

## Additive outliers in INAR(1) models

Mátyás Barczy · Márton Ispány · Gyula Pap ·  
Manuel Scotto · Maria Eduarda Silva

Received: 28 February 2010 / Revised: 22 June 2011  
© Springer-Verlag 2011

**Abstract** In this paper the integer-valued autoregressive model of order one, contaminated with additive outliers is studied in some detail. Moreover, parameter estimation is also addressed. Supposing that the timepoints of the outliers are known but their sizes are unknown, we prove that the conditional least squares (CLS) estimators of the offspring and innovation means are strongly consistent. In contrast, however, the CLS estimators of the outliers' sizes are not strongly consistent, although they converge to a random limit with probability 1. We also prove that the joint CLS estimator of the offspring and innovation means is asymptotically normal. Conditionally on the values of the process at the timepoints neighboring to the outliers' occurrences, the joint CLS estimator of the sizes of the outliers is also asymptotically normal.

---

M. Barczy (✉) · M. Ispány  
Faculty of Informatics, University of Debrecen, Pf. 12,  
Debrecen 4010, Hungary  
e-mail: barczy.matyas@inf.unideb.hu

M. Ispány  
e-mail: ispany.marton@inf.unideb.hu

G. Pap  
University of Szeged, Bolyai Institute, Aradi vértanúk tere 1, Szeged 6720, Hungary  
e-mail: papgy@math.u-szeged.hu

M. Scotto  
Departamento de Matemática, Campus Universitário de Santiago, Universidade de Aveiro,  
3810-193 Aveiro, Portugal  
e-mail: mscotto@ua.pt

M. E. Silva  
Faculdade de Economia, Universidade do Porto, Rua Dr. Roberto Frias s/n, 4200 464 Porto, Portugal  
e-mail: mesilva@fep.up.pt

**Keywords** Integer-valued autoregressive models · Additive outliers · Conditional least squares estimators · Strong consistency · Conditional asymptotic normality

**Mathematics Subject Classification (2000)** 60J80 · 62F12

## 1 Introduction

Recently, there has been considerable interest in integer-valued time series models and a sizeable volume of work is now available in specialized monographs (e.g., Cameron and Trivedi 1998) and review papers (e.g., Weiß 2008). Several integer-valued time series models were proposed in the literature, we only mention the INteger-valued AutoRegressive model of order  $p$  (INAR( $p$ )). This was first introduced by McKenzie (1985) and Al-Osh and Alzaid (1987) for the case  $p = 1$ . The INAR( $p$ ) models have been investigated and extended by several authors, see, e.g., Du and Li (1991), Latour (1988) and for a short survey, see also the introduction of Barczy et al. (2009).

Moreover, topics of major current interest in time series modeling are to detect outliers in sample data and to investigate the impact of outliers on the estimation of conventional ARIMA models. Fox (1972) introduced the notion of additive and innovative outliers and proposed the use of maximum likelihood ratio test to detect them. Chang and Chen (1988) extended Fox's results to ARIMA models and proposed a likelihood ratio test and an iterative procedure for detecting outliers and estimating the model parameters. Some generalizations were obtained by Tsay (1988) for the detection of level shifts and temporary changes. Abraham and Chuang (1993) applied the EM algorithm to the estimation of outliers. Other useful references for outlier detection and estimation in time series models can be found, e.g., in Barczy et al. (2009). It is worth mentioning that all the references above are about continuous-valued processes.

A general motivation for studying outliers for integer-valued time series can be the fact that it may often difficult to remove outliers in the integer-valued case, and hence an important and interesting problem, which has not yet been addressed, is to investigate the impact of outliers on the parameter estimation of series of counts which are represented through integer-valued autoregressive models. This paper aims at giving a contribution towards this direction. A more specialized motivation is the possibility of potential applications, for example in the field of statistical process control (a good description of this topic can be found in Montgomery (2005, Chapter 4, Section 3.7)). In this paper we consider the problem of Conditional Least Squares (CLS) estimation of some parameters of the INAR(1) model contaminated with additive outliers starting from a general initial distribution (having finite second or third moments). We suppose that the timepoints of the outliers are known, but their sizes are unknown. Under the assumption that the second moment of the innovation distribution is finite, we prove that the CLS estimators of the means of the offspring and innovation distributions are strongly consistent, but the joint CLS estimator of the sizes of the outliers is not strongly consistent; nevertheless, it converges to a random limit with probability 1. This random limit depends on the values of the process at the outliers' timepoints and also on the values at the preceding and following timepoints. Under the assumption that the third moment of the innovation distribution is finite, we prove that the joint

CLS estimator of the means of the offspring and innovation distributions is asymptotically normal with the same asymptotic variance as in the case when there are no outliers. Conditionally on the values of the process at the timepoints neighboring to the outliers' occurrences, the joint CLS estimator of the sizes of the outliers is also asymptotically normal. We calculate its asymptotic covariance matrix as well. In this paper we consider one or two additive outliers for INAR(1) models, the general case of finitely many additive outliers may be handled in a similar way, but we renounce to consider it.

The rest of the paper is organized as follows. Section 2 provides a background description of basic theoretical results related to the asymptotic behavior of CLS estimators for the INAR(1) model. In Sect. 3 we consider INAR(1) models contaminated with one or two additive outliers. The cases of one outlier and two outliers are handled separately. Moreover, in case of two outliers we distinguish two subcases, namely, neighbouring or not neighbouring outliers' timepoints. However, we will formulate our results only in the case of two not neighbouring outliers' timepoints estimating the mean of the offspring and innovation distributions and the outliers' sizes. The full presentation of our (other) results can be found in our Arxiv preprint Barczy et al. (2009). In Sect. 4 we give a proof of our results in the above mentioned case.

In a companion paper we examine the INAR(1) model contaminated with one or two innovational outliers, see Barczy et al. (2010) and also Barczy et al. (2009, Section 4).

## 2 The INAR(1) model

Let  $\mathbb{Z}_+$  and  $\mathbb{N}$  denote the set of non-negative integers and positive integers, respectively. Every random variable will be defined on a fixed probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

**Definition 1** Let  $(\varepsilon_k)_{k \in \mathbb{N}}$  be an independent and identically distributed (i.i.d.) sequence of non-negative integer-valued random variables. An INAR(1) time series model is a stochastic process  $(X_k)_{k \in \mathbb{Z}_+}$  satisfying the recursive equation

$$X_k = \sum_{j=1}^{X_{k-1}} \xi_{k,j} + \varepsilon_k, \quad k \in \mathbb{N}, \quad (1)$$

where for all  $k \in \mathbb{N}$ ,  $(\xi_{k,j})_{j \in \mathbb{N}}$  is a sequence of i.i.d. Bernoulli random variables with mean  $\alpha \in [0, 1]$  such that these sequences are mutually independent and independent of the sequence  $(\varepsilon_\ell)_{\ell \in \mathbb{N}}$ , and  $X_0$  is a non-negative integer-valued random variable independent of the sequences  $(\xi_{k,j})_{j \in \mathbb{N}}$ ,  $k \in \mathbb{N}$ , and  $(\varepsilon_\ell)_{\ell \in \mathbb{N}}$ .

We note that the INAR(1) model in (1) can be written in another way using the binomial thinning operator  $\alpha \circ$  (due to Steutel and van Harn 1979), namely,  $X_k = \alpha \circ X_{k-1} + \varepsilon_k$ ,  $k \in \mathbb{N}$ .

