0.0081	0.0079	0.0076	0.1877	0.1334	0.1325	0.0353	0.0179	0.0176
	$\hat{\lambda}_2$	0.0178	−0.0123	−0.0121	0.1911	0.1403	0.1391	0.0368	0.0198	0.0195
	$\hat{\beta}$	0.0142	0.0136	0.0135	0.1437	0.1095	0.1088	0.0209	0.0122	0.0120
(0.3, 0.3, 2, 4, $\frac{1}{8}$)	$\hat{\alpha}_1$	−0.0106	0.0098	0.0089	0.0582	0.0466	0.0467	0.0035	0.0023	0.0023
	$\hat{\alpha}_2$	−0.0085	0.0087	0.0087	0.0555	0.0457	0.0451	0.0032	0.0022	0.0021
	$\hat{\lambda}_1$	0.0313	−0.0225	−0.0223	0.1852	0.1323	0.1322	0.0353	0.0180	0.0180
	$\hat{\lambda}_2$	0.0488	−0.0269	−0.0264	0.3310	0.1339	0.1337	0.1119	0.0187	0.0186
	$\hat{\beta}$	0.0131	0.0128	0.0122	0.0530	0.0411	0.0403	0.0030	0.0019	0.0018
(0.1, 0.3, 2, 2, $\frac{1}{4}$)	$\hat{\alpha}_1$	−0.0051	−0.0040	−0.0036	0.0596	0.0489	0.0480	0.0036	0.0024	0.0023
	$\hat{\alpha}_2$	0.0041	−0.0042	−0.0040	0.0580	0.0477	0.0473	0.0034	0.0023	0.0023
	$\hat{\lambda}_1$	0.0114	0.0076	0.0071	0.1601	0.1193	0.1191	0.0258	0.0143	0.0142
	$\hat{\lambda}_2$	0.0121	0.0083	0.0075	0.1843	0.1287	0.1285	0.0341	0.0166	0.0166
	$\hat{\beta}$	0.0190	0.0155	0.0152	0.1132	0.0861	0.0859	0.0132	0.0077	0.0076
(0.1, 0.3, 2, 4, $\frac{1}{8}$)	$\hat{\alpha}_1$	−0.0092	0.0078	0.0071	0.0600	0.0498	0.0489	0.0037	0.0025	0.0024
	$\hat{\alpha}_2$	−0.0065	−0.0056	−0.0058	0.0561	0.0442	0.0437	0.0032	0.0020	0.0019
	$\hat{\lambda}_1$	0.0189	0.0121	0.0114	0.1583	0.1156	0.1153	0.0254	0.0135	0.0134
	$\hat{\lambda}_2$	0.0271	0.0183	0.0173	0.3391	0.1163	0.1159	0.1157	0.0139	0.0137
	$\hat{\beta}$	0.0194	0.0147	0.0142	0.0476	0.0398	0.0396	0.0026	0.0018	0.0018

Table A6. Bias, SE and MSE of the parameter estimators in Model II ($n = 500$).

