Innovational outliers in INAR(1) models

Mátyás Barczy[*,◇], Márton Ispány[*], Gyula Pap[*],
Manuel Scotto[**] and Maria Eduarda Silva[***]

* University of Debrecen, Faculty of Informatics, Pf. 12, H-4010 Debrecen, Hungary;

** Universidade de Aveiro, Departamento de Matemática, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal;

*** Universidade do Porto, Faculdade de Economia, Rua Dr. Roberto Frias s/n, 4200 464 Porto, Portugal.

Emails: barczy.matyas@inf.unideb.hu (M. Barczy), ispany.marton@inf.unideb.hu (M. Ispány), pap.gyula@inf.unideb.hu (G. Pap), mscotto@ua.pt (M. Scotto), mesilva@fep.up.pt (M. Eduarda Silva).

◇ Corresponding author.

Abstract

We consider integer-valued autoregressive models of order one contaminated with innovational outliers. Assuming that the time points of the outliers are known but their sizes are unknown, we prove that Conditional Least Squares (CLS) estimators of the offspring and innovation means are strongly consistent. In contrast, CLS estimators of the outliers' sizes are not strongly consistent. 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 time points preceding the outliers' occurrences, the joint CLS estimator of the sizes of the outliers is asymptotically normal.

2000 Mathematics Subject Classifications: 60J80, 62F12.

Key words and phrases: INAR(1); innovational outliers; conditional least squares estimators; strong consistency; conditional asymptotic normality.

The authors have been supported by the Hungarian Portuguese Intergovernmental S & T Cooperation Programme for 2008-2009 under Grant No. PT-07/2007. M. Barczy and G. Pap have been partially supported by the Hungarian Scientific Research Fund under Grant No. OTKA T-079128.

1 Introduction

During the last decades there has been considerable interest in integer-valued time series models and a sizeable volume of work is now available in specialized monographs. Among the most successful integer-valued time series models proposed in the literature we only mention the INteger-valued AutoRegressive model of order p (INAR(p)). This model was first introduced by McKenzie [23] and Al-Osh and Alzaid [2] for the case $p = 1$. The INAR(1) and INAR(p) models have been investigated by several authors, see, e.g., Silva and Oliveira [27], Ispány, Pap and van Zuijlen [19], and Drost, van den Akker and Werker [14]. Extensions and generalizations were proposed by Du and Li [13] and Latour [17]. Recently, the so called p-order Rounded INteger-valued AutoRegressive (RINAR(p)) time series model was introduced and studied by Kachour and Yao [21] and Kachour [20].

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. Motivation comes from the need to assess for data quality and to the robustness of subsequent statistical analysis in the presence of discordant observations. Fox [15] introduced the notion of additive and innovational outliers and proposed the use of maximum likelihood ratio test to detect them. Chang and Chen [10] 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 [29] for the detection of level shifts and temporary changes. Random level shifts were studied by Chen and Tiao [11]. Extensions of Tsay's results can be found in Balke [3]. Abraham and Chuang [1] applied the EM algorithm to the estimation of outliers. Other useful references for outlier detection and estimation in time series models are Guttman and Tiao [16], Bustos and Yohai [9], McCulloch and Tsay [22], Peña [24], Sánchez and Peña [26], Perron and Rodriguez [25] and Burridge and Taylor [8].

We emphasize that all references given in the previous paragraph deal with the case of continuous-valued processes. A related interesting problem, which has not yet been addressed, is to investigate the impact of outliers on the parameter estimation for integer-valued autoregressive models. This paper aims at giving a contribution towards this direction. In particular, we consider the problem of Conditional Least Squares (CLS) estimation of some parameters of the INAR(1) model contaminated with innovational outliers, starting from a general initial distribution (having finite second or third moments). We further assume that the time points 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 for the means of the offspring and innovation distributions are strongly consistent, but the CLS estimators of the outliers' sizes are not strongly consistent; nevertheless, they converge to a random limit with probability 1. This random limit depends on the values of the process at the outliers' time points and also on the values at the preceding time points. Moreover, 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 time points preceding the outliers' occurrences, the joint CLS estimator of the sizes of the outliers is also asymptotically normal. The corresponding asymptotic covariance matrix is also calculated.

The remainder 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 Section 3, the INAR(1) model contaminated with one or two innovational outliers, is introduced. The cases of one outlier and two outliers are handled separately. The proofs of our results are given only in case of one model (described in Subsection 3.3), namely, with one outlier estimating the mean of the offspring and innovation distributions and the outlier's size, see Section 4. For completeness, we note that the omitted proofs are available in our Arxiv preprint Barczy, Ispány, Pap, Scotto and Silva [4].

In a companion paper, we examine the INAR(1) model contaminated with additive outliers; see Barczy et al. [4, 5].

2 The INAR(1) model

2.1 The model and some preliminaries

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 $(\Omega, \mathcal{A}, \mathsf{P})$. One way to obtain models for integer-valued data is replacing multiplication in the conventional ARMA models in order to ensure the integer discreteness of the process and to adopt the terms of self-decomposability for integer-valued time series.

2.1.1 Definition. *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* $(X_n)_{n\in\mathbb{Z}_+}$ *satisfying the recursive equation*

$$X_k = \sum_{j=1}^{X_{k-1}} \xi_{k,j} + \varepsilon_k, \qquad k \in \mathbb{N}, \tag{2.1.1}$$

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* $(\varepsilon_\ell)_{\ell\in\mathbb{N}}$.

2.1.1 Remark. The INAR(1) model in (2.1.1) can be written in another way using the binomial thinning operator $\alpha\,\circ$ (due to Steutel and van Harn [28]) which we recall now. Let X be a non-negative integer-valued random variable. Let $(\xi_j)_{j\in\mathbb{N}}$ be a sequence of i.i.d. Bernoulli random variables with mean $\alpha \in [0,1]$. We assume that the sequence $(\xi_j)_{j\in\mathbb{N}}$ is independent of X. The non-negative integer-valued random variable $\alpha \circ X$ is defined by

$$\alpha \circ X := \begin{cases} \sum_{j=1}^{X} \xi_j, & \text{if } X > 0, \\ 0, & \text{if } X = 0. \end{cases}$$

The sequence $(\xi_j)_{j\in\mathbb{N}}$ is called a counting sequence. The model in (2.1.1) takes the form

$$X_k = \alpha \circ X_{k-1} + \varepsilon_k, \qquad k \in \mathbb{N}.$$

In the sequel we will assume that $\alpha \in (0,1)$, $\mathsf{E}X_0^2 < \infty$ and that $\mathsf{E}\varepsilon_1^2 < \infty$, $\mathsf{P}(\varepsilon_1 \neq 0) > 0$. In this case, it is well-known (e.g., Barczy et al. [4, Lemma 5.1]) that there exists a unique stationary distribution of the INAR(1) model in (2.1.1) which will be denoted by $\widetilde{X}$ in the sequel. Let us denote the mean and variance of ε_1 by μ_ε and σ_ε^2, respectively. Clearly, $0 < \mu_\varepsilon < \infty$. It is also well-known that

$$\mathsf{E}\widetilde{X} = \frac{\mu_\varepsilon}{1-\alpha}, \tag{2.1.2}$$

$$\mathsf{E}\widetilde{X}^2 = \frac{\sigma_\varepsilon^2 + \alpha\mu_\varepsilon}{1-\alpha^2} + \frac{\mu_\varepsilon^2}{(1-\alpha)^2}, \tag{2.1.3}$$

see, e.g., Barczy et al. [4, Appendix]. By ergodic theorems (see, e.g., Bhattacharya and Waymire [7, Section II, Theorem 9.4 (d)] or Chung [12, Section I.15, Theorem 2]), we get

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k=\mathsf{E}\widetilde{X}\right)=1, \tag{2.1.4}$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k^2=\mathsf{E}\widetilde{X}^2\right)=1, \tag{2.1.5}$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_{k-1}X_k=\mathsf{E}(\widetilde{X}(\alpha\circ\widetilde{X}+\varepsilon))=\alpha\mathsf{E}\widetilde{X}^2+\mu_\varepsilon\mathsf{E}\widetilde{X}\right)=1, \tag{2.1.6}$$

where ε is a random variable independent of $\widetilde{X}$ with the same distribution as ε_1.

In the sequel, we denote by $\mathcal{F}_k^X$ the σ–algebra generated by the random variables $X_0, X_1, \dots, X_k$.