In the sequel we will assume that  $\alpha \in (0, 1)$ ,  $\mathbb{E}X_0^2 < \infty$  and that  $\mathbb{E}\varepsilon_1^2 < \infty$ ,  $\mathbb{P}(\varepsilon_1 \neq 0) > 0$ . In this case, it is well-known (e.g., Barczy et al. 2009, Lemma 5.1) that there exists a unique stationary distribution of the INAR(1) model in (1). Let us denote the mean and variance of  $\varepsilon_1$  by  $\mu_\varepsilon$  and  $\sigma_\varepsilon^2$ , respectively. Clearly,

$0 < \mu_\varepsilon < \infty$ . We denote by  $\mathcal{F}_k^X$  the  $\sigma$ -algebra generated by the random variables  $X_0, X_1, \dots, X_k$ .

We only consider the joint CLS estimation of  $\alpha$  and  $\mu_\varepsilon$  (details for estimating only  $\alpha$  can be found, e.g., in Barczy et al. 2009). Clearly, for all  $k \in \mathbb{N}$ ,  $\mathbb{E}(X_k | \mathcal{F}_{k-1}^X) = \alpha X_{k-1} + \mu_\varepsilon$ , and thus

$$\sum_{k=1}^n (X_k - \mathbb{E}(X_k | \mathcal{F}_{k-1}^X))^2 = \sum_{k=1}^n (X_k - \alpha X_{k-1} - \mu_\varepsilon)^2, \quad n \in \mathbb{N}. \quad (2)$$

For all  $n \in \mathbb{N}$ , a CLS estimator  $(\hat{\alpha}_n, \hat{\mu}_{\varepsilon,n})$  for the parameter  $(\alpha, \mu_\varepsilon) \in (0, 1) \times (0, \infty)$  can be obtained by minimizing the sum of squares (2) with respect to  $(\alpha, \mu_\varepsilon) \in \mathbb{R}^2$ . One may prove that asymptotically as  $n \rightarrow \infty$ , a unique CLS estimator  $(\hat{\alpha}_n, \hat{\mu}_{\varepsilon,n})$  exists with probability one and

$$\hat{\alpha}_n = \frac{n \sum_{k=1}^n X_{k-1} X_k - (\sum_{k=1}^n X_{k-1}) (\sum_{k=1}^n X_k)}{n \sum_{k=1}^n X_{k-1}^2 - (\sum_{k=1}^n X_{k-1})^2},$$

$$\hat{\mu}_{\varepsilon,n} = \frac{(\sum_{k=1}^n X_{k-1}^2) (\sum_{k=1}^n X_k) - (\sum_{k=1}^n X_{k-1}) (\sum_{k=1}^n X_{k-1} X_k)}{n \sum_{k=1}^n X_{k-1}^2 - (\sum_{k=1}^n X_{k-1})^2},$$

hold asymptotically as  $n \rightarrow \infty$  with probability one, see, e.g., Hall and Heyde (1980, formulae (6.36) and (6.37)). Hereafter by the expression ‘a property holds asymptotically as  $n \rightarrow \infty$  with probability one’ we mean that there exists an event  $S \in \mathcal{A}$  such that  $\mathbb{P}(S) = 1$  and for all  $\omega \in S$  there exists an  $n(\omega) \in \mathbb{N}$  such that the property in question holds for all  $n \geq n(\omega)$ .

It is well-known that  $(\hat{\alpha}_n, \hat{\mu}_{\varepsilon,n})$  is a strongly consistent estimator of  $(\alpha, \mu_\varepsilon)$  as  $n \rightarrow \infty$  for all  $(\alpha, \mu_\varepsilon) \in (0, 1) \times (0, \infty)$ , see, e.g., Hall and Heyde (1980, Section 6.3). Moreover, if  $\mathbb{E}X_0^3 < \infty$  and  $\mathbb{E}\varepsilon_1^3 < \infty$ , by Hall and Heyde (1980, formula (6.44)),

$$\begin{bmatrix} \sqrt{n}(\hat{\alpha}_n - \alpha) \\ \sqrt{n}(\hat{\mu}_{\varepsilon,n} - \mu_\varepsilon) \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, B_{\alpha,\varepsilon}\right) \quad \text{as } n \rightarrow \infty,$$

where

$$B_{\alpha,\varepsilon} := \begin{bmatrix} \mathbb{E}\tilde{X}^2 & \mathbb{E}\tilde{X} \\ \mathbb{E}\tilde{X} & 1 \end{bmatrix}^{-1} A_{\alpha,\varepsilon} \begin{bmatrix} \mathbb{E}\tilde{X}^2 & \mathbb{E}\tilde{X} \\ \mathbb{E}\tilde{X} & 1 \end{bmatrix}^{-1}$$

$$= \frac{1}{(\text{Var } \tilde{X})^2} \begin{bmatrix} 1 & -\mathbb{E}\tilde{X} \\ -\mathbb{E}\tilde{X} & \mathbb{E}\tilde{X}^2 \end{bmatrix} A_{\alpha,\varepsilon} \begin{bmatrix} 1 & -\mathbb{E}\tilde{X} \\ -\mathbb{E}\tilde{X} & \mathbb{E}\tilde{X}^2 \end{bmatrix}, \quad (3)$$

$$A_{\alpha,\varepsilon} := \alpha(1 - \alpha) \begin{bmatrix} \mathbb{E}\tilde{X}^3 & \mathbb{E}\tilde{X}^2 \\ \mathbb{E}\tilde{X}^2 & \mathbb{E}\tilde{X} \end{bmatrix} + \sigma_\varepsilon^2 \begin{bmatrix} \mathbb{E}\tilde{X}^2 & \mathbb{E}\tilde{X} \\ \mathbb{E}\tilde{X} & 1 \end{bmatrix}, \quad (4)$$

and  $\tilde{X}$  denotes a random variable with the unique stationary distribution of the INAR(1) model in (1). Further, we note that

$$\mathbb{E}\tilde{X} = \frac{\mu_\varepsilon}{1-\alpha}, \quad (5)$$

$$\mathbb{E}\tilde{X}^2 = \frac{\sigma_\varepsilon^2 + \alpha\mu_\varepsilon}{1-\alpha^2} + \frac{\mu_\varepsilon^2}{(1-\alpha)^2}, \quad (6)$$

see, e.g., the Appendix of Barczy et al. (2009) (where one can also find an explicit expression for  $\mathbb{E}\tilde{X}^3$ ). Finally, we recall that, by ergodic theorems (see, e.g., Bhattacharya and Waymire 1990, Sect. II, Theorem 9.4d),

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = \mathbb{E}\tilde{X}\right) = 1, \quad (7)$$

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k^2 = \mathbb{E}\tilde{X}^2\right) = 1, \quad (8)$$

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_{k-1}X_k = \alpha\mathbb{E}\tilde{X}^2 + \mu_\varepsilon\mathbb{E}\tilde{X}\right) = 1. \quad (9)$$

### 3 The INAR(1) model with additive outliers

For all  $k, \ell \in \mathbb{Z}_+$ , let

$$\delta_{k,\ell} := \begin{cases} 1 & \text{if } k = \ell, \\ 0 & \text{if } k \neq \ell. \end{cases}$$

We introduce, below, the INAR(1) model contaminated with additive outliers.