True value ($\alpha_1, \alpha_2, \lambda_1, \lambda_2, \beta$)		Bias			SE			MSE		
		CLS	CML	MEL	CLS	CML	MEL	CLS	CML	MEL
(0.1, 0.1, 2, 2, $\frac{1}{4}$)	$\hat{\alpha}_1$	−0.0025	0.0026	0.0024	0.0462	0.0347	0.0345	0.0021	0.0012	0.0012
	$\hat{\alpha}_2$	−0.0020	−0.0018	−0.0016	0.0457	0.0338	0.0337	0.0021	0.0011	0.0011
	$\hat{\lambda}_1$	0.0021	0.0017	0.0014	0.1231	0.0901	0.0896	0.0152	0.0081	0.0080
	$\hat{\lambda}_2$	0.0065	−0.0059	−0.0058	0.1260	0.0933	0.0930	0.0159	0.0087	0.0087
	$\hat{\beta}$	0.0083	0.0075	0.0073	0.0861	0.0659	0.0656	0.0075	0.0044	0.0044
(0.1, 0.1, 2, 4, $\frac{1}{8}$)	$\hat{\alpha}_1$	−0.0064	0.0061	0.0060	0.0458	0.0342	0.0340	0.0021	0.0012	0.0012
	$\hat{\alpha}_2$	−0.0033	−0.0029	−0.0027	0.0435	0.0314	0.0310	0.0019	0.0010	0.0010
	$\hat{\lambda}_1$	0.0070	0.0066	0.0063	0.1200	0.0916	0.0915	0.0144	0.0084	0.0084
	$\hat{\lambda}_2$	0.0098	0.0092	0.0088	0.1249	0.0957	0.0954	0.0157	0.0092	0.0092
	$\hat{\beta}$	0.0087	−0.0083	−0.0081	0.0324	0.0278	0.0275	0.0011	0.0008	0.0008
(0.3, 0.3, 2, 2, $\frac{1}{4}$)	$\hat{\alpha}_1$	−0.0023	−0.0017	−0.0017	0.0467	0.0343	0.0341	0.0022	0.0012	0.0012
	$\hat{\alpha}_2$	−0.0049	−0.0035	−0.0033	0.0446	0.0321	0.0318	0.0020	0.0010	0.0010
	$\hat{\lambda}_1$	0.0050	0.0044	0.0042	0.1472	0.1026	0.1021	0.0217	0.0105	0.0104
	$\hat{\lambda}_2$	0.0105	−0.0062	−0.0059	0.1501	0.1088	0.1085	0.0226	0.0119	0.0118
	$\hat{\beta}$	0.0093	0.0086	0.0084	0.1090	0.0854	0.0853	0.0120	0.0074	0.0073
(0.3, 0.3, 2, 4, $\frac{1}{8}$)	$\hat{\alpha}_1$	−0.0062	0.0060	0.0057	0.0434	0.0335	0.0332	0.0019	0.0012	0.0011
	$\hat{\alpha}_2$	−0.0044	0.0045	0.0042	0.0443	0.0339	0.0336	0.0020	0.0012	0.0011
	$\hat{\lambda}_1$	0.0139	−0.0051	−0.0048	0.1369	0.0926	0.0924	0.0189	0.0086	0.0086
	$\hat{\lambda}_2$	0.0185	−0.0078	−0.0077	0.1336	0.0908	0.0905	0.0182	0.0083	0.0082
	$\hat{\beta}$	0.0081	−0.0069	0.0065	0.0441	0.0377	0.0372	0.0020	0.0015	0.0014
(0.1, 0.3, 2, 2, $\frac{1}{4}$)	$\hat{\alpha}_1$	−0.0046	−0.0038	−0.0035	0.0476	0.0352	0.0348	0.0023	0.0013	0.0012
	$\hat{\alpha}_2$	−0.0038	−0.0036	−0.0034	0.0434	0.0316	0.0315	0.0019	0.0010	0.0010
	$\hat{\lambda}_1$	0.0091	0.0064	0.0061	0.1287	0.0831	0.0829	0.0166	0.0069	0.0069
	$\hat{\lambda}_2$	0.0096	0.0077	0.0074	0.1466	0.1045	0.1044	0.0216	0.0110	0.0110
	$\hat{\beta}$	0.0093	0.0089	0.0083	0.0910	0.0743	0.0741	0.0084	0.0056	0.0056
(0.1, 0.3, 2, 4, $\frac{1}{8}$)	$\hat{\alpha}_1$	−0.0087	0.0071	0.0069	0.0452	0.0359	0.0356	0.0021	0.0013	0.0013
	$\hat{\alpha}_2$	−0.0039	−0.0040	−0.0036	0.0441	0.0346	0.0344	0.0020	0.0012	0.0012
	$\hat{\lambda}_1$	0.0090	0.0085	0.0081	0.1211	0.0810	0.0803	0.0147	0.0066	0.0065
	$\hat{\lambda}_2$	0.0153	0.0063	0.0062	0.1262	0.0837	0.0831	0.0162	0.0070	0.0069
	$\hat{\beta}$	0.0085	0.0082	0.0077	0.0352	0.0293	0.0287	0.0013	0.0009	0.0009

Table A7. One-step-ahead prediction errors for Model I and Model II ($n = 500$) averaged over 2000 replications.

Model	Method	MSPE ₁	MSPE ₂	AMSPE	MAE ₁	MAE ₂	AMAE
Model I	CLS	2.5340	2.5361	2.5351	1.2428	1.2430	1.2427
	CML	2.5225	2.5230	2.5228	1.2341	1.2335	1.2338
	MEL	2.5241	2.5224	2.5233	1.2327	1.2331	1.2329
Model II	CLS	3.4175	3.4265	3.4220	1.4248	1.4286	1.4267
	CML	3.4098	3.4011	3.4055	1.4164	1.4103	1.4134
	MEL	3.4082	3.4056	3.4069	1.4149	1.4101	1.4125

Table A8. Parameter estimates, SE, AIC, and BIC.