2.2 Estimation of the mean of the offspring distribution

First we concentrate on the CLS estimation of the parameter α. Clearly, for all $k\in\mathbb{N}$, $\mathsf{E}(X_k \,|\, \mathcal{F}_{k-1}^X)=\alpha X_{k-1}+\mu_\varepsilon$, and thus

$$\sum_{k=1}^{n}\big(X_k-\mathsf{E}(X_k\,|\,\mathcal{F}_{k-1}^X)\big)^2=\sum_{k=1}^{n}\big(X_k-\alpha X_{k-1}-\mu_\varepsilon\big)^2, \qquad n\in\mathbb{N}. \tag{2.2.1}$$

For all $n\in\mathbb{N}$, a CLS estimator $\widetilde{\alpha}_n$ for the parameter $\alpha\in(0,1)$ can be obtained by minimizing the sum of squares (2.2.1) with respect to $\alpha\in\mathbb{R}$. One may check that asymptotically as $n\to\infty$, a unique CLS estimator $\widetilde{\alpha}_n$ exists with probability one and

$$\widetilde{\alpha}_n=\frac{\sum_{k=1}^{n}(X_k-\mu_\varepsilon)X_{k-1}}{\sum_{k=1}^{n}X_{k-1}^2} \tag{2.2.2}$$

holds asymptotically as $n\to\infty$ with probability one. Hereafter by the expression '*a property holds asymptotically as* $n\to\infty$ *with probability one*' we mean that there exists an event $B\in\mathcal{A}$ such that $\mathsf{P}(B)=1$ and for all $\omega\in B$ there exists an $n(\omega)\in\mathbb{N}$ such that the property in question holds for all $n\geqslant n(\omega)$. The reason why (2.2.2) holds only asymptotically as $n\to\infty$ with probability one and not for all $n\in\mathbb{N}$ and $\omega\in\Omega$ is that for all $n\in\mathbb{N}$, the probability that the denominator $\sum_{k=1}^{n}X_{k-1}^2$ equals zero is positive (provided that $\mathsf{P}(X_0=0)>0$ and $\mathsf{P}(\varepsilon_1=0)>0$), but $\mathsf{P}(\lim_{n\to\infty}\sum_{k=1}^{n}X_{k-1}^2=\infty)=1$ (which follows by (2.1.5)). Using the same arguments as in Hall and Heyde [18, Section 6.3], one can easily check that $\widetilde{\alpha}_n$ is a strongly consistent estimator of α as $n\to\infty$ for all $\alpha\in(0,1)$. Indeed, by (2.1.4)–(2.1.6), we get $\mathsf{P}(\lim_{n\to\infty}\widetilde{\alpha}_n=\alpha)=1$.

Furthermore, if $\mathsf{E}X_0^3 < \infty$ and $\mathsf{E}\varepsilon_1^3 < \infty$, then using the same arguments as in Hall and Heyde [18, Section 6.3], it follows easily that

$$\sqrt{n}(\widetilde{\alpha}_n - \alpha) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, \sigma^2_{\alpha,\varepsilon}) \qquad \text{as} \quad n \to \infty, \tag{2.2.3}$$

where $\stackrel{\mathcal{L}}{\longrightarrow}$ denotes convergence in distribution and

$$\sigma^2_{\alpha,\varepsilon} := \frac{\alpha(1-\alpha)\mathsf{E}\widetilde{X}^3 + \sigma_\varepsilon^2 \mathsf{E}\widetilde{X}^2}{(\mathsf{E}\widetilde{X}^2)^2}, \tag{2.2.4}$$

with

$$\begin{aligned}\mathsf{E}\widetilde{X}^3 = &\frac{\mathsf{E}\varepsilon^3 - 3\sigma_\varepsilon^2(1+\mu_\varepsilon) - \mu_\varepsilon^3 + 2\mu_\varepsilon}{1-\alpha^3} + 3\frac{\sigma_\varepsilon^2 + \alpha\mu_\varepsilon}{1-\alpha^2} - 2\frac{\mu_\varepsilon}{1-\alpha} \\ &+ 3\frac{\mu_\varepsilon(\sigma_\varepsilon^2 + \alpha\mu_\varepsilon)}{(1-\alpha)(1-\alpha^2)} + \frac{\mu_\varepsilon^3}{(1-\alpha)^3}.\end{aligned} \tag{2.2.5}$$

The proof of (2.2.5) can be found in the Appendix of our Arxiv preprint Barczy et al. [4].

We remark that one uses in fact Corollary 3.1 in Hall and Heyde [18] to derive (2.2.3). It is important to point out that the moment conditions $\mathsf{E}X_0^3 < \infty$ and $\mathsf{E}\varepsilon_1^3 < \infty$ are needed to check the conditions of this corollary (the so called conditional Lindeberg condition and an analogous condition on the conditional variance).

2.3 Estimation of the mean of the offspring and innovation distributions

Now we consider the joint CLS estimation of α and μ_ε. For all $n \in \mathbb{N}$, a CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n})$ for the parameter $(\alpha, \mu_\varepsilon) \in (0,1) \times (0,\infty)$ can be obtained by minimizing the sum of squares (2.2.1) with respect to $(\alpha, \mu_\varepsilon) \in \mathbb{R}^2$. One may prove that asymptotically as $n \to \infty$, a unique CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n})$ exists with probability one and

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

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

hold asymptotically as $n \to \infty$ with probability one, see, e.g., Hall and Heyde [18, formulae (6.36) and (6.37)]. It is well-known that $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n})$ is a strongly consistent estimator of $(\alpha, \mu_\varepsilon)$ as $n \to \infty$ for all $(\alpha, \mu_\varepsilon) \in (0,1) \times (0,\infty)$, see, e.g., Hall and Heyde [18, Section 6.3].

Moreover, if $\mathsf{E}X_0^3 < \infty$ and $\mathsf{E}\varepsilon_1^3 < \infty$, by Hall and Heyde [18, formula (6.44)],

$$\begin{bmatrix} \sqrt{n}(\widehat{\alpha}_n - \alpha) \\ \sqrt{n}(\widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon) \end{bmatrix} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, B_{\alpha,\varepsilon} \right) \qquad \text{as } n \to \infty, \tag{2.3.1}$$

where

$$\begin{aligned} B_{\alpha,\varepsilon} &:= \begin{bmatrix} \mathsf{E}\widetilde{X}^2 & \mathsf{E}\widetilde{X} \\ \mathsf{E}\widetilde{X} & 1 \end{bmatrix}^{-1} A_{\alpha,\varepsilon} \begin{bmatrix} \mathsf{E}\widetilde{X}^2 & \mathsf{E}\widetilde{X} \\ \mathsf{E}\widetilde{X} & 1 \end{bmatrix}^{-1} \\ &= \frac{1}{(\operatorname{Var}\widetilde{X})^2} \begin{bmatrix} 1 & -\mathsf{E}\widetilde{X} \\ -\mathsf{E}\widetilde{X} & \mathsf{E}\widetilde{X}^2 \end{bmatrix} A_{\alpha,\varepsilon} \begin{bmatrix} 1 & -\mathsf{E}\widetilde{X} \\ -\mathsf{E}\widetilde{X} & \mathsf{E}\widetilde{X}^2 \end{bmatrix}, \end{aligned} \tag{2.3.2}$$

with

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

and $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1).

3 The INAR(1) model with innovational outliers

3.1 The model and some preliminaries

For all $k, \ell \in \mathbb{Z}_+$, let $\delta_{k,\ell} := 0$ if $k \neq \ell$ and $\delta_{k,k} := 1$.

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

3.1.1 Definition. *Let $(\varepsilon_\ell)_{\ell\in\mathbb{N}}$ be an i.i.d. sequence of non-negative integer-valued random variables. A stochastic process $(Y_k)_{k\in\mathbb{Z}_+}$ is called an INAR(1) model with finitely many innovational outliers if*

$$Y_k = \sum_{j=1}^{Y_{k-1}} \xi_{k,j} + \eta_k, \qquad k \in \mathbb{N},$$

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* Y_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* $(\varepsilon_\ell)_{\ell\in\mathbb{N}}$, *and*

$$\eta_k := \varepsilon_k + \sum_{i=1}^{I} \delta_{k,s_i}\theta_i, \qquad k\in\mathbb{Z}_+,$$

where $I\in\mathbb{N}$, $s_i,\theta_i\in\mathbb{N}$, $i=1,\dots,I$ *such that* $s_i\neq s_j$ *if* $i\neq j$, $i,j=1,\dots,I$. *We assume that* $\mathsf{E}Y_0^2<\infty$ *and that* $\mathsf{E}\varepsilon_1^2<\infty$, $\mathsf{P}(\varepsilon_1\neq 0)>0$.

In case of one (innovational) outlier a more suitable representation of $(Y_k)_{k\in\mathbb{Z}_+}$ is given in the following proposition.

3.1.1 Proposition. *Let* $(Y_k)_{k\in\mathbb{Z}_+}$ *be an INAR(1) model with one innovational outlier* $\theta_1:=\theta$ *at time point* $s_1:=s$. *Then for all* $\omega\in\Omega$ *and* $k\in\mathbb{Z}_+$, $Y_k(\omega)=X_k(\omega)+Z_k(\omega)$, *where* $(X_k)_{k\in\mathbb{Z}_+}$ *is an INAR(1) process given by*

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

with $X_0:=Y_0$, *and*

$$Z_k := \begin{cases} 0 & \text{if } k=0,1,\dots,s-1,\\ \theta & \text{if } k=s,\\ \sum_{j=X_{k-1}+1}^{X_{k-1}+Z_{k-1}} \xi_{k,j} & \text{if } k\geqslant s+1. \end{cases}$$

Furthermore, the processes X *and* Z *are independent, and* $\mathsf{P}(\lim_{k\to\infty} Z_k=0)=1$ *and* $Z_k \xrightarrow{L_p} 0$ *as* $k\to\infty$ *for all* $p\in\mathbb{N}$, *where* $\xrightarrow{L_p}$ *denotes convergence in* L_p.

Proof. See Subsection 4.1. □

For our later purposes we need to calculate the first and second moments of Z.