**Definition 2** A stochastic process  $(Y_k)_{k \in \mathbb{Z}_+}$  is called an INAR(1) model contaminated with finitely many additive outliers if

$$Y_k = X_k + \sum_{i=1}^I \delta_{k,s_i} \theta_i, \quad k \in \mathbb{Z}_+,$$

where  $(X_k)_{k \in \mathbb{Z}_+}$  is an INAR(1) process given by (1) with  $\alpha \in (0, 1)$ ,  $\mathbb{E}X_0^2 < \infty$ ,  $\mathbb{E}\varepsilon_1^2 < \infty$ ,  $\mathbb{P}(\varepsilon_1 \neq 0) > 0$ , and  $I \in \mathbb{N}$ ,  $s_i, \theta_i \in \mathbb{N}$ ,  $i = 1, \dots, I$  such that  $s_i \neq s_j$  if  $i \neq j$ ,  $i, j = 1, \dots, I$ .

Notice that  $\theta_i$ ,  $i = 1, \dots, I$ , represents the  $i$ th additive outlier's size and  $\delta_{k,s_i}$  is an impulse taking the value 1 if  $k = s_i$  and 0 otherwise. Roughly speaking, an additive outlier can be interpreted as a measurement error at time  $s_i$ ,  $i = 1, \dots, I$ , or as an impulse due to some unspecified exogenous source. Note also that  $Y_0 = X_0$ . Let  $\mathcal{F}_k^Y$  be the  $\sigma$ -algebra generated by the random variables  $Y_0, Y_1, \dots, Y_k$ .

In what follows we formulate our results *only* in the case of two not neighbouring outliers' timepoints estimating the mean of the offspring and innovation distributions and the outliers' sizes. Our other results are presented in Barczy et al. (2009).

First we present a formula for  $E(Y_k | \mathcal{F}_{k-1}^Y)$  in case of  $I = 2$  and arbitrary known timepoints  $s_1, s_2 \in \mathbb{N}$ ,  $s_1 \neq s_2$ . Since  $Y_k = X_k + \delta_{k,s_1}\theta_1 + \delta_{k,s_2}\theta_2$ ,  $k \in \mathbb{Z}_+$ , we get for all  $k \in \mathbb{N}$ ,

$$\begin{aligned} E(Y_k | \mathcal{F}_{k-1}^Y) &= \alpha X_{k-1} + \mu_\varepsilon + \delta_{k,s_1}\theta_1 + \delta_{k,s_2}\theta_2 \\ &= \alpha Y_{k-1} + \mu_\varepsilon + (-\alpha\delta_{k-1,s_1} + \delta_{k,s_1})\theta_1 + (-\alpha\delta_{k-1,s_2} + \delta_{k,s_2})\theta_2. \end{aligned} \quad (10)$$

In the sequel we also suppose that  $s_1 < s_2 - 1$ , i.e., the timepoints  $s_1$  and  $s_2$  are not neighbouring. Then, it follows by (10) that for all  $n \geq s_2 + 1$ ,  $n \in \mathbb{N}$ ,

$$\begin{aligned} \sum_{k=1}^n (Y_k - E(Y_k | \mathcal{F}_{k-1}^Y))^2 \\ = \sum_{\substack{k=1 \\ k \notin \{s_1, s_1+1, s_2, s_2+1\}}}^n (Y_k - \alpha Y_{k-1} - \mu_\varepsilon)^2 \\ + (Y_{s_1} - \alpha Y_{s_1-1} - \mu_\varepsilon - \theta_1)^2 + (Y_{s_1+1} - \alpha Y_{s_1} - \mu_\varepsilon + \alpha\theta_1)^2 \\ + (Y_{s_2} - \alpha Y_{s_2-1} - \mu_\varepsilon - \theta_2)^2 + (Y_{s_2+1} - \alpha Y_{s_2} - \mu_\varepsilon + \alpha\theta_2)^2. \end{aligned} \quad (11)$$

By minimizing the sum of squares (11) with respect to  $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in \mathbb{R}^4$ , a CLS estimator  $(\hat{\alpha}_n^\dagger, \hat{\mu}_{\varepsilon,n}^\dagger, \hat{\theta}_{1,n}^\dagger, \hat{\theta}_{2,n}^\dagger)$  for the parameter  $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0, 1) \times (0, \infty) \times \mathbb{N}^2$  can be obtained for all  $n \geq s_2 + 1$ . In Sect. 4 we prove that asymptotically as  $n \rightarrow \infty$ , a unique CLS estimator  $(\hat{\alpha}_n^\dagger, \hat{\mu}_{\varepsilon,n}^\dagger, \hat{\theta}_{1,n}^\dagger, \hat{\theta}_{2,n}^\dagger)$  exists with probability one.

The next result shows that  $\hat{\alpha}_n^\dagger$  is a strongly consistent estimator of  $\alpha$ ,  $\hat{\mu}_{\varepsilon,n}^\dagger$  is a strongly consistent estimator of  $\mu_\varepsilon$ , whereas  $\hat{\theta}_{1,n}^\dagger$  and  $\hat{\theta}_{2,n}^\dagger$  fail to be strongly consistent estimators of  $\theta_1$  and  $\theta_2$ , respectively.

**Theorem 1** Consider the CLS estimators  $(\hat{\alpha}_n^\dagger, \hat{\mu}_{\varepsilon,n}^\dagger, \hat{\theta}_{1,n}^\dagger, \hat{\theta}_{2,n}^\dagger)_{n \in \mathbb{N}}$  of the parameter  $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0, 1) \times (0, \infty) \times \mathbb{N}^2$ . Then the sequences  $(\hat{\alpha}_n^\dagger)_{n \in \mathbb{N}}$  and  $(\hat{\mu}_{\varepsilon,n}^\dagger)_{n \in \mathbb{N}}$  are strongly consistent for all  $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0, 1) \times (0, \infty) \times \mathbb{N}^2$ , i.e.,

$$P\left(\lim_{n \rightarrow \infty} \hat{\alpha}_n^\dagger = \alpha\right) = 1 \quad \forall (\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0, 1) \times (0, \infty) \times \mathbb{N}^2, \quad (12)$$

$$P\left(\lim_{n \rightarrow \infty} \hat{\mu}_{\varepsilon,n}^\dagger = \mu_\varepsilon\right) = 1 \quad \forall (\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0, 1) \times (0, \infty) \times \mathbb{N}^2, \quad (13)$$

whereas the sequences  $(\hat{\theta}_{1,n}^\dagger)_{n \in \mathbb{N}}$  and  $(\hat{\theta}_{2,n}^\dagger)_{n \in \mathbb{N}}$  are not strongly consistent for any  $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0, 1) \times (0, \infty) \times \mathbb{N}^2$ , namely,

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \hat{\theta}_{i,n}^\dagger = Y_{s_i} - \frac{\alpha}{1 + \alpha^2} (Y_{s_i-1} + Y_{s_i+1}) - \frac{1 - \alpha}{1 + \alpha^2} \mu_\varepsilon \right) = 1, \quad i = 1, 2, \quad (14)$$

for all  $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0, 1) \times (0, \infty) \times \mathbb{N}^2$ .

*Proof* The proof can be found in Sect. 4.  $\square$

The asymptotic distributions of the CLS estimators are given in the next theorem.