Model	CML				MEL			
	Estimate	SE	AIC	BIC	Estimate	SE	AIC	BIC
BINAR(1)-I	$\hat{\alpha}_1 = 0.3942$	0.0692	947.0418	961.8909	$\hat{\alpha}_1 = 0.3945$	0.0683	943.4658	958.3149
	$\hat{\alpha}_2 = 0.1925$	0.0742			$\hat{\alpha}_2 = 0.1876$	0.0737		
	$\hat{\lambda}_1 = 0.8874$	0.1677			$\hat{\lambda}_1 = 0.8559$	0.0166		
	$\hat{\lambda}_2 = 0.9036$	0.1874			$\hat{\lambda}_2 = 0.8894$	0.1843		
	$\hat{\phi} = 0.0843$	0.2139			$\hat{\phi} = 0.0917$	0.2105		

Table A8. Cont.

Model	CML				MEL			
	Estimate	SE	AIC	BIC	Estimate	SE	AIC	BIC
BINAR(1)-II	$\hat{\alpha}_1 = 0.3695$	0.0779	922.5176	937.3667	$\hat{\alpha}_1 = 0.3581$	0.0775	919.2084	934.0575
	$\hat{\alpha}_2 = 0.1782$	0.0724			$\hat{\alpha}_2 = 0.1669$	0.0722		
	$\hat{\lambda}_1 = 0.9237$	0.1927			$\hat{\lambda}_1 = 0.9176$	0.1923		
	$\hat{\lambda}_2 = 0.9179$	0.1889			$\hat{\lambda}_2 = 0.9101$	0.1887		
	$\hat{\beta} = 0.1014$	0.1228			$\hat{\beta} = 0.0998$	0.1230		
IINAR(1)-I	$\hat{\alpha}_1 = 0.3869$	0.0598	963.1124	977.9615	$\hat{\alpha}_1 = 0.3872$	0.0587	961.9123	976.7614
	$\hat{\alpha}_2 = 0.1890$	0.0771			$\hat{\alpha}_2 = 0.1894$	0.0763		
	$\hat{\lambda}_1 = 0.8688$	0.1684			$\hat{\lambda}_1 = 0.8573$	0.1679		
	$\hat{\lambda}_2 = 0.8901$	0.2225			$\hat{\lambda}_2 = 0.8766$	0.2218		
IINAR(1)-II	$\hat{\alpha}_1 = 0.3687$	0.0861	938.3437	953.1928	$\hat{\alpha}_1 = 0.3548$	0.0852	935.3807	950.2298
	$\hat{\alpha}_2 = 0.1742$	0.0854			$\hat{\alpha}_2 = 0.1860$	0.0849		
	$\hat{\lambda}_1 = 0.9255$	0.2038			$\hat{\lambda}_1 = 0.9103$	0.2029		
	$\hat{\lambda}_2 = 0.9183$	0.2013			$\hat{\lambda}_2 = 0.9247$	0.2011		
	$\hat{\beta} = 0.1009$	0.1344			$\hat{\beta} = 0.1136$	0.1327		

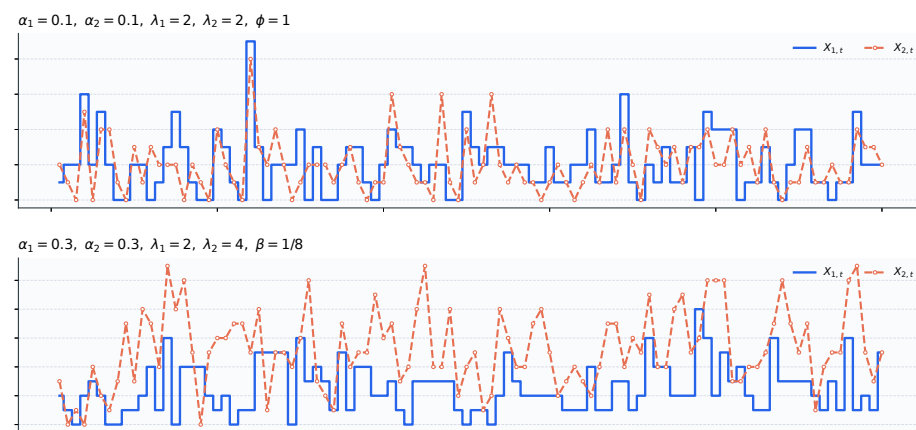


Figure A1. Plots of sample paths for Model I and Model II.