3.1.2 Proposition. *We have*

$$\mathsf{E}Z_{s+k} = \theta\alpha^k, \qquad k\in\mathbb{Z}_+, \tag{3.1.1}$$

$$\mathsf{E}Z_{s+k}^2 = \theta^2\alpha^{2k} - \theta\alpha^k(\alpha^k-1), \qquad k\in\mathbb{Z}_+, \tag{3.1.2}$$

$$\mathsf{E}(Z_{s+k-1}Z_{s+k}) = \alpha\mathsf{E}Z_{s+k-1}^2 = \theta^2\alpha^{2k-1} - \theta\alpha^k(\alpha^{k-1}-1), \quad k\in\mathbb{N}. \tag{3.1.3}$$

Proof. See Subsection 4.1. □

In the sequel we denote by $\mathcal{F}_k^Y$ the σ–algebra generated by the random variables $Y_0, Y_1, \dots, Y_k$. For pairwise distinct positive integers $s_1, \dots, s_I$ we use the notation

$$\sum_{k=1}^{n}{}^{(s_1,\dots,s_I)} := \sum_{\substack{k=1\\ k\neq s_1,\dots,k\neq s_I}}^{n} .$$

3.2 One outlier case: estimation of the mean of the offspring distribution and the outlier's size

First we suppose that $I = 1$ and that the relevant time point $s_1 := s$ is known. We concentrate on the CLS estimation of the parameter (α, θ), where $\theta := \theta_1$. An easy calculation shows that

$$\mathsf{E}(Y_k \mid \mathcal{F}_{k-1}^Y) = \alpha Y_{k-1} + \mathsf{E}\eta_k = \alpha Y_{k-1} + \mu_\varepsilon + \delta_{k,s}\theta, \qquad k \in \mathbb{N}.$$

Hence for all $n \geqslant s$,

$$\begin{aligned}&\sum_{k=1}^{n}\big(Y_k - \mathsf{E}(Y_k \mid \mathcal{F}_{k-1}^Y)\big)^2\\ &\qquad = \sum_{k=1}^{n}{}^{(s)} \big(Y_k - \alpha Y_{k-1} - \mu_\varepsilon\big)^2 + \big(Y_s - \alpha Y_{s-1} - \mu_\varepsilon - \theta\big)^2.\end{aligned} \tag{3.2.1}$$

For all $n \geqslant s$, a CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_n)$ for the parameter $(\alpha, \theta) \in (0,1)\times\mathbb{N}$ can be obtained by minimizing the sum of squares (3.2.1) with respect to $(\alpha,\theta) \in \mathbb{R}^2$. Barczy et al. [4, Lemma 4.2.1] showed that asymptotically as $n \to \infty$, a unique CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_n)$ exists with probability one.

3.2.1 Theorem. *The CLS estimator* $\widetilde{\alpha}_n$ *of* α *is strongly consistent for all* $(\alpha,\theta) \in (0,1)\times\mathbb{N}$, *i.e.,* $\mathsf{P}(\lim_{n\to\infty} \widetilde{\alpha}_n = \alpha) = 1$. *The CLS estimator* $\widetilde{\theta}_n$ *of* θ *is not strongly consistent for any* $(\alpha,\theta) \in (0,1)\times\mathbb{N}$, *namely,*

$$\mathsf{P}\left(\lim_{n\to\infty} \widetilde{\theta}_n = Y_s - \alpha Y_{s-1} - \mu_\varepsilon\right) = 1, \qquad \forall\, (\alpha,\theta) \in (0,1)\times\mathbb{N}.$$

Proof. The proof can be found in Barczy et al. [4, Theorem 4.2.1]. □

3.2.2 Theorem. *Under the additional assumptions* $\mathsf{E}Y_0^3 < \infty$ *and* $\mathsf{E}\varepsilon_1^3 < \infty$, *we have*

$$\sqrt{n}(\widetilde{\alpha}_n - \alpha) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{\alpha,\varepsilon}^2) \qquad \text{as } n \to \infty,$$

with $\sigma^2_{\alpha,\varepsilon}$ *defined as in* (2.2.4)*. Moreover, conditionally on the value* Y_{s-1}*,*

$$\sqrt{n}\big(\widetilde{\theta}_n - \lim_{k\to\infty}\widetilde{\theta}_k\big) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, Y_{s-1}^2\sigma^2_{\alpha,\varepsilon}) \qquad \text{as } n\to\infty.$$

Proof. The proof can be found in Barczy et al. [4, Theorem 4.2.2]. □

3.3 One outlier case: estimation of the mean of the offspring and innovation distributions and the outlier's size

We suppose that $I = 1$ and that $s_1 := s$ is known. We consider the CLS estimation of $(\alpha, \mu_\varepsilon, \theta)$, where $\theta := \theta_1$. For all $n \geqslant s$, a CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n)$ for the parameter $(\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}$ can be obtained by minimizing the sum of squares (3.2.1) with respect to $(\alpha, \mu_\varepsilon, \theta) \in \mathbb{R}^3$. In Subsection 4.2 we prove that asymptotically as $n\to\infty$, a unique CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n)$ exists with probability one.

3.3.1 Theorem. *The CLS estimator* $\widehat{\alpha}_n$ *of* α *and* $\widehat{\mu}_{\varepsilon,n}$ *of* μ_ε *are strongly consistent for all* $(\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}$*, i.e.,*

$$\mathsf{P}\left(\lim_{n\to\infty}\widehat{\alpha}_n = \alpha\right) = 1, \qquad \forall\,(\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}, \tag{3.3.1}$$

$$\mathsf{P}\left(\lim_{n\to\infty}\widehat{\mu}_{\varepsilon,n} = \mu_\varepsilon\right) = 1, \qquad \forall\,(\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}. \tag{3.3.2}$$

The CLS estimator $\widehat{\theta}_n$ *of* θ *is not strongly consistent for any* $(\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}$*, namely,*

$$\mathsf{P}\left(\lim_{n\to\infty}\widehat{\theta}_n = Y_s - \alpha Y_{s-1} - \mu_\varepsilon\right) = 1, \qquad \forall\,(\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}. \tag{3.3.3}$$

Proof. See Subsection 4.2. □

3.3.2 Theorem. *Under the additional assumptions* $\mathsf{E}Y_0^3 < \infty$ *and* $\mathsf{E}\varepsilon_1^3 < \infty$*, we have*

$$\begin{bmatrix} \sqrt{n}(\widehat{\alpha}_n - \alpha) \\ \sqrt{n}(\widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon) \end{bmatrix} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, B_{\alpha,\varepsilon}\right) \qquad \text{as } n\to\infty, \tag{3.3.4}$$

with $B_{\alpha,\varepsilon}$ *defined as in* (2.3.2)*. Moreover, conditionally on the value* Y_{s-1}*,*

$$\sqrt{n}(\widehat{\theta}_n - \lim_{k\to\infty}\widehat{\theta}_k) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, [Y_{s-1}\ 0]B_{\alpha,\varepsilon}[Y_{s-1}\ 0]^\top) \qquad \text{as } n\to\infty. \tag{3.3.5}$$

Proof. See Subsection 4.2. □

3.4 Two outliers case: estimation of the mean of the offspring distribution and the outliers' sizes

In this subsection we assume that $I = 2$ and that the relevant time points $s_1, s_2 \in \mathbb{N}$, $s_1 \neq s_2$, are known. We concentrate on the CLS estimation of $(\alpha, \theta_1, \theta_2)$. We have

$$\mathsf{E}(Y_k \,|\, \mathcal{F}_{k-1}^Y) = \alpha Y_{k-1} + \mu_\varepsilon + \delta_{k,s_1}\theta_1 + \delta_{k,s_2}\theta_2, \qquad k \in \mathbb{N}.$$

Hence for all $n \geqslant \max(s_1, s_2)$,

$$\sum_{k=1}^{n} \big(Y_k - \mathsf{E}(Y_k \,|\, \mathcal{F}_{k-1}^Y)\big)^2 = \sum_{k=1}^{n}{}^{(s_1,s_2)} \big(Y_k - \alpha Y_{k-1} - \mu_\varepsilon\big)^2 + \big(Y_{s_1} - \alpha Y_{s_1-1} - \mu_\varepsilon - \theta_1\big)^2 + \big(Y_{s_2} - \alpha Y_{s_2-1} - \mu_\varepsilon - \theta_2\big)^2. \tag{3.4.1}$$

By minimizing the sum of squares (3.4.1) with respect to $(\alpha, \theta_1, \theta_2) \in \mathbb{R}^3$, a CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_{1,n}, \widetilde{\theta}_{2,n})$ for the parameter $(\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2$ can be obtained for all $n \geqslant \max(s_1, s_2)$. Barczy et al. [4, Lemma 4.4.1] showed that asymptotically as $n \to \infty$, a unique CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_{1,n}, \widetilde{\theta}_{2,n})$ exists with probability one.

