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;
** University of Szeged, Bolyai Institute, H-6720 Szeged, Aradi vértanúk tere 1, 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.

e–mails: barczy.matyas@inf.unideb.hu (M. Barczy), ispany.marton@inf.unideb.hu (M. Ispány), papgy@math.u-szeged.hu (G. Pap), mscotto@ua.pt (M. Scotto), mesilva@fep.up.pt (M. E. Silva).

◇ Corresponding author.

Abstract

In this paper the integer-valued autoregressive model of order one, contaminated with additive or innovational outliers is studied in some detail. Moreover, parameter estimation is also addressed. Supposing that the time points 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. This random limit depends on the values of the process at the outliers' time points and on the values at the preceding time points and in case of additive outliers also on the values at the following time points. We also prove that the joint CLS estimator of the offspring and innovation means is asymptotically normal. Conditionally on the above described values of the process, the joint CLS estimator of the sizes of the outliers is also asymptotically normal.

2000 Mathematics Subject Classifications: 60J80, 62F12.

Key words and phrases: integer-valued autoregressive models, additive and 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. Mátyás Barczy and Gyula Pap were partially supported by the Hungarian Scientific Research Fund under Grants No. OTKA-T-079128.

Contents

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., MacDonald and Zucchini [47], Cameron and Trivedi [21], and Steutel and van Harn [61]) and review papers (e.g., McKenzie [49], Jung and Tremayne [40], and Weiß [64]). Motivation to include discrete data models comes from the need to account for the discrete nature of certain data sets, often counts of events, objects or individuals. Examples of applications can be found in the analysis of time series of count data that are generated from stock transactions (Quoreshi [54]), where each transaction refers to a trade between a buyer and a seller in a volume of stocks for a given price, in optimal alarm systems to predict whether a count process will upcross a certain level and give an alarm whenever the upcrossing level is predicted (Monteiro, Pereira and Scotto [50]), international tourism demand (Brännäs and Nordström [16]), experimental biology (Zhou and Basawa [69]), social science (McCabe and Martin [45]), and queueing systems (Ahn, Gyemin and Jongwoo [2]).

Several integer-valued time series models were proposed in the literature, we mention the INteger-valued AutoRegressive model of order p (INAR(p)) and the INteger-valued Moving Average model of order q (INMA(q)). The former was first introduced by McKenzie [48] and Al-Osh and Alzaid [3] for the case $p = 1$. The INAR(1) and INAR(p) models have been investigated by several authors, see, e.g., Silva and Oliveira [56], [57], Silva and Silva [58], Ispány, Pap and van Zuijlen [37], [38], [39], Drost, van den Akker and Werker [28] (local asymptotic normality for INAR(p) models), Györfi, Ispány, Pap and Varga [34], Ispány [36], Drost, van den Akker and Werker [26], [27] (nearly unstable INAR(1) models and semiparametric INAR(p) models), Bu and McCabe [18] (model selection) and Andersson and Karlis [7] (missing values). Empirical relevant extensions have been suggested by Brännäs [13] (explanatory variables), Blundell [12] (panel data), Brännäs and Hellström [15] (extended dependence structure), and more recently by Silva, Silva, Pereira and Silva [59] (replicated data) and by Weiß [66] (combined INAR(p) models). Extensions and generalizations were proposed by Du and Li [30] and Latour [44]. The INMA(q) model was proposed by Al-Osh and Alzaid [4], and subsequently studied by Brännäs and Hall [14] and Weiß [65]. Related models were introduced by Aly and Bouzar [5], [6], Zhu and Joe [70] and more recently by Weiß [65]. Extensions for random coefficients integer-valued autoregressive models have been proposed by Zheng, Basawa and Datta [67], [68] who investigated basic probabilistic and statistical properties of these models. Zheng and co-authors illustrated their performance in the analysis of epileptic seizure counts (e.g., Latour [44]) and in the analysis of the monthly number cases of poliomyelitis in the US for the period 1970-1983. Doukhan, Latour and Oraichi [25] introduced the class of non-negative integer-valued bilinear time series, some of their results concerning the existence of stationary solutions were extended by Drost, van den Akker and Werker [29]. Recently, the so called p-order Rounded INteger-valued AutoRegressive (RINAR(p)) time series model was introduced and studied by Kachour and Yao [42] and Kachour [41].