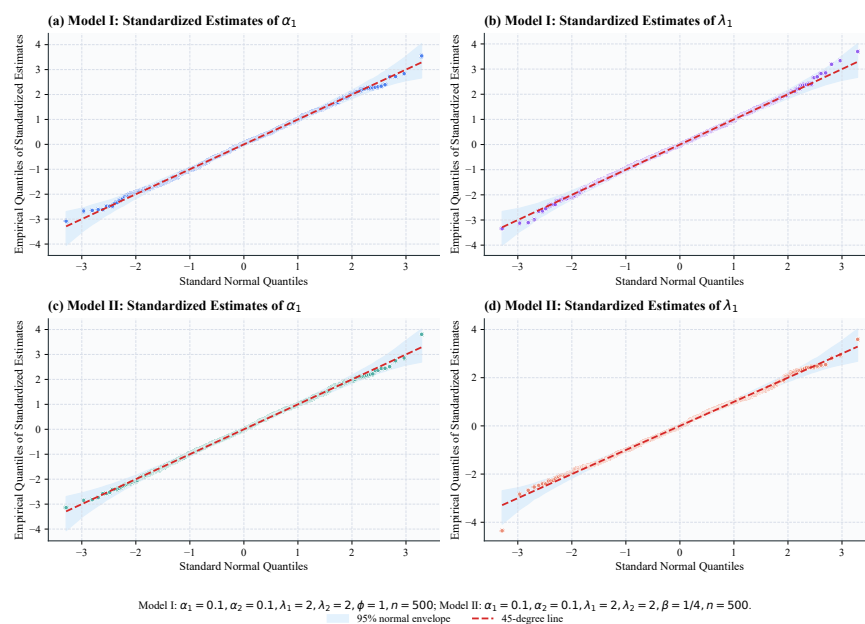


Figure A2. Q-Q Plots for some parameters in Model I and Model II.

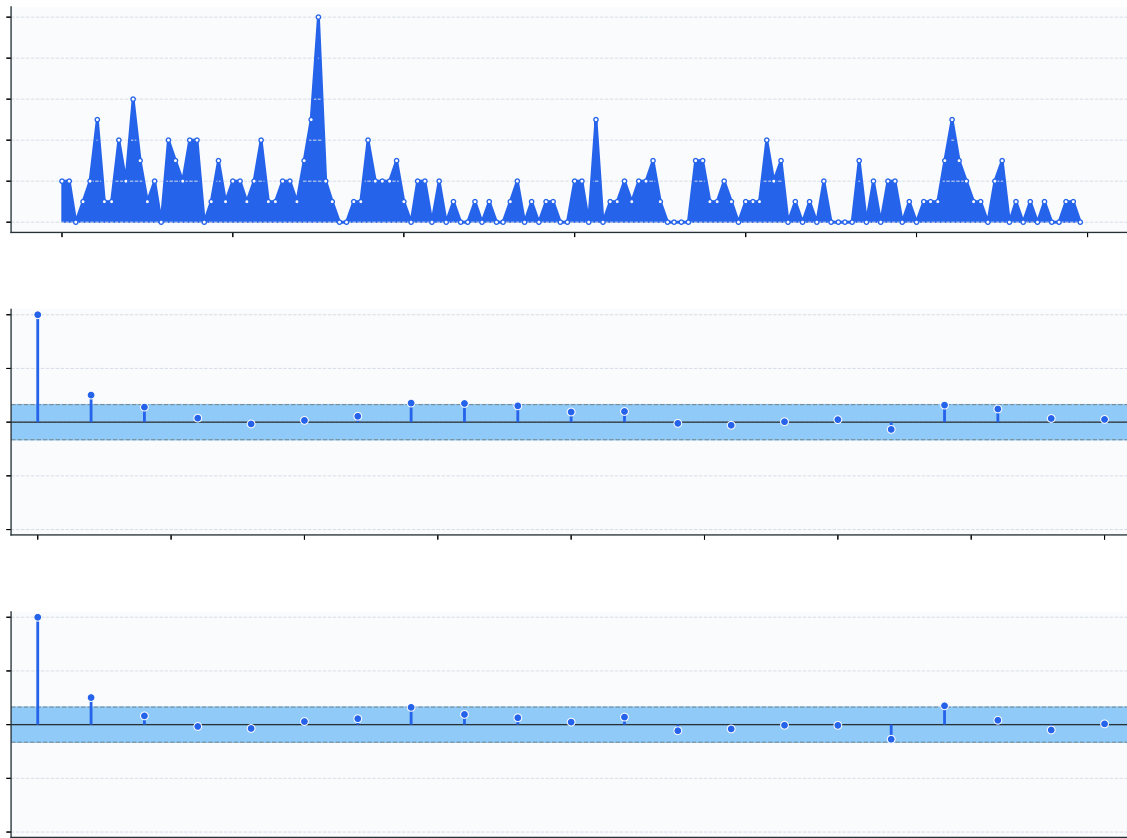


Figure A3. Sample path plot, sample ACF plot, and sample PACF plot for the burglary data.