3.4.1 Theorem. *The CLS estimator $\widetilde{\alpha}_n$ of α is strongly consistent for all $(\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2$, i.e., $\mathsf{P}(\lim_{n\to\infty} \widetilde{\alpha}_n = \alpha) = 1$. For $i = 1, 2$, the CLS estimator $\widetilde{\theta}_{i,n}$ of θ_i is not strongly consistent for any $(\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2$, namely,*

$$\mathsf{P}\left(\lim_{n\to\infty} \widetilde{\theta}_{i,n} = Y_{s_i} - \alpha Y_{s_i-1} - \mu_\varepsilon\right) = 1, \qquad \forall\, (\alpha, \theta_1, \theta_1) \in (0,1) \times \mathbb{N}^2.$$

Proof. The proof can be found in Barczy et al. [4, Theorem 4.4.1]. □

3.4.2 Theorem. *Under the additional assumptions $\mathsf{E}Y_0^3 < \infty$ and $\mathsf{E}\varepsilon_1^3 < \infty$, we have*

$$\sqrt{n}(\widetilde{\alpha}_n - \alpha) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, \sigma_{\alpha,\varepsilon}^2) \qquad \text{as } n \to \infty,$$

with $\sigma_{\alpha,\varepsilon}^2$ defined as in (2.2.4). Moreover, conditionally on the values Y_{s_1-1} and Y_{s_2-1}, as $n \to \infty$,

$$\begin{bmatrix} \sqrt{n}\big(\widetilde{\theta}_{1,n} - \lim_{k\to\infty} \widetilde{\theta}_{1,k}\big) \\ \sqrt{n}\big(\widetilde{\theta}_{2,n} - \lim_{k\to\infty} \widetilde{\theta}_{2,k}\big) \end{bmatrix} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_{\alpha,\varepsilon}^2 \begin{bmatrix} Y_{s_1-1}^2 & Y_{s_1-1}Y_{s_2-1} \\ Y_{s_1-1}Y_{s_2-1} & Y_{s_2-1}^2 \end{bmatrix} \right).$$

Proof. The proof can be found in Barczy et al. [4, Theorem 4.4.2]. □

3.5 Two outliers case: estimation of the mean of the offspring and innovation distributions and the outliers' sizes

In this subsection we assume that $I = 2$ and that the relevant time points $s_1, s_2 \in \mathbb{N}$, $s_1 \neq s_2$, are known. We consider the CLS estimation of $(\alpha, \mu_\varepsilon, \theta_1, \theta_2)$. For all $n \geqslant \max(s_1, s_2)$, a CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_{1,n}, \widehat{\theta}_{2,n})$ for the parameter $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0,1) \times (0,\infty) \times \mathbb{N}^2$ can be obtained by minimizing the sum of squares (3.4.1) with respect to $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in \mathbb{R}^4$. Barczy et al. [4, Lemma 4.5.1] showed that asymptotically as $n \to \infty$, a unique CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_{1,n}, \widehat{\theta}_{2,n})$ exists with probability one.

3.5.1 Theorem. *The CLS estimator $\widehat{\alpha}_n$ of α and $\widehat{\mu}_{\varepsilon,n}$ of μ_ε are strongly consistent for all $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0,1) \times (0,\infty) \times \mathbb{N}^2$, i.e., $\mathsf{P}(\lim_{n\to\infty} \widehat{\alpha}_n = \alpha) = 1$ and $\mathsf{P}(\lim_{n\to\infty} \widehat{\mu}_{\varepsilon,n} = \mu_\varepsilon) = 1$. For $i = 1, 2$, the CLS estimator $\widehat{\theta}_{i,n}$ of θ_i is not strongly consistent for any $(\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0,1) \times (0,\infty) \times \mathbb{N}^2$, namely,*

$$\mathsf{P}\left(\lim_{n\to\infty} \widehat{\theta}_{i,n} = Y_{s_i} - \alpha Y_{s_i-1} - \mu_\varepsilon\right) = 1, \qquad \forall\, (\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0,1) \times (0,\infty) \times \mathbb{N}^2.$$

Proof. The proof can be found in Barczy et al. [4, Theorem 4.5.1]. □

3.5.2 Theorem. *Under the additional assumptions $\mathsf{E}Y_0^3 < \infty$ and $\mathsf{E}\varepsilon_1^3 < \infty$, we have*

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

with $B_{\alpha,\varepsilon}$ defined as in (2.3.2). Moreover, conditionally on the values Y_{s_1-1} and Y_{s_2-1},

$$\begin{bmatrix} \sqrt{n}\big(\widehat{\theta}_{1,n} - \lim_{k\to\infty} \widehat{\theta}_{1,k}\big) \\ \sqrt{n}\big(\widehat{\theta}_{2,n} - \lim_{k\to\infty} \widehat{\theta}_{2,k}\big) \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, C_{\alpha,\varepsilon} B_{\alpha,\varepsilon} C_{\alpha,\varepsilon}^\top \right) \qquad \text{as } n \to \infty,$$

where

$$C_{\alpha,\varepsilon} := \begin{bmatrix} Y_{s_1-1} & 1 \\ Y_{s_2-1} & 1 \end{bmatrix}.$$

Proof. The proof can be found in Barczy et al. [4, Theorem 4.5.2]. □

4 Proofs

4.1 Proofs for Subsection 3.1

In proving Proposition 3.1.1, we need the following result.

4.1.1 Lemma. *Let* $(X_n)_{n\in\mathbb{Z}_+}$ *and* $(Z_n)_{n\in\mathbb{Z}_+}$ *be two (not necessarily homogeneous) Markov chains with state space* $\mathbb{Z}_+$*. Let us suppose that* $(X_n, Z_n)_{n\in\mathbb{Z}_+}$ *is a Markov chain,* X_0 *and* Z_0 *are independent, and that for all* $n \in \mathbb{N}$ *and* $i, j, k, \ell \in \mathbb{Z}_+$ *such that* $\mathsf{P}(X_{n-1} = k,\, Z_{n-1} = \ell) > 0$,

$$\begin{aligned}
&\mathsf{P}(X_n = i, Z_n = j \,|\, X_{n-1} = k, Z_{n-1} = \ell)\\
&\qquad = \mathsf{P}(X_n = i \,|\, X_{n-1} = k)\mathsf{P}(Z_n = j \,|\, Z_{n-1} = \ell).
\end{aligned}$$

Then $(X_n)_{n\in\mathbb{Z}_+}$ *and* $(Z_n)_{n\in\mathbb{Z}_+}$ *are independent.*

Proof. For all $n \in \mathbb{N}$ and $i_0, i_1, \dots, i_n, j_0, j_1, \dots, j_n \in \mathbb{Z}_+$, we get

$$\begin{aligned}
&\mathsf{P}(X_n = i_n, \dots, X_0 = i_0, Z_n = j_n, \dots, Z_0 = j_0)\\
&= \mathsf{P}(X_n = i_n, Z_n = j_n \,|\, X_{n-1} = i_{n-1}, Z_{n-1} = j_{n-1}) \cdots\\
&\quad \times \mathsf{P}(X_1 = i_1, Z_1 = j_1 \,|\, X_0 = i_0, Z_0 = j_0)\mathsf{P}(X_0 = i_0, Z_0 = j_0)\\
&= \mathsf{P}(X_n = i_n \,|\, X_{n-1} = i_{n-1}) \cdots \mathsf{P}(X_1 = i_1 \,|\, X_0 = i_0)\mathsf{P}(X_0 = i_0)\\
&\quad \times \mathsf{P}(Z_n = j_n \,|\, Z_{n-1} = j_{n-1}) \cdots \mathsf{P}(Z_1 = j_1 \,|\, Z_0 = j_0)\mathsf{P}(Z_0 = j_0)\\
&= \mathsf{P}(X_n = i_n, \dots, X_0 = i_0)\mathsf{P}(Z_n = j_n, \dots, Z_0 = j_0),
\end{aligned}$$

which yields that $X_n, \dots, X_0$ and $Z_n, \dots, Z_0$ are independent. One can think it over that this implies the statement. □

Proof of Proposition 3.1.1. Clearly, $Y_j = X_j + Z_j = X_j$ for $j = 0, 1, \dots, s-1$, and by induction, one can easily check that $Y_k = X_k + Z_k$ for all $k \geqslant s$.

In proving the independence of the processes X and Z, it is enough to check that the conditions of Lemma 4.1.1 are satisfied. For all $n > s$, $i_{n-1}, i_n, j_{n-1}, j_n \in \mathbb{Z}_+$ and for all $B \in \sigma(\xi_{i,j} : i = 1, \dots, n-2,\ j \in \mathbb{N})$ with the property that the event $A := \{X_{n-1} =$

$i_{n-1}, Z_{n-1} = j_{n-1}\} \cap B$ has positive probability, we get

$$\begin{aligned}
\mathsf{P}(X_n = i_n, Z_n = j_n \,|\, A) &= \mathsf{P}\left(\sum_{j=1}^{i_{n-1}} \xi_{n,j} + \varepsilon_n = i_n, \sum_{j=i_{n-1}+1}^{i_{n-1}+j_{n-1}} \xi_{n,j} = j_n \,\middle|\, A \right) \\
&= \mathsf{P}\left(\sum_{j=1}^{i_{n-1}} \xi_{n,j} + \varepsilon_n = i_n, \sum_{j=i_{n-1}+1}^{i_{n-1}+j_{n-1}} \xi_{n,j} = j_n \right) \qquad (4.1.1) \\
&= \mathsf{P}\left(\sum_{j=1}^{i_{n-1}} \xi_{n,j} + \varepsilon_n = i_n \right) \mathsf{P}\left(\sum_{j=i_{n-1}+1}^{i_{n-1}+j_{n-1}} \xi_{n,j} = j_n \right),
\end{aligned}$$

where we used the measurability of (X_{n-1}, Z_{n-1}) with respect to the σ–algebra $\sigma(\xi_{i,j} : i = 1, \ldots, n-1,\ j \in \mathbb{N})$ and that the random variables ε_n, $(\xi_{n,1}, \ldots, \xi_{n,i_{n-1}})$ and $(\xi_{n,i_{n-1}+1}, \ldots, \xi_{n,i_{n-1}+j_{n-1}})$ are independent of this σ–algebra and also from each other. Hence, for all $n > s$,

$$\mathsf{P}(X_n = i_n, Z_n = j_n \,|\, A) = \mathsf{P}(X_n = i_n, Z_n = j_n \,|\, X_{n-1} = i_{n-1}, Z_{n-1} = j_{n-1}). \qquad (4.1.2)$$