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 [32] introduced the notion of additive and innovational outliers and proposed the use of maximum likelihood ratio test to detect them. Chang and Chen [22] 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 [62] for the detection of level shifts and temporary changes. Random level shifts were studied by Chen and Tiao [23]. Extensions of Tsay's results can be found in Balke [9]. 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 [33], Bustos and Yohai [20], McCulloch and Tsay [46], Peña [52], Sánchez and Peña [55], Perron and Rodriguez [53] and Burridge and Taylor [19].

It is worth mentioning that all references given in the previous paragraph deal with the case of 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 [51, 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 or innovational outliers starting from a general initial distribution (having finite second or third moments). We suppose 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 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' time points and on the values at the preceding time points and in case of additive outliers also on the values at the following time points. 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 above described values of the process, 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 present results in the case of one or two additive or innovational outliers for INAR(1) models, the general case of finitely many additive or innovational 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 with the asymptotic behavior of CLS estimator for the INAR(1) model. In Sections 3 and 4 we consider INAR(1) models contaminated with one or two additive or innovational outliers, respectively. The cases of one outlier and two outliers are handled separately. Section 5 is an appendix containing the proofs of some auxiliary results.

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 and stability 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 [60]) 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 INAR(1) 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 always assume that $\mathsf{E}X_0^2 < \infty$ and that $\mathsf{E}\varepsilon_1^2 < \infty$, $\mathsf{P}(\varepsilon_1 \neq 0) > 0$. 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 easy to show that

$$\lim_{k\to\infty} \mathsf{E}X_k = \frac{\mu_\varepsilon}{1-\alpha}, \qquad \lim_{k\to\infty} \operatorname{Var} X_k = \frac{\sigma_\varepsilon^2 + \alpha\mu_\varepsilon}{1-\alpha^2}, \qquad \text{if } \alpha \in (0,1), \tag{2.1.2}$$

and that $\lim_{k\to\infty} \mathsf{E}X_k = \lim_{k\to\infty} \operatorname{Var} X_k = \infty$ if $\alpha = 1$ (e.g., Ispány, Pap and van Zuijlen [37, page 751]). The case $\alpha \in (0,1)$ is called *stable* or *asymptotically stationary*, whereas the case

$\alpha = 1$ is called *unstable*. For the stable case, there exists a unique stationary distribution of the INAR(1) model in (2.1.1), see Lemma 5.1 in the Appendix.

In the sequel we assume that $\alpha \in (0,1)$, and we denote by $\mathcal{F}_k^X$ the σ–algebra generated by the random variables $X_0, X_1, \ldots, X_k$.

2.2 Estimation of the mean of the offspring distribution

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

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

For all $n \in \mathbb{N}$, we define the function $Q_n : \mathbb{R}^{n+1} \times \mathbb{R} \to \mathbb{R}$, as

$$Q_n(x_0, x_1, \ldots, x_n; \alpha') := \sum_{k=1}^{n} \big(x_k - \alpha' x_{k-1} - \mu_\varepsilon\big)^2, \qquad x_0, x_1, \ldots, x_n, \alpha' \in \mathbb{R}.$$

By definition, for all $n \in \mathbb{N}$, a CLS estimator for the parameter $\alpha \in (0,1)$ is a measurable function $\widetilde{\alpha}_n : \mathbb{R}^{n+1} \to \mathbb{R}$ such that

$$Q_n(x_0, x_1, \ldots, x_n; \widetilde{\alpha}_n(x_0, x_1, \ldots, x_n)) = \inf_{\alpha' \in \mathbb{R}} Q_n(x_0, x_1, \ldots, x_n; \alpha') \qquad \forall\ (x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1}.$$

It is well-known that

$$\widetilde{\alpha}_n(X_0, X_1, \ldots, X_n) = \frac{\sum_{k=1}^n (X_k - \mu_\varepsilon) X_{k-1}}{\sum_{k=1}^n X_{k-1}^2} \tag{2.2.1}$$

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 $S \in \mathcal{A}$ such that $\mathsf{P}(S) = 1$ and for all $\omega \in S$ there exists an $n(\omega) \in \mathbb{N}$ such that the property in question holds for all $n \geqslant n(\omega)$. The reason why (2.2.1) 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 the later formula (2.2.6)). In what follows we simply denote $\widetilde{\alpha}_n(X_0, X_1, \ldots, X_n)$ by $\widetilde{\alpha}_n$. Using the same arguments as in Hall and Heyde [35, Section 6.3], one can easily check that $\widetilde{\alpha}_n$ is a strongly consistent estimator of α as $n \to \infty$ for all $\alpha \in (0,1)$, i.e.,