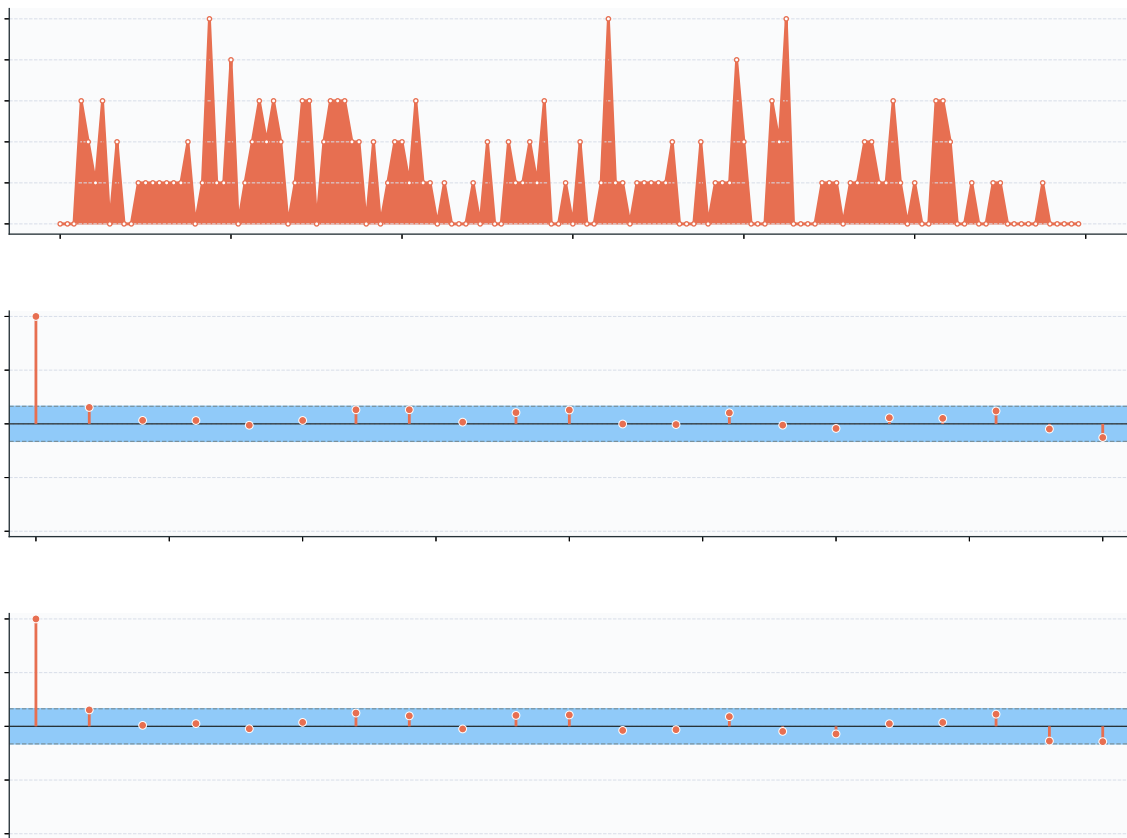
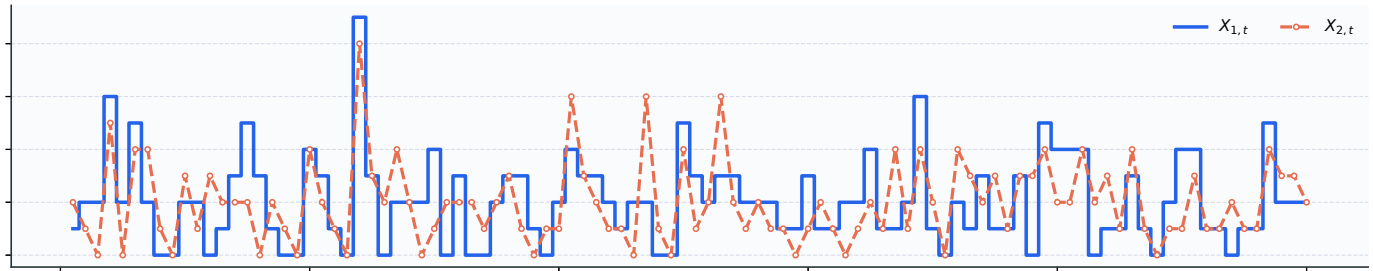


Figure A4. Sample path plot, sample ACF plot, and sample PACF plot for the robbery data.