Since $Z_0 = Z_1 = \cdots = Z_{s-1} = 0$, $Z_s = \theta$, and $(X_n)_{n\in\mathbb{Z}_+}$ is a Markov chain, we have (4.1.2) is satisfied also for $n = 1, 2, \ldots, s$, which yields that $(X_n, Z_n)_{n\in\mathbb{Z}_+}$ is a Markov chain. Since $Z_0 = 0$, X_0 and Z_0 are independent. Similar arguments along with the result in (4.1.1), with the special choice $B := \Omega$ lead to

$$\begin{aligned}
&\mathsf{P}(X_n = i_n, Z_n = j_n \,|\, X_{n-1} = i_{n-1}, Z_{n-1} = j_{n-1}) \\
&= \mathsf{P}\left(\sum_{j=1}^{i_{n-1}} \xi_{n,j} + \varepsilon_n = i_n \,\middle|\, X_{n-1} = i_{n-1} \right) \mathsf{P}\left(\sum_{j=1}^{j_{n-1}} \xi_{n,j+i_{n-1}} = j_n \,\middle|\, Z_{n-1} = j_{n-1} \right) \\
&= \mathsf{P}(X_n = i_n \,|\, X_{n-1} = i_{n-1}) \mathsf{P}(Z_n = j_n \,|\, Z_{n-1} = j_{n-1}),
\end{aligned}$$

which yields that the conditions of Lemma 4.1.1 are satisfied.

Since

$$Z_{k+1} = \sum_{j=X_k+1}^{X_k+Z_k} \xi_{k+1,j} \leqslant \sum_{j=X_k+1}^{X_k+Z_k} 1 = Z_k, \qquad k \geqslant s,$$

the nonnegative sequence $(Z_k(\omega))_{k\geqslant s+1}$ is monotone decreasing for all $\omega \in \Omega$, thus $(Z_k(\omega))_{k\in\mathbb{Z}_+}$ converges for all $\omega \in \Omega$. Hence, if we check that Z_k converges in probability to 0 as $k \to \infty$, then, by Riesz's theorem, we get $\mathsf{P}(\lim_{k\to\infty} Z_k = 0) = 1$. Let $\mathcal{F}_k^{X,Z}$ be the σ–algebra generated by the random variables $Z_0, Z_1, \ldots, Z_k$ and $X_0, X_1, \ldots, X_k$. Using that $\mathsf{E}(Z_k \,|\, \mathcal{F}_{k-1}^{X,Z}) = \alpha Z_{k-1}$, $k \geqslant s+1$, we get $\mathsf{E} Z_k = \alpha \mathsf{E} Z_{k-1}$, $k \geqslant s+1$, and hence

$\mathsf{E}Z_{s+k} = \alpha^k \mathsf{E}Z_s = \theta\alpha^k$, $k \geqslant 0$. For all $\varepsilon > 0$, by Markov's inequality,

$$\mathsf{P}(Z_{s+k} \geqslant \varepsilon) \leqslant \frac{\mathsf{E}Z_{s+k}}{\varepsilon} = \frac{\theta\alpha^k}{\varepsilon} \to 0 \qquad \text{as} \quad k \to \infty,$$

as desired.

Since the sequence $(Z_k(\omega))_{k\geqslant s+1}$ is monotone decreasing for all $\omega \in \Omega$, we get for all $p \in \mathbb{N}$ and for any constant $M > 0$, the sequence $\big(|Z_k|^p \mathbb{1}_{\{|Z_k|\geqslant M\}}\big)_{k\geqslant s+1}$ is monotone decreasing. Hence

$$\sup_{k\geqslant s+1} \mathsf{E}\left(|Z_k|^p \mathbb{1}_{\{|Z_k|\geqslant M\}}\right) = \mathsf{E}\left(|Z_{s+1}|^p \mathbb{1}_{\{|Z_{s+1}|\geqslant M\}}\right) \to 0 \qquad \text{as} \quad M \to \infty,$$

which yields the uniformly integrability of $(Z_k^p)_{k\in\mathbb{N}}$. By Theorem 3.6 in Bhattacharya and Waymire [7, Chapter 0], we conclude that $Z_k \xrightarrow{L_p} 0$ as $k \to \infty$, i.e., $\lim_{k\to\infty} \mathsf{E}Z_k^p = 0$. This completes the proof. □

Proof of Proposition 3.1.2. In the proof of Proposition 3.1.1 we have already checked (3.1.1). Using that for all $k \geqslant s+1$,

$$\mathsf{E}\Big((Z_k - \alpha Z_{k-1})^2 \,\big|\, \mathcal{F}_{k-1}^{X,Z}\Big) = \mathsf{E}\left(\left(\sum_{j=X_{k-1}+1}^{X_{k-1}+Z_{k-1}} (\xi_{k,j} - \alpha)\right)^2 \,\middle|\, \mathcal{F}_{k-1}^{X,Z}\right) = \alpha(1-\alpha)Z_{k-1},$$

we get for all $k \geqslant s+1$,

$$\mathsf{E}\Big(Z_k^2 \,\big|\, \mathcal{F}_{k-1}^{X,Z}\Big) = \mathsf{E}\Big(\big((Z_k - \alpha Z_{k-1}) + \alpha Z_{k-1}\big)^2 \,\big|\, \mathcal{F}_{k-1}^{X,Z}\Big) = \alpha(1-\alpha)Z_{k-1} + \alpha^2 Z_{k-1}^2,$$

and hence $\mathsf{E}Z_k^2 = \alpha^2 \mathsf{E}Z_{k-1}^2 + \alpha(1-\alpha)\mathsf{E}Z_{k-1}$, $k \geqslant s+1$. Then

$$\begin{bmatrix} \mathsf{E}Z_k \\ \mathsf{E}Z_k^2 \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ \alpha(1-\alpha) & \alpha^2 \end{bmatrix} \begin{bmatrix} \mathsf{E}Z_{k-1} \\ \mathsf{E}Z_{k-1}^2 \end{bmatrix}, \qquad k \geqslant s+1,$$

and hence, by an easy calculation, we have (3.1.2). Finally, for all $k \in \mathbb{N}$,

$$\begin{aligned} \mathsf{E}(Z_{s+k-1}Z_{s+k}) &= \mathsf{E}\big(\mathsf{E}(Z_{s+k-1}Z_{s+k} \,|\, \mathcal{F}_{s+k-1}^{X,Z})\big) = \mathsf{E}\big(Z_{s+k-1}\mathsf{E}(Z_{s+k} \,|\, \mathcal{F}_{s+k-1}^{X,Z})\big) \\ &= \mathsf{E}\big(Z_{s+k-1}\alpha Z_{s+k-1}\big) = \alpha \mathsf{E}Z_{s+k-1}^2, \end{aligned}$$

which yields (3.1.3). This completes the proof. □

4.2 Proofs for Subsection 3.3

We retain the notations introduced in Subsection 3.3 and for all $n \in \mathbb{N}$, $y_0, \dots, y_n \in \mathbb{R}$ and $\omega \in \Omega$, let us put

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

First we give a proof that asymptotically as $n \to \infty$, a unique CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n)$ exists with probability one. Motivated by (3.2.1), for all $n \geqslant s$, $n \in \mathbb{N}$, we define the function $Q_n : \mathbb{R}^{n+1} \times \mathbb{R}^3 \to \mathbb{R}$ by

$$Q_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') := \sum_{k=1}^{n}{}^{(s)} \big(y_k - \alpha' y_{k-1} - \mu'_\varepsilon\big)^2 + \big(y_s - \alpha' y_{s-1} - \mu'_\varepsilon - \theta'\big)^2$$

for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $\alpha', \mu'_\varepsilon, \theta' \in \mathbb{R}$. By definition, for all $n \geqslant s$, a CLS estimator for the parameter $(\alpha, \mu_\varepsilon, \theta) \in (0,1) \times (0,\infty) \times \mathbb{N}$ is a measurable function $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n) : \mathbb{R}^{n+1} \to \mathbb{R}^3$ such that

$$Q_n(\mathbf{y}_n; \widehat{\alpha}_n(\mathbf{y}_n), \widehat{\mu}_{\varepsilon,n}(\mathbf{y}_n), \widehat{\theta}_n(\mathbf{y}_n)) = \inf_{(\alpha', \mu'_\varepsilon, \theta') \in \mathbb{R}^3} Q_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'), \qquad \forall\ \mathbf{y}_n \in \mathbb{R}^{n+1}.$$

The next lemma is about the unique existence of the CLS estimator of $(\alpha, \mu_\varepsilon, \theta)$.