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\sum_{k=1}^n (X_k - \mu_\varepsilon) X_{k-1}}{\sum_{k=1}^n X_{k-1}^2} = \alpha\right) = 1, \qquad \forall\, \alpha \in (0,1). \tag{2.2.2}$$

Namely, if $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1), then

$$\mathsf{E}\widetilde{X} = \frac{\mu_\varepsilon}{1-\alpha}, \tag{2.2.3}$$

$$\mathsf{E}\widetilde{X}^2 = \frac{\sigma_\varepsilon^2 + \alpha\mu_\varepsilon}{1-\alpha^2} + \frac{\mu_\varepsilon^2}{(1-\alpha)^2}. \tag{2.2.4}$$

For the proofs of (2.2.3) and (2.2.4), see the Appendix. By the existence of a unique stationary distribution, we obtain that $\{i \in \mathbb{Z}_+ : i \geqslant i_{\min}\}$ with

$$i_{\min} := \min\Big\{i \in \mathbb{Z}_+ : \mathsf{P}(\varepsilon_1 = i) > 0\Big\}$$

is a positive recurrent class of the Markov chain $(X_k)_{k\in\mathbb{Z}_+}$ (see, e.g., Bhattacharya and Waymire [11, Section II, Theorem 9.4 (c)] or Chung [24, Section I.6, Theorem 4 and Section I.7, Theorem 2]). By ergodic theorems (see, e.g., Bhattacharya and Waymire [11, Section II, Theorem 9.4 (d)] or Chung [24, Section I.15, Theorem 2]), we get

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k = \mathsf{E}\widetilde{X}\right) = 1, \tag{2.2.5}$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k^2 = \mathsf{E}\widetilde{X}^2\right) = 1, \tag{2.2.6}$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_{k-1}X_k = \mathsf{E}(\widetilde{X}(\alpha\circ\widetilde{X}+\varepsilon)) = \alpha\mathsf{E}\widetilde{X}^2 + \mu_\varepsilon\mathsf{E}\widetilde{X}\right) = 1, \tag{2.2.7}$$

where ε is a random variable independent of $\widetilde{X}$ with the same distribution as ε_1. (For (2.2.7), one uses that the distribution of $(\widetilde{X}, \alpha\circ\widetilde{X}+\varepsilon)$ is the unique stationary distribution of the Markov chain $(X_k, X_{k+1})_{k\in\mathbb{Z}_+}$.) By (2.2.5)–(2.2.7),

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

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 [35, Section 6.3], it follows easily that

$$\sqrt{n}(\widetilde{\alpha}_n - \alpha) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, \sigma^2_{\alpha,\varepsilon}) \qquad \text{as } n\to\infty, \tag{2.2.8}$$

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

$$\sigma^2_{\alpha,\varepsilon} := \frac{\alpha(1-\alpha)\mathsf{E}\widetilde{X}^3 + \sigma_\varepsilon^2\mathsf{E}\widetilde{X}^2}{(\mathsf{E}\widetilde{X}^2)^2}, \tag{2.2.9}$$

with

$$\begin{aligned}\mathsf{E}\widetilde{X}^3 = {} & \frac{\mathsf{E}\varepsilon^3 - 3\sigma_\varepsilon^2(1+\mu_\varepsilon) - \mu_\varepsilon^3 + 2\mu_\varepsilon}{1-\alpha^3} + 3\frac{\sigma_\varepsilon^2 + \alpha\mu_\varepsilon}{1-\alpha^2} - 2\frac{\mu_\varepsilon}{1-\alpha} \\ & + 3\frac{\mu_\varepsilon(\sigma_\varepsilon^2 + \alpha\mu_\varepsilon)}{(1-\alpha)(1-\alpha^2)} + \frac{\mu_\varepsilon^3}{(1-\alpha)^3}.\end{aligned} \tag{2.2.10}$$

For the proof of (2.2.10), see the Appendix.