$$\alpha_1 = 0.1, \alpha_2 = 0.1, \lambda_1 = 2, \lambda_2 = 2, \phi = 1$$



$$\alpha_1 = 0.3, \alpha_2 = 0.3, \lambda_1 = 2, \lambda_2 = 4, \beta = 1/8$$

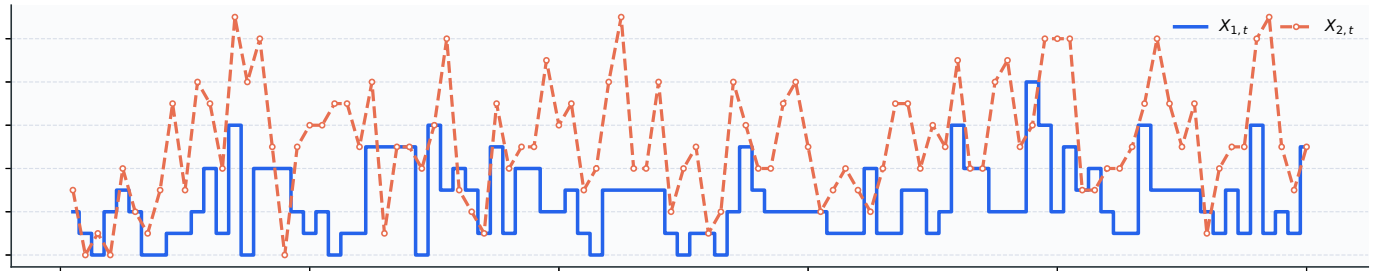


Figure A5. Sample CCF plot for the burglary and robbery data.