4.2.1 Lemma. *There exists an event $A \in \mathcal{A}$ such that $\mathsf{P}(A) = 1$ and for all $\omega \in A$ there exists an $n(\omega) \in \mathbb{N}$ with the property that the function*

$$\mathbb{R}^3 \ni (\alpha', \mu'_\varepsilon, \theta') \mapsto Q_n(\mathbf{Y}_n(\omega); \alpha', \mu'_\varepsilon, \theta')$$

is strictly convex for all $n \geqslant n(\omega)$, and

$$\frac{\partial Q_n}{\partial \alpha'}(\mathbf{Y}_n(\omega); \alpha', \mu'_\varepsilon, \theta') = 0, \qquad \frac{\partial Q_n}{\partial \mu'_\varepsilon}(\mathbf{Y}_n(\omega); \alpha', \mu'_\varepsilon, \theta') = 0, \qquad \frac{\partial Q_n}{\partial \theta'}(\mathbf{Y}_n(\omega); \alpha', \mu'_\varepsilon, \theta') = 0, \tag{4.2.1}$$

has a unique solution with respect to $(\alpha', \mu'_\varepsilon, \theta') \in \mathbb{R}^3$ for all $n \geqslant n(\omega)$. Consequently, for all $\omega \in A$ and $n \geqslant n(\omega)$, the function $\mathbb{R}^3 \ni (\alpha', \mu'_\varepsilon, \theta') \mapsto Q_n(\mathbf{Y}_n(\omega); \alpha', \mu'_\varepsilon, \theta')$ attains its minimum at this unique solution.

Lemma 4.2.1 states that asymptotically as $n \to \infty$, the system of equations (4.2.1) has a unique solution with respect to $(\alpha', \mu'_\varepsilon, \theta') \in \mathbb{R}^3$ with probability one, which is nothing else but $(\widehat{\alpha}_n(\mathbf{Y}_n(\omega)), \widehat{\mu}_{\varepsilon,n}(\mathbf{Y}_n(\omega)), \widehat{\theta}_n(\mathbf{Y}_n(\omega)))$. In the sequel we simply denote $(\widehat{\alpha}_n(\mathbf{Y}_n(\omega)), \widehat{\mu}_{\varepsilon,n}(\mathbf{Y}_n(\omega)), \widehat{\theta}_n(\mathbf{Y}_n(\omega)))$ by $(\widehat{\alpha}_n(\omega), \widehat{\mu}_{\varepsilon,n}(\omega), \widehat{\theta}_n(\omega))$.

Proof of Lemma 4.2.1. In proving the strict convexity of the function in question, it is enough to check that there exists an event $A \in \mathcal{A}$ such that $\mathsf{P}(A) = 1$ and for all $\omega \in A$

there exists an $n(\omega) \in \mathbb{N}$ with the property that (3×3) Hessian matrix

$$H_n(\omega, \alpha', \mu'_\varepsilon, \theta') := \begin{bmatrix} \frac{\partial^2 Q_n}{\partial(\alpha')^2} & \frac{\partial^2 Q_n}{\partial\mu'_\varepsilon \partial\alpha'} & \frac{\partial^2 Q_n}{\partial\theta' \partial\alpha'} \\ \frac{\partial^2 Q_n}{\partial\alpha' \partial\mu'_\varepsilon} & \frac{\partial^2 Q_n}{\partial(\mu'_\varepsilon)^2} & \frac{\partial^2 Q_n}{\partial\theta' \partial\mu'_\varepsilon} \\ \frac{\partial^2 Q_n}{\partial\alpha' \partial\theta'} & \frac{\partial^2 Q_n}{\partial\mu'_\varepsilon \partial\theta'} & \frac{\partial^2 Q_n}{\partial(\theta')^2} \end{bmatrix} (\mathbf{Y}_n(\omega); \alpha', \mu'_\varepsilon, \theta')$$

is (strictly) positive definite for all $n \geqslant n(\omega)$ and $(\alpha', \mu'_\varepsilon, \theta') \in \mathbb{R}^3$, see, e.g., Berkovitz [6, Theorem 3.3, Chapter III]. We get for all $(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') \in \mathbb{R}^{n+1} \times \mathbb{R}^3$,

$$\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = -2 \sum_{k=1}^{n}{}^{(s)} \big(y_k - \alpha' y_{k-1} - \mu'_\varepsilon\big) y_{k-1} - 2\big(y_s - \alpha' y_{s-1} - \mu'_\varepsilon - \theta'\big) y_{s-1},$$

$$\frac{\partial Q_n}{\partial \mu'_\varepsilon}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = -2 \sum_{k=1}^{n}{}^{(s)} \big(y_k - \alpha' y_{k-1} - \mu'_\varepsilon\big) - 2\big(y_s - \alpha' y_{s-1} - \mu'_\varepsilon - \theta'\big),$$

$$\frac{\partial Q_n}{\partial \theta'}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = -2\big(y_s - \alpha' y_{s-1} - \mu'_\varepsilon - \theta'\big).$$

In the sequel we simply write $H_n(\alpha', \mu'_\varepsilon, \theta')$ instead of $H_n(\omega, \alpha', \mu'_\varepsilon, \theta')$. One can easily obtain that the matrix $H_n(\alpha', \mu'_\varepsilon, \theta')$ has the following leading principal minors

$$\Delta_{1,n}(\alpha', \mu'_\varepsilon, \theta') := 2 \sum_{k=1}^n Y_{k-1}^2, \qquad \Delta_{2,n}(\alpha', \mu'_\varepsilon, \theta') := 4 \left(n \sum_{k=1}^n Y_{k-1}^2 - \left(\sum_{k=1}^n Y_{k-1} \right)^2 \right),$$

$$\Delta_{3,n}(\alpha', \mu'_\varepsilon, \theta') := 8 \left((n-1) \sum_{k=1}^n Y_{k-1}^2 + 2Y_{s-1} \sum_{k=1}^n Y_{k-1} - n(Y_{s-1})^2 - \left(\sum_{k=1}^n Y_{k-1} \right)^2 \right).$$

We show that for all $(\alpha', \mu'_\varepsilon, \theta') \in \mathbb{R}^3$,

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\Delta_{1,n}(\alpha', \mu'_\varepsilon, \theta')}{n} = 2\mathsf{E}\widetilde{X}^2 \right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\Delta_{2,n}(\alpha', \mu'_\varepsilon, \theta')}{n^2} = 4 \operatorname{Var} \widetilde{X} \right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\Delta_{3,n}(\alpha', \mu'_\varepsilon, \theta')}{n^2} = 8 \operatorname{Var} \widetilde{X} \right) = 1,$$

where $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1). By (2.1.4), (2.1.5) and Proposition 3.1.1, it is enough to check that

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n Z_{k-1} = 0 \right) = 1, \qquad \mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n Z_{k-1}^2 = 0 \right) = 1, \tag{4.2.2}$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n X_{k-1} Z_{k-1} = 0 \right) = 1. \tag{4.2.3}$$

By Proposition 3.1.1, $\mathsf{P}(\lim_{n\to\infty} Z_n = 0) = 1$, which readily follows (4.2.2). Using Cauchy-Schwartz's inequality, we get

$$\frac{1}{n}\left|\sum_{k=1}^{n} X_{k-1}Z_{k-1}\right| \leqslant \sqrt{\frac{1}{n}\sum_{k=1}^{n} X_{k-1}^2 \frac{1}{n}\sum_{k=1}^{n} Z_{k-1}^2} \to \sqrt{\mathsf{E}\widetilde{X}^2}\sqrt{\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} Z_{k-1}^2} = 0$$

almost surely as $n\to\infty$, which implies (4.2.3). Hence

$$\mathsf{P}\left(\lim_{n\to\infty} \Delta_{i,n}(\alpha', \mu_\varepsilon', \theta') = \infty\right) = 1, \qquad \forall\ (\alpha', \mu_\varepsilon', \theta') \in \mathbb{R}^3, \qquad i = 1, 2, 3,$$

which yields that $H_n(\alpha', \mu_\varepsilon', \theta')$ has positive leading principal minors asymptotically as $n\to\infty$ with probability one. Then $H_n(\alpha', \mu_\varepsilon', \theta')$ is (strictly) positive definite asymptotically as $n\to\infty$ with probability one. Hence, using also that for all $n\in\mathbb{N}$ and $\mathbf{y}_n \in \mathbb{R}^{n+1}$, the function $\mathbb{R}^3 \ni (\alpha', \mu_\varepsilon', \theta') \mapsto Q_n(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta')$ is continuous and

$$\lim_{(\alpha')^2+(\mu_\varepsilon')^2+(\theta')^2\to\infty} Q_n(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta') = \infty,$$

we get the (random) function $\mathbb{R}^3 \ni (\alpha', \mu_\varepsilon', \theta') \mapsto Q_n(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta')$ attains its minimum asymptotically as $n\to\infty$ with probability one. Using the strict convexity of the function in question, we get (4.2.1) has a unique solution with respect to $(\alpha', \mu_\varepsilon', \theta') \in \mathbb{R}^3$ for all $n \geqslant n(\omega)$, see, e.g., Berkovitz [6, Theorem 3.1 and Corollary 1, Chapter IV]. This completes the proof. □

Proof of Theorem 3.3.1. Let us introduce the notation

$$D_n := (n-1)\sum_{k=1}^{n}{}^{(s)} Y_{k-1}^2 - \left(\sum_{k=1}^{n}{}^{(s)} Y_{k-1}\right)^2, \qquad n\in\mathbb{N}.$$

By (2.1.4), (2.1.5), (4.2.2), (4.2.3) and Proposition 3.1.1, we get

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{D_n}{n^2} = \operatorname{Var}\widetilde{X}\right) = 1, \tag{4.2.4}$$

where $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1).