**Theorem 2** Under the additional assumptions  $\mathbb{E}X_0^3 < \infty$  and  $\mathbb{E}\varepsilon_1^3 < \infty$ , we have

$$\begin{bmatrix} \sqrt{n}(\hat{\alpha}_n^\dagger - \alpha) \\ \sqrt{n}(\hat{\mu}_{\varepsilon,n}^\dagger - \mu_\varepsilon) \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, B_{\alpha,\varepsilon} \right) \quad \text{as } n \rightarrow \infty \quad (15)$$

with  $B_{\alpha,\varepsilon}$  defined as in (3). Moreover, conditionally on the values  $Y_{s_1-1}$ ,  $Y_{s_2-1}$  and  $Y_{s_1+1}$ ,  $Y_{s_2+1}$ ,

$$\begin{bmatrix} \sqrt{n}(\hat{\theta}_{1,n}^\dagger - \lim_{k \rightarrow \infty} \hat{\theta}_{1,k}^\dagger) \\ \sqrt{n}(\hat{\theta}_{2,n}^\dagger - \lim_{k \rightarrow \infty} \hat{\theta}_{2,k}^\dagger) \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, C_{\alpha,\varepsilon} B_{\alpha,\varepsilon} C_{\alpha,\varepsilon}^\top \right) \quad \text{as } n \rightarrow \infty, \quad (16)$$

where

$$C_{\alpha,\varepsilon} := \frac{1}{(1 + \alpha^2)^2} \begin{bmatrix} (\alpha^2 - 1)(Y_{s_1-1} + Y_{s_1+1}) + (1 + 2\alpha - \alpha^2)\mu_\varepsilon & (\alpha - 1)(1 + \alpha^2) \\ (\alpha^2 - 1)(Y_{s_2-1} + Y_{s_2+1}) + (1 + 2\alpha - \alpha^2)\mu_\varepsilon & (\alpha - 1)(1 + \alpha^2) \end{bmatrix}.$$

*Proof* The proof can be found in Sect. 4.  $\square$

## 4 Proofs

We retain the notations introduced earlier. Moreover, let us define

$$\mathbf{Y}_n := (Y_0, Y_1, \dots, Y_n), \quad \mathbf{y}_n := (y_0, y_1, \dots, y_n),$$

$$\mathbf{Y}_n(\omega) := (Y_0(\omega), Y_1(\omega), \dots, Y_n(\omega))$$

for all  $n \in \mathbb{N}$ ,  $y_0, \dots, y_n \in \mathbb{R}$  and  $\omega \in \Omega$ . First we give a proof that asymptotically as  $n \rightarrow \infty$ , a unique CLS estimator  $(\hat{\alpha}_n^\dagger, \hat{\mu}_{\varepsilon,n}^\dagger, \hat{\theta}_{1,n}^\dagger, \hat{\theta}_{2,n}^\dagger)$  exists with probability one. Motivated by (11), for all  $n \geq s_2 + 1$ ,  $n \in \mathbb{N}$ , we define the function  $Q_n^\dagger: \mathbb{R}^{n+1} \times \mathbb{R}^4 \rightarrow \mathbb{R}$ ,

$$\begin{aligned}
Q_n^\dagger(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) &:= \sum_{\substack{k=1 \\ k \notin \{s_1, s_1+1, s_2, s_2+1\}}}^n (y_k - \alpha' y_{k-1} - \mu'_\varepsilon)^2 \\
&+ (y_{s_1} - \alpha' y_{s_1-1} - \mu'_\varepsilon - \theta'_1)^2 + (y_{s_1+1} - \alpha' y_{s_1} - \mu'_\varepsilon + \alpha' \theta'_1)^2 \\
&+ (y_{s_2} - \alpha' y_{s_2-1} - \mu'_\varepsilon - \theta'_2)^2 + (y_{s_2+1} - \alpha' y_{s_2} - \mu'_\varepsilon + \alpha' \theta'_2)^2,
\end{aligned}$$

for all  $\mathbf{y}_n \in \mathbb{R}^{n+1}$ ,  $\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2 \in \mathbb{R}$ . By definition, for all  $n \geq s_2 + 1$ , a CLS estimator for the parameter  $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0, 1) \times (0, \infty) \times \mathbb{N}^2$  is a measurable function

$$(\hat{\alpha}_n^\dagger, \hat{\mu}_{\varepsilon,n}^\dagger, \hat{\theta}_{1,n}^\dagger, \hat{\theta}_{2,n}^\dagger) : S_n \rightarrow \mathbb{R}^4$$

such that

$$\begin{aligned}
&Q_n^\dagger(\mathbf{y}_n; \hat{\alpha}_n^\dagger(\mathbf{y}_n), \hat{\mu}_{\varepsilon,n}^\dagger(\mathbf{y}_n), \hat{\theta}_{1,n}^\dagger(\mathbf{y}_n), \hat{\theta}_{2,n}^\dagger(\mathbf{y}_n)) \\
&= \inf_{(\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \in \mathbb{R}^4} Q_n^\dagger(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \quad \forall \mathbf{y}_n \in S_n,
\end{aligned}$$

where  $S_n$  is suitable subset of  $\mathbb{R}^{n+1}$  (defined in the proof of Lemma 1). We note that we do not define the CLS estimator  $(\hat{\alpha}_n^\dagger, \hat{\mu}_{\varepsilon,n}^\dagger, \hat{\theta}_{1,n}^\dagger, \hat{\theta}_{2,n}^\dagger)$  for all samples  $\mathbf{y}_n \in \mathbb{R}^{n+1}$ .

The next result is about the existence and uniqueness of  $(\hat{\alpha}_n^\dagger, \hat{\mu}_{\varepsilon,n}^\dagger, \hat{\theta}_{1,n}^\dagger, \hat{\theta}_{2,n}^\dagger)$ .

**Lemma 1** *There exist subsets  $S_n \subset \mathbb{R}^{n+1}$ ,  $n \geq \max(5, s_2 + 1)$  with the following properties:*

- (i) *there exists a unique CLS estimator  $(\hat{\alpha}_n^\dagger, \hat{\mu}_{\varepsilon,n}^\dagger, \hat{\theta}_{1,n}^\dagger, \hat{\theta}_{2,n}^\dagger) : S_n \rightarrow \mathbb{R}^4$ ,*
- (ii) *for all  $\mathbf{y}_n \in S_n$ ,  $(\hat{\alpha}_n^\dagger(\mathbf{y}_n), \hat{\mu}_{\varepsilon,n}^\dagger(\mathbf{y}_n), \hat{\theta}_{1,n}^\dagger(\mathbf{y}_n), \hat{\theta}_{2,n}^\dagger(\mathbf{y}_n))$  is the unique solution of the system of equations*

$$\begin{aligned}
\frac{\partial Q_n^\dagger}{\partial \alpha'}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) &= 0, & \frac{\partial Q_n^\dagger}{\partial \mu'_\varepsilon}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) &= 0, \\
\frac{\partial Q_n^\dagger}{\partial \theta'_1}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) &= 0, & \frac{\partial Q_n^\dagger}{\partial \theta'_2}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) &= 0,
\end{aligned} \tag{17}$$

- (iii)  $\mathbf{Y}_n \in S_n$  holds asymptotically as  $n \rightarrow \infty$  with probability one.