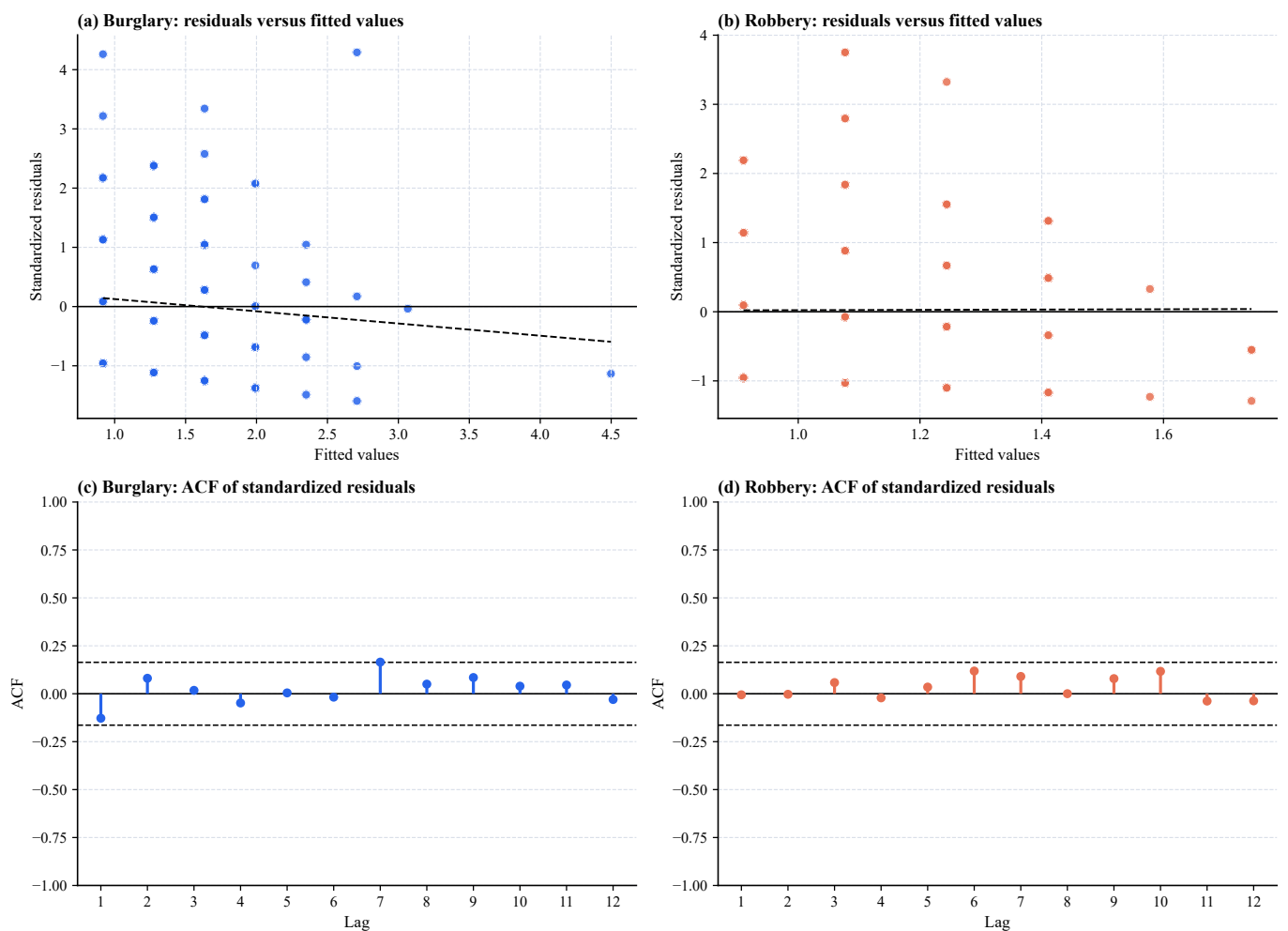


Figure A6. Residual diagnostics for the selected BINAR(1)-II model with MEL estimation. Panels (a,b) show standardized residuals versus fitted values, and panels (c,d) show the autocorrelation functions of the standardized residuals.

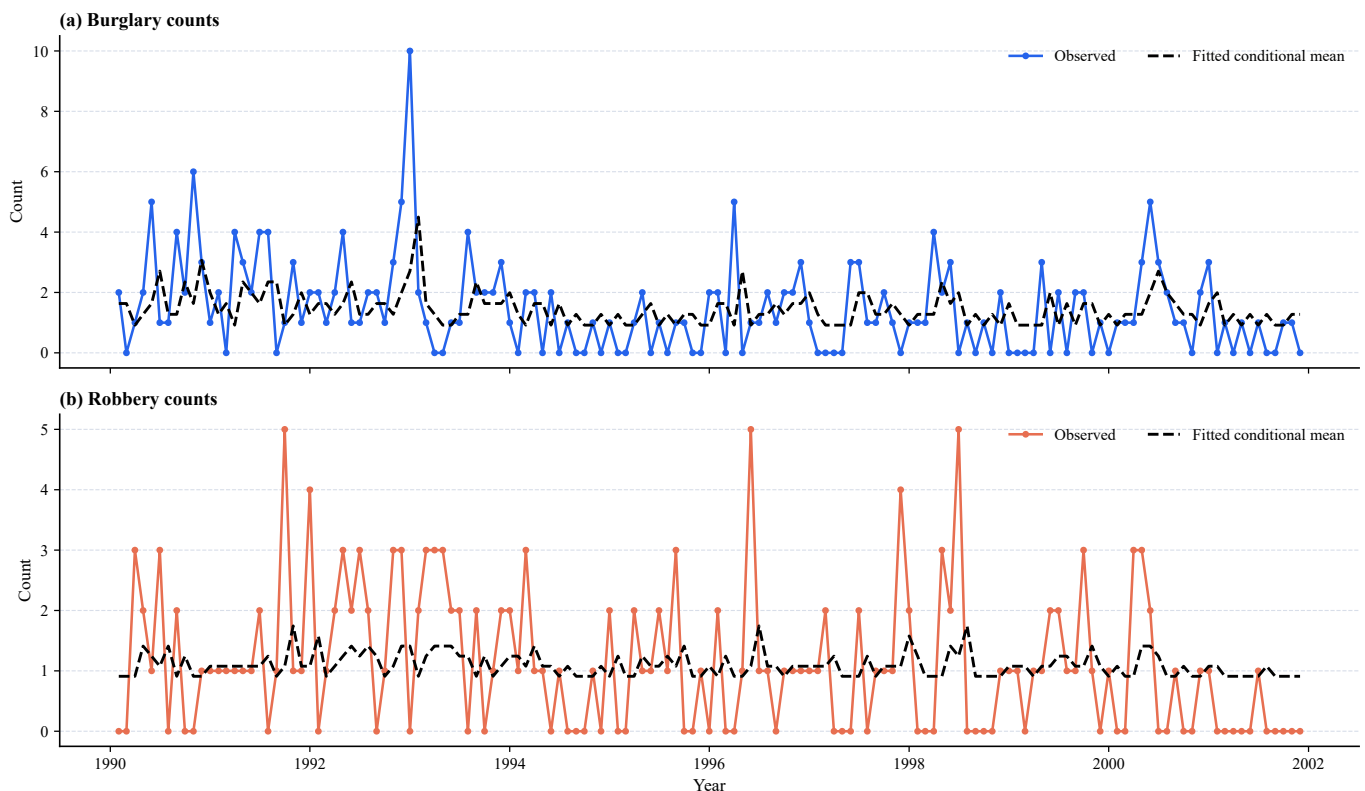


Figure A7. Observed and fitted values for the selected BINAR(1)-II model with MEL estimation. The fitted values represent one-step-ahead conditional means.