Solving the system of equations (4.2.1) and applying Proposition 3.1.1, the following equalities hold asymptotically as $n\to\infty$ with probability one,

$$\begin{bmatrix}\widehat{\alpha}_n\\ \widehat{\mu}_{\varepsilon,n}\end{bmatrix} = \frac{1}{D_n}\begin{bmatrix}K_n\\ L_n\end{bmatrix}, \qquad \widehat{\theta}_n = Y_s - \widehat{\alpha}_n Y_{s-1} - \widehat{\mu}_{\varepsilon,n}, \tag{4.2.5}$$

where

$$K_n := (n-1)\sum_{k=1}^{n}{}^{(s)}(X_{k-1}+Z_{k-1})(X_k+Z_k) - \sum_{k=1}^{n}{}^{(s)}(X_k+Z_k)\sum_{k=1}^{n}{}^{(s)}(X_{k-1}+Z_{k-1}),$$

$$L_n := \sum_{k=1}^{n}{}^{(s)}(X_{k-1}+Z_{k-1})^2\sum_{k=1}^{n}{}^{(s)}(X_k+Z_k) - \sum_{k=1}^{n}{}^{(s)}(X_{k-1}+Z_{k-1})\sum_{k=1}^{n}{}^{(s)}(X_{k-1}+Z_{k-1})(X_k+Z_k).$$

Using (2.1.4), (2.1.5), (2.1.6), (4.2.2), (4.2.3) and that

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}Z_{k-1}Z_k = 0\right) = 1,$$

we obtain

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{K_n}{n^2} = \alpha\mathsf{E}\widetilde{X}^2 + \mu_\varepsilon\mathsf{E}\widetilde{X} - (\mathsf{E}\widetilde{X})^2\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{L_n}{n^2} = \mathsf{E}\widetilde{X}^2\mathsf{E}\widetilde{X} - \mathsf{E}\widetilde{X}(\alpha\mathsf{E}\widetilde{X}^2 + \mu_\varepsilon\mathsf{E}\widetilde{X})\right) = 1.$$

By (2.1.2), (2.1.3) and (4.2.4), we get

$$\mathsf{P}\left(\lim_{n\to\infty}\widehat{\alpha}_n = \lim_{n\to\infty}\frac{K_n}{D_n} = \frac{\alpha\operatorname{Var}\widetilde{X} + (\alpha-1)(\mathsf{E}\widetilde{X})^2 + \mu_\varepsilon\mathsf{E}\widetilde{X}}{\operatorname{Var}\widetilde{X}} = \alpha\right) = 1,$$

and

$$\mathsf{P}\left(\lim_{n\to\infty}\widehat{\mu}_{\varepsilon,n} = \lim_{n\to\infty}\frac{L_n}{D_n} = \frac{(1-\alpha)\mathsf{E}\widetilde{X}\mathsf{E}\widetilde{X}^2 - \mu_\varepsilon(\mathsf{E}\widetilde{X})^2}{\operatorname{Var}\widetilde{X}} = \mu_\varepsilon\right) = 1.$$

Finally, using (4.2.5), (3.3.1) and (3.3.2) we have (3.3.3). This completes the proof. □

Proof of Theorem 3.3.2. By (4.2.5) and Proposition 3.1.1, asymptotically as $n\to\infty$ with probability one $\widehat{\alpha}_n - \alpha$ and $\widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon$ take the forms

$$\frac{1}{D_n}\left((n-1)\sum_{k=1}^{n}{}^{(s)}(X_k-\alpha X_{k-1})X_{k-1} - \sum_{k=1}^{n}{}^{(s)}(X_k-\alpha X_{k-1})\sum_{k=1}^{n}{}^{(s)}X_{k-1} + R_n\right)$$

and

$$\frac{1}{D_n}\left(\sum_{k=1}^{n}{}^{(s)}X_{k-1}^2\sum_{k=1}^{n}{}^{(s)}(X_k-\mu_\varepsilon) - \sum_{k=1}^{n}{}^{(s)}X_{k-1}\sum_{k=1}^{n}{}^{(s)}(X_k-\mu_\varepsilon)X_{k-1} + S_n\right),$$

respectively, where

$$
\begin{aligned}
R_n := &(n-1)\sum_{k=1}^{n}{}^{(s)}(Z_k - \alpha Z_{k-1})(X_{k-1} + Z_{k-1}) + (n-1)\sum_{k=1}^{n}{}^{(s)}(X_k - \alpha X_{k-1})Z_{k-1} \\
&- \sum_{k=1}^{n}{}^{(s)}(Z_k - \alpha Z_{k-1})\sum_{k=1}^{n}{}^{(s)}(X_{k-1} + Z_{k-1}) - \sum_{k=1}^{n}{}^{(s)}(X_k - \alpha X_{k-1})\sum_{k=1}^{n}{}^{(s)} Z_{k-1},
\end{aligned}
$$

and

$$
\begin{aligned}
S_n := &\sum_{k=1}^{n}{}^{(s)}(2X_{k-1}Z_{k-1} + Z_{k-1}^2)\sum_{k=1}^{n}{}^{(s)}(X_k + Z_k - \mu_\varepsilon) + \sum_{k=1}^{n}{}^{(s)} X_{k-1}^2 \sum_{k=1}^{n}{}^{(s)} Z_k \\
&- \sum_{k=1}^{n}{}^{(s)} Z_{k-1}\sum_{k=1}^{n}{}^{(s)}(X_k + Z_k - \mu_\varepsilon)(X_{k-1} + Z_{k-1}) \\
&- \sum_{k=1}^{n}{}^{(s)} X_{k-1}\sum_{k=1}^{n}{}^{(s)} Z_k(X_{k-1} + Z_{k-1}) - \sum_{k=1}^{n}{}^{(s)} X_{k-1}\sum_{k=1}^{n}{}^{(s)}(X_k - \mu_\varepsilon)Z_{k-1}.
\end{aligned}
$$

By (2.3.1), (4.2.4) and Slutsky's lemma (see, e.g., Lemma 2.8 in van der Vaart [30]), to prove (3.3.4) it is enough to check that

$$
\frac{R_n}{n^{3/2}} \stackrel{\mathsf{P}}{\longrightarrow} 0 \qquad \text{as} \quad n \to \infty, \tag{4.2.6}
$$

$$
\frac{S_n}{n^{3/2}} \stackrel{\mathsf{P}}{\longrightarrow} 0 \qquad \text{as} \quad n \to \infty, \tag{4.2.7}
$$

where $\stackrel{\mathsf{P}}{\longrightarrow}$ denotes convergence in probability. In order to prove (4.2.6) it is enough to check that

$$
\frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}(Z_k - \alpha Z_{k-1})X_{k-1} \stackrel{L_2}{\longrightarrow} 0 \qquad \text{as} \quad n \to \infty, \tag{4.2.8}
$$

$$
\frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}(Z_k - \alpha Z_{k-1})Z_{k-1} \stackrel{L_2}{\longrightarrow} 0 \qquad \text{as} \quad n \to \infty, \tag{4.2.9}
$$

$$
\frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}(X_k - \alpha X_{k-1})Z_{k-1} \stackrel{L_2}{\longrightarrow} 0 \qquad \text{as} \quad n \to \infty, \tag{4.2.10}
$$

$$
\frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}(Z_k - \alpha Z_{k-1}) \cdot \frac{1}{n}\sum_{k=1}^{n}{}^{(s)}(X_{k-1} + Z_{k-1}) \stackrel{\mathsf{P}}{\longrightarrow} 0 \qquad \text{as} \quad n \to \infty, \tag{4.2.11}
$$

$$
\frac{1}{n}\sum_{k=1}^{n}{}^{(s)}(X_k - \alpha X_{k-1}) \cdot \frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)} Z_{k-1} \stackrel{\mathsf{P}}{\longrightarrow} 0 \qquad \text{as} \quad n \to \infty. \tag{4.2.12}
$$

To prove (4.2.8), it is enough to check that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} \mathsf{E}\big[(Z_k - \alpha Z_{k-1})X_{k-1}\big] = 0, \qquad (4.2.13)$$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left(\sum_{k=1}^{n}{}^{(s)} (Z_k - \alpha Z_{k-1})X_{k-1}\right)^2 = 0. \qquad (4.2.14)$$

Indeed, if $(\eta_n)_{n\in\mathbb{N}}$ is a sequence of square integrable random variables such that $\lim_{n\to\infty} \mathsf{E}\eta_n = 0$ and $\lim_{n\to\infty} \mathsf{E}\eta_n^2 = 0$, then η_n converges in L_2 to 0 as $n \to \infty$. Since $\mathsf{E}Z_k = \alpha \mathsf{E}Z_{k-1}$, $k \geqslant s+1$, and the processes X and Z are independent, we have (4.2.13). Using that $Z_0 = \cdots = Z_{s-1} = 0$, we also get

$$\mathsf{E}\left(\sum_{k=1}^{n}{}^{(s)} (Z_k - \alpha Z_{k-1})X_{k-1}\right)^2 = \sum_{k=s+1}^{n} \mathsf{E}(Z_k - \alpha Z_{k-1})^2 \mathsf{E}X_{k-1}^2, \qquad n \geqslant s+1.$$

Since $\lim_{k\to\infty} \mathsf{E}X_k^2 = \mathsf{E}\widetilde{X}^2$ (see, e.g., Ispány, Pap and van Zuijlen [19, page 751]), there exists some $L > 0$ such that $\mathsf{E}X_k^2 < L$ for all $k \in \mathbb{N}$. By Proposition 3.1.1, $\lim_{k\to\infty} \mathsf{E}(Z_k - \alpha Z_{k-1})^2 \leqslant \lim_{k\to\infty} 2\mathsf{E}(Z_k^2 + \alpha^2 Z_{k-1}^2) = 0$, and hence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}(Z_k - \alpha Z_{k-1})^2 = 0.$$

This yields that

$$\frac{1}{n} \sum_{k=s+1}^{n} \mathsf{E}(Z_k - \alpha Z_{k-1})^2 \mathsf{E}X_{k-1}^2 \leqslant \frac{L}{n} \sum_{k=s+1}^{n} \mathsf{E}(Z_k - \alpha Z_{k-1})^2 \to 0 \qquad \text{as } n \to \infty,$$

which completes the proof of (4.2.14).