*Proof* For any fixed  $\mathbf{y}_n \in \mathbb{R}^{n+1}$ ,  $n \geq \max(5, s_2 + 1)$  and  $\alpha' \in \mathbb{R}$ , the quadratic function  $\mathbb{R}^3 \ni (\mu'_\varepsilon, \theta'_1, \theta'_2) \mapsto Q_n^\dagger(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$  can be written in the form

$$\begin{aligned}
Q_n^\dagger(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) &= \left( \begin{bmatrix} \mu'_\varepsilon \\ \theta'_1 \\ \theta'_2 \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right)^\top A_n(\alpha') \\
&\times \left( \begin{bmatrix} \mu'_\varepsilon \\ \theta'_1 \\ \theta'_2 \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right) + \hat{Q}_n^\dagger(\mathbf{y}_n; \alpha'),
\end{aligned}$$



where

$$t_n(\mathbf{y}_n; \alpha') := \begin{bmatrix} \sum_{k=1}^n (y_k - \alpha' y_{k-1}) \\ (1 + (\alpha')^2) y_{s_1} - \alpha' (y_{s_1-1} + y_{s_1+1}) \\ (1 + (\alpha')^2) y_{s_2} - \alpha' (y_{s_2-1} + y_{s_2+1}) \end{bmatrix},$$

$$\widehat{Q}_n^+(\mathbf{y}_n; \alpha') := \sum_{k=1}^n (y_k - \alpha' y_{k-1})^2 - t_n(\mathbf{y}_n; \alpha')^\top A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha'),$$

and the matrix

$$A_n(\alpha') := \begin{bmatrix} n & 1 - \alpha' & 1 - \alpha' \\ 1 - \alpha' & 1 + (\alpha')^2 & 0 \\ 1 - \alpha' & 0 & 1 + (\alpha')^2 \end{bmatrix}$$

is strictly positive definite for all  $n \geq 5$  and  $\alpha' \in \mathbb{R}$ .

The inverse matrix  $A_n(\alpha')^{-1}$  takes the form

$$\frac{1}{D_n(\alpha')} \begin{bmatrix} (1 + (\alpha')^2)^2 & -(1 - \alpha')(1 + (\alpha')^2) & -(1 - \alpha')(1 + (\alpha')^2) \\ -(1 - \alpha')(1 + (\alpha')^2) & n(1 + (\alpha')^2) - (1 - \alpha')^2 & (1 - \alpha')^2 \\ -(1 - \alpha')(1 + (\alpha')^2) & (1 - \alpha')^2 & n(1 + (\alpha')^2) - (1 - \alpha')^2 \end{bmatrix},$$

where

$$D_n(\alpha') := (1 + (\alpha')^2)((n - 2)(\alpha')^2 + 4\alpha' + n - 2).$$

The polynomial  $\mathbb{R} \ni \alpha' \mapsto D_n(\alpha')$  is of order 4 with leading coefficient  $n - 2$ .

We have  $\widehat{Q}_n^+(\mathbf{y}_n; \alpha') = R_n(\mathbf{y}_n; \alpha')/D_n(\alpha')$ , where  $\mathbb{R} \ni \alpha' \mapsto R_n(\mathbf{y}_n; \alpha')$  is a polynomial of order 6 with leading coefficient

$$\begin{aligned} c_n(\mathbf{y}_n) &:= (n - 2) \sum_{k=1}^n y_{k-1}^2 - \left( \sum_{k=1}^n y_{k-1} \right)^2 - (n - 1)(y_{s_1}^2 + y_{s_2}^2) \\ &\quad + 2(y_{s_1} + y_{s_2}) \sum_{k=1}^n y_{k-1} - 2y_{s_1} y_{s_2}. \end{aligned}$$

Let

$$\widehat{S}_n^+ := \left\{ \mathbf{y}_n \in \mathbb{R}^{n+1} : c_n(\mathbf{y}_n) > 0 \right\}.$$

For  $\mathbf{y}_n \in \widehat{S}_n^+$ , we have  $\lim_{|\alpha'| \rightarrow \infty} \widehat{Q}_n^+(\mathbf{y}_n; \alpha') = \infty$  and the continuous function  $\mathbb{R} \ni \alpha' \mapsto \widehat{Q}_n^+(\mathbf{y}_n; \alpha')$  attains its infimum. Consequently, for all  $n \geq 5$  there exists a CLS estimator  $(\widehat{\alpha}_n^\dagger, \widehat{\mu}_{\varepsilon, n}^\dagger, \widehat{\theta}_{1, n}^\dagger, \widehat{\theta}_{2, n}^\dagger) : \widehat{S}_n^+ \rightarrow \mathbb{R}^4$ , where

$$\widehat{Q}_n^\dagger(\mathbf{y}_n; \widehat{\alpha}_n^\dagger(\mathbf{y}_n)) = \inf_{\alpha' \in \mathbb{R}} \widehat{Q}_n^\dagger(\mathbf{y}_n; \alpha') \quad \forall \mathbf{y}_n \in \widehat{S}_n^\dagger,$$

$$\begin{bmatrix} \widehat{\mu}_{\varepsilon,n}^\dagger(\mathbf{y}_n) \\ \widehat{\theta}_{1,n}^\dagger(\mathbf{y}_n) \\ \widehat{\theta}_{2,n}^\dagger(\mathbf{y}_n) \end{bmatrix} = A_n(\widehat{\alpha}_n^\dagger(\mathbf{y}_n))^{-1} t_n(\mathbf{y}_n; \widehat{\alpha}_n^\dagger(\mathbf{y}_n)), \quad \mathbf{y}_n \in \widehat{S}_n^\dagger, \quad (18)$$

and for all  $\mathbf{y}_n \in \widehat{S}_n^\dagger$ ,  $(\widehat{\alpha}_n^\dagger(\mathbf{y}_n), \widehat{\mu}_{\varepsilon,n}^\dagger(\mathbf{y}_n), \widehat{\theta}_{1,n}^\dagger(\mathbf{y}_n), \widehat{\theta}_{2,n}^\dagger(\mathbf{y}_n))$  is a solution of the system of equations (17).

By (7) and (8), we get  $\mathbf{P}(\lim_{n \rightarrow \infty} n^{-2} c_n(\mathbf{Y}_n) = \text{Var } \widetilde{X}) = 1$ , where  $\widetilde{X}$  denotes a random variable with the unique stationary distribution of the INAR(1) model in (1). Hence  $\mathbf{Y}_n \in \widehat{S}_n^\dagger$  holds asymptotically as  $n \rightarrow \infty$  with probability one.

Now we turn to find sets  $S_n \subset \widehat{S}_n^\dagger$ ,  $n \geq \max(5, s_2 + 1)$  such that the system of equations (17) has a unique solution with respect to  $(\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$  for all  $\mathbf{y}_n \in S_n$ . Let us denote by  $H_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$  the  $(4 \times 4)$  Hessian matrix of  $Q_n^\dagger$  consisting of the second order partial derivatives of  $Q_n^\dagger$  with respect to the variables  $\alpha', \mu'_\varepsilon, \theta'_1$  and  $\theta'_2$ . Further let  $\Delta_{i,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$  be its  $i$ -th order leading principal minor,  $i = 1, 2, 3, 4$ , and, for all  $n \geq \max(5, s_2 + 1)$ , let

$$S_n := \left\{ \mathbf{y}_n \in \widehat{S}_n^\dagger : \Delta_{i,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) > 0, \right. \\ \left. i = 1, 2, 3, 4, \forall (\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \in \mathbb{R}^4 \right\}.$$

By Berkovitz (2002, Theorem 3.3, Chapter III), the function  $\mathbb{R}^4 \ni (\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \mapsto Q_n^\dagger(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$  is strictly convex for all  $\mathbf{y}_n \in S_n$ . Since it was already proved that the system of equations (17) has a solution for all  $\mathbf{y}_n \in \widehat{S}_n^\dagger$ , we obtain that this solution is unique for all  $\mathbf{y}_n \in S_n$ .