References

1. Steutel, F.W.; Van Harn, K. Discrete analogues of self-decomposability and stability. *Ann. Probab.* **1979**, *7*, 893–899.
2. McKenzie, E. Some simple models for discrete variate time series. *Water Resour. Bull.* **1985**, *21*, 645–650.
3. Al-Osh, M.A.; Alzaid, A.A. First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **1987**, *8*, 261–275.
4. Latour, A. The multivariate GINAR(p) process. *Adv. Appl. Probab.* **1997**, *29*, 228–248.
5. Pedeli, X.; Karlis, D. A bivariate INAR(1) process with application. *Stat. Model.* **2011**, *11*, 325–349.
6. Pedeli, X.; Karlis, D. On estimation of the bivariate Poisson INAR process. *Commun. Stat. Simul. Comput.* **2013**, *42*, 514–533.
7. Liu, Y.; Wang, D.; Zhang, H.; Shi, N. Bivariate zero-truncated Poisson INAR(1) process. *J. Korean Stat. Soc.* **2016**, *45*, 260–275.
8. Yang, K.; Zhao, Y.; Li, H.; Wang, D. On bivariate threshold Poisson integer-valued autoregressive processes. *Metrika* **2023**, *86*, 931–963. <https://doi.org/10.1007/s00184-023-00899-0>.
9. Li, H.; Yang, K.; Zhao, S.; Wang, D. First-order random coefficients integer-valued threshold autoregressive processes. *AStA Adv. Stat. Anal.* **2018**, *102*, 305–331.
10. Li, H.; Wang, H.; Yang, K.; Sun, J.; Liu, Y. A nonparametric Bayesian analysis for meningococcal disease counts based on integer-valued threshold time series models. *Commun. Stat. Theory Methods* **2022**, *52*, 8223–8240.
11. Buteikis, A.; Leipus, R. A copula-based bivariate integer-valued autoregressive process with application. *Mod. Stoch. Theory Appl.* **2019**, *6*, 227–249.
12. Karlis, D.; Pedeli, X. Flexible bivariate INAR(1) processes using copulas. *Commun. Stat. Theory Methods* **2013**, *42*, 723–740.
13. Owen, A. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **1988**, *75*, 237–249.
14. Owen, A. Empirical likelihood ratio confidence regions in time series models. *Ann. Stat.* **1990**, *18*, 90–120.
15. Owen, A. Empirical likelihood for linear models. *Ann. Stat.* **1991**, *19*, 1725–1747.
16. Chuang, C.S.; Chan, N.H. Empirical likelihood for autoregressive models with applications to unstable time series. *Stat. Sin.* **2002**, *12*, 387–407.
17. Zhu, F.; Wang, D. Empirical likelihood for linear and log-linear INGARCH models. *J. Korean Stat. Soc.* **2015**, *44*, 150–160.
18. Ding, X.; Wang, D. Empirical likelihood inference for INAR(1) model with explanatory variables. *J. Korean Stat. Soc.* **2016**, *45*, 623–632.
19. Zhang, H.; Wang, D.; Zhu, F. The empirical likelihood for first-order random coefficient integer-valued autoregressive processes. *Commun. Stat. Theory Methods* **2011**, *40*, 492–509.

20. Zhang, J.; Li, B.; Wang, Y.; et al. An empirical likelihood-based unified test for the integer-valued AR(1) models. *J. Stat. Plan. Inference* **2024**, 232, 106149.
21. Liu, Y.; Xiao, Y. Confidence regions of parameters for BINAR(1) process based on empirical likelihood methods. *J. Jilin Univ. Sci. Ed.* **2022**, 60, 1335–1341.
22. ICPSR. *Crime Incidents in Pittsburgh, 1990–2001*; ICPSR: Ann Arbor, MI, USA, 2006. <https://doi.org/10.3886/ICPSR04545.v1>.
23. Klimko, L.A.; Nelson, P.I. On conditional least squares estimation for stochastic processes. *Ann. Statist.* **1978**, 6, 629–642.
24. Kocherlakota, S.; Kocherlakota, K. *Bivariate Discrete Distributions*; CRC Press: New York, NY, USA, 1992.
25. Boucher, J.P.; Denuit, M.; Guillen, M. Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. *Variance* **2008**, 2, 135–162.
26. Hall, P.; Heyde, C.C. *Martingale Limit Theory and Its Application*; Academic Press: New York, NY, USA, 1980.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.