We remark that one uses in fact Corollary 3.1 in Hall and Heyde [35] to derive (2.2.8). 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 $n \in \mathbb{N}$, we define the function $Q_n : \mathbb{R}^{n+1} \times \mathbb{R}^2 \to \mathbb{R}$, as

$$Q_n(x_0, x_1, \ldots, x_n; \alpha', \mu'_\varepsilon) := \sum_{k=1}^{n} \big(x_k - \alpha' x_{k-1} - \mu'_\varepsilon\big)^2, \qquad x_0, x_1, \ldots, x_n, \alpha', \mu'_\varepsilon \in \mathbb{R}.$$

By definition, for all $n \in \mathbb{N}$, a CLS estimator for the parameter $(\alpha, \mu_\varepsilon) \in (0,1) \times (0, \infty)$ is a measurable function $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}) : \mathbb{R}^{n+1} \to \mathbb{R}^2$ such that

$$\begin{aligned} &Q_n(x_0, x_1, \ldots, x_n; \widehat{\alpha}_n(x_0, x_1, \ldots, x_n), \widehat{\mu}_{\varepsilon,n}(x_0, x_1, \ldots, x_n)) \\ &\qquad = \inf_{(\alpha', \mu'_\varepsilon) \in \mathbb{R}^2} Q_n(x_0, x_1, \ldots, x_n; \alpha', \mu'_\varepsilon) \qquad \forall\ (x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1}. \end{aligned}$$

It is well-known that

$$\begin{aligned} &\sum_{k=1}^{n} (X_k - \widehat{\alpha}_n X_{k-1} - \widehat{\mu}_{\varepsilon,n}) X_{k-1} = 0, \\ &\sum_{k=1}^{n} (X_k - \widehat{\alpha}_n X_{k-1} - \widehat{\mu}_{\varepsilon,n}) = 0, \end{aligned}$$

hold asymptotically as $n \to \infty$ with probability one, or equivalently

$$\begin{bmatrix} \sum_{k=1}^n X_{k-1}^2 & \sum_{k=1}^n X_{k-1} \\ \sum_{k=1}^n X_{k-1} & n \end{bmatrix} \begin{bmatrix} \widehat{\alpha}_n \\ \widehat{\mu}_{\varepsilon,n} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n X_{k-1} X_k \\ \sum_{k=1}^n X_k \end{bmatrix}$$

holds asymptotically as $n \to \infty$ with probability one. Using that, by (2.2.5) and (2.2.6),

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n^2} \left(n \sum_{k=1}^{n} X_{k-1}^2 - \left(\sum_{k=1}^{n} X_{k-1} \right)^2 \right) = \mathsf{E}\widetilde{X}^2 - (\mathsf{E}\widetilde{X})^2 = \operatorname{Var} \widetilde{X} > 0 \right) = 1,$$

we get

$$\widehat{\alpha}_n(X_0, X_1, \ldots, X_n) = \frac{n \sum_{k=1}^n X_{k-1} X_k - \left(\sum_{k=1}^n X_{k-1}\right)\left(\sum_{k=1}^n X_k\right)}{n \sum_{k=1}^n X_{k-1}^2 - \left(\sum_{k=1}^n X_{k-1}\right)^2},$$

$$\begin{aligned} \widehat{\mu}_{\varepsilon,n}(X_0, X_1, \ldots, X_n) &= \frac{\left(\sum_{k=1}^n X_{k-1}^2\right)\left(\sum_{k=1}^n X_k\right) - \left(\sum_{k=1}^n X_{k-1}\right)\left(\sum_{k=1}^n X_{k-1} X_k\right)}{n \sum_{k=1}^n X_{k-1}^2 - \left(\sum_{k=1}^n X_{k-1}\right)^2} \\ &= \frac{1}{n} \left(\sum_{k=1}^{n} X_k - \widehat{\alpha}_n \sum_{k=1}^{n} X_{k-1} \right), \end{aligned}$$

hold asymptotically as $n \to \infty$ with probability one, see, e.g., Hall and Heyde [35, formulae (6.36) and (6.37)]. In the sequel we simply denote $\widehat{\alpha}_n(X_0, X_1, \ldots, X_n)$ and $\widehat{\mu}_{\varepsilon,n}(X_0, X_1, \ldots, X_n)$ by $\widehat{\alpha}_n$ and $\widehat{\mu}_{\varepsilon,n}$, respectively. 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 [35, Section 6.3]. Moreover, if $\mathsf{E}X_0^3 < \infty$ and $\mathsf{E}\varepsilon_1^3 < \infty$, by Hall and Heyde [35, formula (6.44)],