One can also check that  $\mathbf{Y}_n \in S_n$  holds asymptotically as  $n \rightarrow \infty$  with probability one, see, Barczy et al. (2009, Lemma 3.6.1).  $\square$

By Lemma 1,  $(\widehat{\alpha}_n^\dagger(\mathbf{Y}_n), \widehat{\mu}_{\varepsilon,n}^\dagger(\mathbf{Y}_n), \widehat{\theta}_{1,n}^\dagger(\mathbf{Y}_n), \widehat{\theta}_{2,n}^\dagger(\mathbf{Y}_n))$  exists uniquely asymptotically as  $n \rightarrow \infty$  with probability one. In the sequel we will simply denote it by  $(\widehat{\alpha}_n^\dagger, \widehat{\mu}_{\varepsilon,n}^\dagger, \widehat{\theta}_{1,n}^\dagger, \widehat{\theta}_{2,n}^\dagger)$ .

**Proof of Theorem 1** The aim of the following discussion is to show that the sequences  $(\widehat{\theta}_{1,n}^\dagger - \theta_1)_{n \in \mathbb{N}}$  and  $(\widehat{\theta}_{2,n}^\dagger - \theta_2)_{n \in \mathbb{N}}$  are bounded with probability one. First we note that for all  $\mathbf{y}_n \in \mathbb{R}^{n+1}$  and  $(\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \in \mathbb{R}^4$ , we have

$$\begin{aligned} \frac{\partial Q_n^\dagger}{\partial \theta'_i}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \\ = -2(y_{s_i} - \alpha' y_{s_i-1} - \mu'_\varepsilon - \theta'_i) + 2\alpha'(y_{s_i+1} - \alpha' y_{s_i} - \mu'_\varepsilon + \alpha' \theta'_i), \quad i = 1, 2. \end{aligned} \quad (19)$$

By (17), (19) and Lemma 1, we get

$$\hat{\theta}_{i,n}^{\dagger} = Y_{s_i} - \frac{\hat{\alpha}_n^{\dagger}}{1 + (\hat{\alpha}_n^{\dagger})^2} (Y_{s_i-1} + Y_{s_i+1}) - \frac{1 - \hat{\alpha}_n^{\dagger}}{1 + (\hat{\alpha}_n^{\dagger})^2} \hat{\mu}_{\varepsilon,n}^{\dagger} \quad i = 1, 2. \quad (20)$$

By (18) and the explicit form of the inverse matrix  $A_n(\alpha')^{-1}$ , we obtain

$$\begin{bmatrix} \hat{\mu}_{\varepsilon,n}^{\dagger} \\ \hat{\theta}_{1,n}^{\dagger} \\ \hat{\theta}_{2,n}^{\dagger} \end{bmatrix} = \frac{1}{D_n(\hat{\alpha}_n^{\dagger})} \begin{bmatrix} G_n \\ H_n \\ J_n \end{bmatrix},$$

where

$$\begin{aligned} G_n &:= -(1 - \hat{\alpha}_n^{\dagger})(1 + (\hat{\alpha}_n^{\dagger})^2) \\ &\quad \times \left( (1 + (\hat{\alpha}_n^{\dagger})^2)(Y_{s_1} + Y_{s_2}) - \hat{\alpha}_n^{\dagger}(Y_{s_1-1} + Y_{s_1+1} + Y_{s_2-1} + Y_{s_2+1}) \right) \\ &\quad + (1 + (\hat{\alpha}_n^{\dagger})^2)^2 \sum_{k=1}^n (Y_k - \hat{\alpha}_n^{\dagger} Y_{k-1}), \\ H_n &:= (n(1 + (\hat{\alpha}_n^{\dagger})^2) - (1 - \hat{\alpha}_n^{\dagger})^2) \left( (1 + (\hat{\alpha}_n^{\dagger})^2) Y_{s_1} - \hat{\alpha}_n^{\dagger}(Y_{s_1-1} + Y_{s_1+1}) \right) \\ &\quad + (1 - \hat{\alpha}_n^{\dagger})^2 \left( (1 + (\hat{\alpha}_n^{\dagger})^2) Y_{s_2} - \hat{\alpha}_n^{\dagger}(Y_{s_2-1} + Y_{s_2+1}) \right) \\ &\quad - (1 - \hat{\alpha}_n^{\dagger})(1 + (\hat{\alpha}_n^{\dagger})^2) \sum_{k=1}^n (Y_k - \hat{\alpha}_n^{\dagger} Y_{k-1}), \\ J_n &:= (1 - \hat{\alpha}_n^{\dagger})^2 \left( (1 + (\hat{\alpha}_n^{\dagger})^2) Y_{s_1} - \hat{\alpha}_n^{\dagger}(Y_{s_1-1} + Y_{s_1+1}) \right) \\ &\quad + (n(1 + (\hat{\alpha}_n^{\dagger})^2) - (1 - \hat{\alpha}_n^{\dagger})^2) \left( (1 + (\hat{\alpha}_n^{\dagger})^2) Y_{s_2} - \hat{\alpha}_n^{\dagger}(Y_{s_2-1} + Y_{s_2+1}) \right) \\ &\quad - (1 - \hat{\alpha}_n^{\dagger})(1 + (\hat{\alpha}_n^{\dagger})^2) \sum_{k=1}^n (Y_k - \hat{\alpha}_n^{\dagger} Y_{k-1}). \end{aligned}$$

Using (7) and that for all  $p_i \in \mathbb{R}$ ,  $i = 0, \dots, 4$ ,

$$\sup_{x \in \mathbb{R}, n \geq 5} \frac{n(p_4 x^4 + p_3 x^3 + p_2 x^2 + p_1 x + p_0)}{(1 + x^2)((n - 2)x^2 + 4x + n - 2)} < \infty,$$

one can think it over that  $H_n/D_n(\hat{\alpha}_n^{\dagger})$ ,  $n \in \mathbb{N}$ , and  $J_n/D_n(\hat{\alpha}_n^{\dagger})$ ,  $n \in \mathbb{N}$ , are bounded with probability one, which yields also that the sequences  $(\hat{\theta}_{1,n}^{\dagger} - \theta_1)_{n \in \mathbb{N}}$  and  $(\hat{\theta}_{2,n}^{\dagger} - \theta_2)_{n \in \mathbb{N}}$  are bounded with probability one.

Again by Lemma 1 and Eqs. 17 we get that

$$\begin{bmatrix} \hat{\alpha}_n^{\dagger} \\ \hat{\mu}_{\varepsilon,n}^{\dagger} \end{bmatrix} = \begin{bmatrix} a_n & b_n \\ b_n & n \end{bmatrix}^{-1} \begin{bmatrix} c_n \\ d_n \end{bmatrix} \quad (21)$$

holds asymptotically as  $n \rightarrow \infty$  with probability one, where

$$a_n := \sum_{k=1}^n X_{k-1}^2 + (\theta_1 - \hat{\theta}_{1,n}^\dagger)(\theta_1 - \hat{\theta}_{1,n}^\dagger + 2X_{s_1}) + (\theta_2 - \hat{\theta}_{2,n}^\dagger)(\theta_2 - \hat{\theta}_{2,n}^\dagger + 2X_{s_2}),$$

$$b_n := \sum_{k=1}^n X_{k-1} + \theta_1 - \hat{\theta}_{1,n}^\dagger + \theta_2 - \hat{\theta}_{2,n}^\dagger,$$

$$c_n := \sum_{k=1}^n X_{k-1}X_k + (\theta_1 - \hat{\theta}_{1,n}^\dagger)(X_{s_1-1} + X_{s_1+1}) + (\theta_2 - \hat{\theta}_{2,n}^\dagger)(X_{s_2-1} + X_{s_2+1}),$$

$$d_n := \sum_{k=1}^n X_k + \theta_1 - \hat{\theta}_{1,n}^\dagger + \theta_2 - \hat{\theta}_{2,n}^\dagger.$$

Here we emphasize that the  $(2 \times 2)$ -matrix in (21) is invertible asymptotically as  $n \rightarrow \infty$  with probability one, since one can check that

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} \frac{a_n}{n} = \mathbf{E}\tilde{X}^2\right) = 1, \quad \mathbf{P}\left(\lim_{n \rightarrow \infty} \frac{b_n}{n} = \mathbf{E}\tilde{X}\right) = 1, \quad (22)$$

which implies that  $\mathbf{P}(\lim_{n \rightarrow \infty} \frac{1}{n^2}(na_n - b_n^2) = \mathbf{E}\tilde{X}^2 - (\mathbf{E}\tilde{X})^2 = \text{Var } \tilde{X}) = 1$ , and hence  $\mathbf{P}(\lim_{n \rightarrow \infty} (na_n - b_n^2) = \infty) = 1$ .