In a similar way one can prove (4.2.9) and (4.2.10). For a detailed proof, see Theorem 4.2.2 in Barczy et al. [4]. For (4.2.9), we only note that for all $n \geqslant s+1$,

$$\frac{1}{n} \mathsf{E}\left(\sum_{k=1}^{n}{}^{(s)} (Z_k - \alpha Z_{k-1})Z_{k-1}\right)^2 = \frac{\alpha(1-\alpha)}{n} \sum_{k=s+1}^{n} \mathsf{E}Z_{k-1}^3,$$

which tends to 0 as $n \to \infty$, by Proposition 3.1.1.

To prove (4.2.11), using that

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n}{}^{(s)} (X_{k-1} + Z_{k-1}) = \mathsf{E}\widetilde{X}\right) = 1,$$

it is enough to check that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} (Z_k - \alpha Z_{k-1}) \xrightarrow{L_2} 0 \qquad \text{as } n \to \infty,$$

which can be verified as earlier. Equality in (4.2.12) can be handled in the same way.

Now we turn to prove (4.2.7). Using (2.1.5) and that

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}{}^{(s)}(X_k+Z_k-\mu_\varepsilon)=\mathsf{E}\widetilde{X}-\mu_\varepsilon\right)=1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}{}^{(s)}(X_k+Z_k-\mu_\varepsilon)(X_{k-1}+Z_{k-1})=\alpha\mathsf{E}\widetilde{X}^2\right)=1,$$

it is enough to verify that

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}Z_{k-1}\stackrel{L_1}{\longrightarrow}0\qquad\text{as}\quad n\to\infty,\tag{4.2.15}$$

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}X_{k-1}Z_{k-1}\stackrel{L_1}{\longrightarrow}0\qquad\text{as}\quad n\to\infty,\tag{4.2.16}$$

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}Z_{k-1}^2\stackrel{L_1}{\longrightarrow}0\qquad\text{as}\quad n\to\infty,\tag{4.2.17}$$

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}Z_{k-1}Z_k\stackrel{L_1}{\longrightarrow}0\qquad\text{as}\quad n\to\infty.\tag{4.2.18}$$

To check (4.2.15), using that $Z_k\geqslant 0$, $k\in\mathbb{N}$, by Markov's inequality, it is enough to show that

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{k=s+1}^{n}\mathsf{E}Z_{k-1}=0.$$

Since, by (3.1.1), $\mathsf{E}Z_{s+k}=\theta\alpha^k$, $k\geqslant 0$, we have

$$\frac{1}{\sqrt{n}}\sum_{k=s+1}^{n}\mathsf{E}Z_{k-1}\leqslant\frac{\theta}{\sqrt{n}}\sum_{k=0}^{n}\alpha^k=\frac{\theta}{\sqrt{n}}\frac{\alpha^{n+1}-1}{\alpha-1}\to 0\qquad\text{as}\quad n\to\infty.\tag{4.2.19}$$

This completes the proof of (4.2.15).

To prove (4.2.16), using that the processes X and Z are non-negative, by Markov's inequality, it is enough to show that

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}\mathsf{E}(X_{k-1}Z_{k-1})=0.$$

Using that the processes X and Z are independent and $\lim_{k\to\infty}\mathsf{E}X_{k-1}=\mathsf{E}\widetilde{X}$ (see, e.g., Ispány, Pap and van Zuijlen [19, page 751]), as in the proof of (4.2.8), it is enough to check (4.2.19), which was already done.

Using Proposition 3.1.2, by similar arguments as earlier, one can check (4.2.17) and (4.2.18).

Finally, using (4.2.5) and (3.3.3), we get

$$\sqrt{n}(\widehat{\theta}_n - \lim_{k\to\infty} \widehat{\theta}_k) = -\sqrt{n}(\widehat{\alpha}_n - \alpha)Y_{s-1} - \sqrt{n}(\widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon)$$

$$= \begin{bmatrix} -Y_{s-1} & -1 \end{bmatrix} \begin{bmatrix} \sqrt{n}(\widehat{\alpha}_n - \alpha) \\ \sqrt{n}(\widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon) \end{bmatrix},$$

and hence, by (3.3.4), we have (3.3.5). This completes the proof. □

References

[1] Abraham, B. and Chuang, A. (1993). Expectation-maximization algorithms and the estimation of time series models in the presence of outliers. *Journal of Time Series Analysis* **14,** 221–234.

[2] Al-Osh, M. A. and Alzaid, A. A. (1987). First order integer-valued autoregressive INAR(1) process. *Journal of Time Series Analysis* **8**(3), 261–275.

[3] Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business & Economic Statistics* **11**(1), 81–92.

[4] Barczy, M., Ispány, M., Pap, G., Scotto, M. and Silva, M. E. (2009). Outliers in INAR(1) models. *Arxiv*, URL: http://arxiv.org/abs/0903.2421

[5] Barczy, M., Ispány, M., Pap, G., Scotto, M. and Silva, M. E. (2009). Additive outliers in INAR(1) models. Submitted to *Methodol. Comput. Appl. Probab.*

[6] Berkovitz, L. D. (2002). *Convexity and optimization in* $\mathbb{R}^n$. Wiley, New York.

[7] Bhattacharya, R. N. and Waymire, E. C. (1990). *Stochastic processes with applications.* Wiley, New York.

[8] Burridge, P. and Taylor, A. M. (2006). Additive outlier detection via extreme value theory. *Journal of Time Series Analysis* **27,** 685–701.

[9] Bustos, O. H. and Yohai, V. J. (1986). Robust estimates for ARMA models. *Journal of the American Statistical Association* **81,** 155–168.

[10] Chang, G. C. I. T. and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics* **30,** 193–204.

[11] Chen, C. and Tiao, G. C. (1990). Random level-shift time series models, ARIMA approximations, and level-shift detection. *Journal of Business & Economic Statistics* **8,** 83–97.

[12] Chung, K. L. (1960). *Markov chains with stationary transition probabilities.* Springer, Berlin-Göttingen-Heidelberg.

[13] Du, J. G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis* **12,** 129–142.

[14] Drost, F. C., van den Akker, R. and Werker, B. J. M. (2008). Local asymptotic normality and efficient estimation for INAR(p) models. *Journal of Time Series Analysis* **29**(5), 783–801.

[15] Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **34,** 350–363.

[16] Guttman, I. and Tiao, G. C. (1978). Effect of correlation on the estimation of a mean in the presence of spurious observations. *The Canadian Journal of Statistics* **6,** 229–247.

[17] Latour, A. (1988). Existence and stochastic structure of a non-negative integer-valued autoregressive processes. *Journal of Time Series Analysis* **4,** 439–455.

[18] Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic Press, New-York-London.

[19] Ispány, M., Pap, G. and van Zuijlen, M. (2003). Asymptotic inference for nearly unstable INAR(1) models. *Journal of Applied Probability* **40**(3), 750–765.

[20] Kachour, M. (2009). p-order rounded integer-valued autoregressive (RINAR(p)) process. *Arxiv*, URL: http://arxiv.org/abs/0902.1598

[21] Kachour, M. and Yao, J. F. (2008). First-order rounded integer-valued autoregressive (RINAR(1)) process. *Preprint*, University of Rennes 1.

[22] McCulloch, R. E. and Tsay, R. S. (1993). Bayesian inference and prediction for mean and variance shifts in autocovariance time series. *Journal of the American Statistical Association* **88,** 986–978.

[23] McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin* **21,** 645–650.

[24] Peña, D. (2001). Outliers, influential observations, and missing data. In: Peña D., Tiao G. C., Tsay R. S. (eds) *A Course in Time Series Analysis.* John Wiley & Sons, 136–170.

[25] Perron, P. and Rodriguez, G. (2003). Searching for additive outliers in non-stationary time series. *Journal of Time Series Analysis* **24,** 193-220.

[26] Sánchez, M. J. and Peña, D. (2003). The identification of multiple outliers in ARIMA models. *Communications in Statistics. Theory and Methods* **32,** 1265–1287.

[27] Silva, M. E. and Oliveira, V. L. (2004). Difference equations for the higher-order moments and cumulants of the INAR(1) model. *Journal of Time Series Analysis* **25**(3), 317–333.

[28] Steutel, F. W and van Harn, K. (1979). Discrete analogues of self–decomposability and stability. *The Annals of Probability* **7**, 893–99.

[29] Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting* **7,** 1–20.

[30] van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.