Further

$$\begin{bmatrix} \hat{\alpha}_n^\dagger - \alpha \\ \hat{\mu}_{\varepsilon,n}^\dagger - \mu_\varepsilon \end{bmatrix} = \begin{bmatrix} a_n & b_n \\ b_n & n \end{bmatrix}^{-1} \begin{bmatrix} e_n \\ f_n \end{bmatrix} \quad (23)$$

holds asymptotically as  $n \rightarrow \infty$  with probability one, where

$$\begin{aligned} e_n &:= \sum_{k=1}^n X_{k-1}(X_k - \alpha X_{k-1} - \mu_\varepsilon) \\ &\quad + (\theta_1 - \hat{\theta}_{1,n}^\dagger)(X_{s_1-1} + X_{s_1+1} - 2\alpha X_{s_1} - \mu_\varepsilon - \alpha(\theta_1 - \hat{\theta}_{1,n}^\dagger)) \\ &\quad + (\theta_2 - \hat{\theta}_{2,n}^\dagger)(X_{s_2-1} + X_{s_2+1} - 2\alpha X_{s_2} - \mu_\varepsilon - \alpha(\theta_2 - \hat{\theta}_{2,n}^\dagger)), \\ f_n &:= \sum_{k=1}^n (X_k - \alpha X_{k-1} - \mu_\varepsilon) + (1 - \alpha)(\theta_1 - \hat{\theta}_{1,n}^\dagger + \theta_2 - \hat{\theta}_{2,n}^\dagger). \end{aligned}$$

Then, using again (5), (7), (8), (9) and that the sequences  $(\hat{\theta}_{1,n}^\dagger - \theta_1)_{n \in \mathbb{N}}$  and  $(\hat{\theta}_{2,n}^\dagger - \theta_2)_{n \in \mathbb{N}}$  are bounded with probability one, we get

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} \frac{e_n}{n} = \alpha \mathbf{E}\tilde{X}^2 + \mu_\varepsilon \mathbf{E}\tilde{X} - \alpha \mathbf{E}\tilde{X}^2 - \mu_\varepsilon \mathbf{E}\tilde{X} = 0\right) = 1,$$

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} \frac{f_n}{n} = \mathbf{E}\tilde{X} - \alpha \mathbf{E}\tilde{X} - \mu_\varepsilon = 0\right) = 1.$$

Hence, by (23), we obtain

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \begin{bmatrix} \hat{\alpha}_n^\dagger - \alpha \\ \hat{\mu}_{\varepsilon,n}^\dagger - \mu_\varepsilon \end{bmatrix} = \begin{bmatrix} \mathbb{E}\tilde{X}^2 & \mathbb{E}\tilde{X} \\ \mathbb{E}\tilde{X} & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = 1,$$

which yields (12) and (13). Then (12), (13) and (20) imply (14).  $\square$

*Proof of Theorem 2* By (3) and (23), to prove (15) it is enough to show that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \begin{bmatrix} a_n & b_n \\ b_n & n \end{bmatrix} \begin{bmatrix} \mathbb{E}\tilde{X}^2 & \mathbb{E}\tilde{X} \\ \mathbb{E}\tilde{X} & 1 \end{bmatrix}\right) = 1, \quad (24)$$

$$\frac{1}{\sqrt{n}} \begin{bmatrix} e_n \\ f_n \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, A_{\alpha,\varepsilon}\right) \quad \text{as } n \rightarrow \infty, \quad (25)$$

where  $\tilde{X}$  is a random variable having the unique stationary distribution of the INAR(1) model in (1) and the  $(2 \times 2)$ -matrix  $A_{\alpha,\varepsilon}$  is defined in (4). By (22), we have (24). By formula (6.43) in Hall and Heyde (1980, Sect. 6.3),

$$\begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{k=1}^n X_{k-1}(X_k - \alpha X_{k-1} - \mu_\varepsilon) \\ \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - \alpha X_{k-1} - \mu_\varepsilon) \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, A_{\alpha,\varepsilon}\right) \quad \text{as } n \rightarrow \infty.$$

Hence using that the sequences  $(\hat{\theta}_{1,n}^\dagger - \theta_1)_{n \in \mathbb{N}}$  and  $(\hat{\theta}_{2,n}^\dagger - \theta_2)_{n \in \mathbb{N}}$  are bounded with probability one, by Slutsky's lemma, we get (25).

Now we turn to prove (16). Using the notation  $B_n^\dagger := (1 + (\hat{\alpha}_n^\dagger)^2)I_2$ , where  $I_2$  denotes the  $(2 \times 2)$  identity matrix, by (20), we have

$$\begin{bmatrix} \hat{\theta}_{1,n}^\dagger \\ \hat{\theta}_{2,n}^\dagger \end{bmatrix} = (B_n^\dagger)^{-1} \begin{bmatrix} (1 + (\hat{\alpha}_n^\dagger)^2)Y_{s_1} - \hat{\alpha}_n^\dagger(Y_{s_1-1} + Y_{s_1+1}) - (1 - \hat{\alpha}_n^\dagger)\hat{\mu}_{\varepsilon,n}^\dagger \\ (1 + (\hat{\alpha}_n^\dagger)^2)Y_{s_2} - \hat{\alpha}_n^\dagger(Y_{s_2-1} + Y_{s_2+1}) - (1 - \hat{\alpha}_n^\dagger)\hat{\mu}_{\varepsilon,n}^\dagger \end{bmatrix}$$

holds asymptotically as  $n \rightarrow \infty$  with probability one. Then Theorem 1 yields that

$\mathbb{P}(\lim_{n \rightarrow \infty} B_n^\dagger = (1 + \alpha^2)I_2 =: B^\dagger) = 1$ . By (14), we have

$$\begin{aligned} & \begin{bmatrix} \sqrt{n}(\hat{\theta}_{1,n}^\dagger - \lim_{k \rightarrow \infty} \hat{\theta}_{1,k}^\dagger) \\ \sqrt{n}(\hat{\theta}_{2,n}^\dagger - \lim_{k \rightarrow \infty} \hat{\theta}_{2,k}^\dagger) \end{bmatrix} \\ &= \sqrt{n}(B_n^\dagger)^{-1} \left( \begin{bmatrix} (1 + (\hat{\alpha}_n^\dagger)^2)Y_{s_1} - \hat{\alpha}_n^\dagger(Y_{s_1-1} + Y_{s_1+1}) - (1 - \hat{\alpha}_n^\dagger)\hat{\mu}_{\varepsilon,n}^\dagger \\ (1 + (\hat{\alpha}_n^\dagger)^2)Y_{s_2} - \hat{\alpha}_n^\dagger(Y_{s_2-1} + Y_{s_2+1}) - (1 - \hat{\alpha}_n^\dagger)\hat{\mu}_{\varepsilon,n}^\dagger \end{bmatrix} \right. \\ & \quad \left. - B_n^\dagger(B^\dagger)^{-1} \begin{bmatrix} (1 + \alpha^2)Y_{s_1} - \alpha(Y_{s_1-1} + Y_{s_1+1}) - (1 - \alpha)\mu_\varepsilon \\ (1 + \alpha^2)Y_{s_2} - \alpha(Y_{s_2-1} + Y_{s_2+1}) - (1 - \alpha)\mu_\varepsilon \end{bmatrix} \right), \end{aligned}$$

and hence

$$\begin{aligned}
 & \begin{bmatrix} \sqrt{n}(\hat{\theta}_{1,n}^\dagger - \lim_{k \rightarrow \infty} \hat{\theta}_{1,k}^\dagger) \\ \sqrt{n}(\hat{\theta}_{2,n}^\dagger - \lim_{k \rightarrow \infty} \hat{\theta}_{2,k}^\dagger) \end{bmatrix} \\
 &= \sqrt{n}(B_n^\dagger)^{-1} \begin{bmatrix} (\hat{\alpha}_n^\dagger + \alpha)Y_{s_1} - (Y_{s_1-1} + Y_{s_1+1}) + \hat{\mu}_{\varepsilon,n}^\dagger \alpha - 1 \\ (\hat{\alpha}_n^\dagger + \alpha)Y_{s_2} - (Y_{s_2-1} + Y_{s_2+1}) + \hat{\mu}_{\varepsilon,n}^\dagger \alpha - 1 \end{bmatrix} \begin{bmatrix} \hat{\alpha}_n^\dagger - \alpha \\ \hat{\mu}_{\varepsilon,n}^\dagger - \mu_\varepsilon \end{bmatrix} \\
 &+ \sqrt{n}(B_n^\dagger)^{-1}(B^\dagger - B_n^\dagger)(B^\dagger)^{-1} \begin{bmatrix} (1 + \alpha^2)Y_{s_1} - \alpha(Y_{s_1-1} + Y_{s_1+1}) - (1 - \alpha)\mu_\varepsilon \\ (1 + \alpha^2)Y_{s_2} - \alpha(Y_{s_2-1} + Y_{s_2+1}) - (1 - \alpha)\mu_\varepsilon \end{bmatrix}.
 \end{aligned}$$

Then

$$\begin{bmatrix} \sqrt{n} \left( \hat{\theta}_{1,n}^\dagger - \lim_{k \rightarrow \infty} \hat{\theta}_{1,k}^\dagger \right) \\ \sqrt{n} \left( \hat{\theta}_{2,n}^\dagger - \lim_{k \rightarrow \infty} \hat{\theta}_{2,k}^\dagger \right) \end{bmatrix} = C_{n,\alpha,\varepsilon} \begin{bmatrix} \sqrt{n}(\hat{\alpha}_n^\dagger - \alpha) \\ \sqrt{n}(\hat{\mu}_{\varepsilon,n}^\dagger - \mu_\varepsilon) \end{bmatrix} \quad (26)$$

holds asymptotically as  $n \rightarrow \infty$  with probability one, where  $C_{n,\alpha,\varepsilon}$  is defined by

$$\begin{aligned}
 & (B_n^\dagger)^{-1} \begin{bmatrix} (\hat{\alpha}_n^\dagger + \alpha)Y_{s_1} - Y_{s_1-1} - Y_{s_1+1} + \hat{\mu}_{\varepsilon,n}^\dagger \alpha - 1 \\ (\hat{\alpha}_n^\dagger + \alpha)Y_{s_2} - Y_{s_2-1} - Y_{s_2+1} + \hat{\mu}_{\varepsilon,n}^\dagger \alpha - 1 \end{bmatrix} \\
 & - (\hat{\alpha}_n^\dagger + \alpha)(B_n^\dagger)^{-1}(B^\dagger)^{-1} \begin{bmatrix} (1 + \alpha^2)Y_{s_1} - \alpha(Y_{s_1-1} + Y_{s_1+1}) - (1 - \alpha)\mu_\varepsilon \\ (1 + \alpha^2)Y_{s_2} - \alpha(Y_{s_2-1} + Y_{s_2+1}) - (1 - \alpha)\mu_\varepsilon \end{bmatrix} 0.
 \end{aligned}$$

By (12) and (13), we have  $C_{n,\alpha,\varepsilon}$  converges almost surely as  $n \rightarrow \infty$  to  $C_{\alpha,\varepsilon}$ .

By (26), (15) and Slutsky's lemma, we have (16).  $\square$

**Acknowledgements** The authors have been supported by the Hungarian Portuguese Intergovernmental S & T Cooperation Programme for 2008–2009 under Grant No. PT-07/2007. Mátyás Barczy and Gyula Pap were partially supported by the Hungarian Scientific Research Fund under Grants No. OTKA-T-079128. M. Ispány has been partially supported by TÁMOP 4.2.1/B-09/1/KONV-2010-0007/IK/IT project, which is implemented through the New Hungary Development Plan co-financed by the European Social Fund and the European Regional Development Fund.

## References

- Abraham B, Chuang A (1993) Expectation-maximization algorithms and the estimation of time series models in the presence of outliers. *J Time Ser Anal* 14:221–234
- Al-Osh MA, Alzaid AA (1987) First order integer-valued autoregressive INAR(1) process. *J Time Ser Anal* 8(3):261–275
- Barczy M, Ispány M, Pap G, Scotto M, Silva ME (2009) Outliers in INAR(1) models. Arxiv. URL: <http://arxiv.org/abs/0903.2421>
- Barczy M, Ispány M, Pap G, Scotto M, Silva ME (2010) Innovational outliers in INAR(1) models. *Commun Stat Theory Method* 39:3343–3362
- Bertsekas DP (2002) Convexity and optimization in  $\mathbb{R}^n$ . Wiley, New York
- Bhattacharya RN, Waymire EC (1990) Stochastic processes with applications. Wiley, New York

- 357 Cameron AC, Trivedi P (1998) Regression analysis of count data. Cambridge University Press, Oxford  
358 Chang I, Tiao GC, Chen C (1988) Estimation of time series parameters in the presence of outliers.  
359 Technometrics 30:193–204  
360 Du JG, Li Y (1991) The integer-valued autoregressive (INAR( $p$ )) model. J Time Ser Anal 12:129–142  
361 Fox AJ (1972) Outliers in time series. J Royal Stat Soc B 34:350–363  
362 Hall P, Heyde CC (1980) Martingale limit theory and its application. Academic Press, New York  
363 Latour A (1988) Existence and stochastic structure of a non-negative integer-valued autoregressive  
364 processes. J Time Ser Anal 4:439–455  
365 McKenzie E (1985) Some simple models for discrete variate time series. Water Resour Bull 21:645–650  
366 Montgomery DC (2005) Introduction to statistical quality control, 5th edn. John Wiley & Sons, Inc.,  
367 New York  
368 Steutel FW, van Harn K (1979) Discrete analogues of self-decomposability and stability. Ann Prob 7:893–  
369 899  
370 van Tsay RS (1988) Outliers, level shifts, and variance changes in time series. J Forecast 7:1–20  
371 Weiß CH (2008) Thinning operations for modelling time series of counts—a survey. Adv Stat Anal  
372 92(3):319–341