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

where

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

$$A_{\alpha,\varepsilon} := \alpha(1-\alpha) \begin{bmatrix} \mathsf{E}\widetilde{X}^3 & \mathsf{E}\widetilde{X}^2 \\ \mathsf{E}\widetilde{X}^2 & \mathsf{E}\widetilde{X} \end{bmatrix} + \sigma_\varepsilon^2 \begin{bmatrix} \mathsf{E}\widetilde{X}^2 & \mathsf{E}\widetilde{X} \\ \mathsf{E}\widetilde{X} & 1 \end{bmatrix}, \tag{2.3.3}$$

and $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1). For our later purposes, we sketch a proof of (2.3.1). Using that

$$\begin{bmatrix} \widehat{\alpha}_n \\ \widehat{\mu}_{\varepsilon,n} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n X_{k-1}^2 & \sum_{k=1}^n X_{k-1} \\ \sum_{k=1}^n X_{k-1} & n \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^n X_{k-1}X_k \\ \sum_{k=1}^n X_k \end{bmatrix}$$

holds asymptotically as $n \to \infty$ with probability one, we obtain

$$\begin{aligned} \begin{bmatrix} \widehat{\alpha}_n - \alpha \\ \widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon \end{bmatrix} &= \\ &= \begin{bmatrix} \sum_{k=1}^n X_{k-1}^2 & \sum_{k=1}^n X_{k-1} \\ \sum_{k=1}^n X_{k-1} & n \end{bmatrix}^{-1} \left(\begin{bmatrix} \sum_{k=1}^n X_{k-1}X_k \\ \sum_{k=1}^n X_k \end{bmatrix} - \begin{bmatrix} \sum_{k=1}^n X_{k-1}^2 & \sum_{k=1}^n X_{k-1} \\ \sum_{k=1}^n X_{k-1} & n \end{bmatrix} \begin{bmatrix} \alpha \\ \mu_\varepsilon \end{bmatrix} \right) \\ &= \begin{bmatrix} \sum_{k=1}^n X_{k-1}^2 & \sum_{k=1}^n X_{k-1} \\ \sum_{k=1}^n X_{k-1} & n \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^n (X_k - \alpha X_{k-1} - \mu_\varepsilon) X_{k-1} \\ \sum_{k=1}^n (X_k - \alpha X_{k-1} - \mu_\varepsilon) \end{bmatrix} \end{aligned}$$

holds asymptotically as $n \to \infty$ with probability one. By (2.2.5) and (2.2.6), we have

$$\frac{1}{n} \begin{bmatrix} \sum_{k=1}^n X_{k-1}^2 & \sum_{k=1}^n X_{k-1} \\ \sum_{k=1}^n X_{k-1} & n \end{bmatrix} \to \begin{bmatrix} \mathsf{E}\widetilde{X}^2 & \mathsf{E}\widetilde{X} \\ \mathsf{E}\widetilde{X} & 1 \end{bmatrix} \qquad \text{as } n \to \infty \text{ with probability one,}$$

and, by Hall and Heyde [35, Section 6.3, formula (6.43)],

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

Using Slutsky's lemma, we get (2.3.1).

Let us introduce some notations which will be used throughout the paper. For all $k, \ell \in \mathbb{Z}_+$, let

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

For a sequence of random variables $(\zeta_k)_{k\in\mathbb{N}}$ and for $s_1, \dots, s_I \in \mathbb{N}$, $I \in \mathbb{N}$, we define

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

3 The INAR(1) model with additive outliers

3.1 The model

In this section we only introduce INAR(1) models contaminated with additive outliers.

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

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

where $(X_k)_{k\in\mathbb{Z}_+}$ *is an INAR(1) process given by* (2.1.1) *with* $\alpha \in (0,1)$, $\mathsf{E}X_0^2 < \infty$, $\mathsf{E}\varepsilon_1^2 < \infty$, $\mathsf{P}(\varepsilon_1 \neq 0) > 0$, *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 θ_i, $i = 1, \dots, I$, represents the ith additive outlier's size and δ_{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 σ–algebra generated by the random variables $Y_0, Y_1, \dots, Y_k$. For all $n \in \mathbb{N}$, $y_0, \dots, y_n \in \mathbb{R}$ and $\omega \in \Omega$, let us introduce the notations

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

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

First we assume 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) with $\theta := \theta_1$. An easy calculation shows that

$$\begin{aligned}
\mathsf{E}(Y_k \mid \mathcal{F}_{k-1}^Y) &= \alpha X_{k-1} + \mu_\varepsilon + \delta_{k,s}\theta = \alpha(Y_{k-1} - \delta_{k-1,s}\theta) + \mu_\varepsilon + \delta_{k,s}\theta \\
&= \alpha Y_{k-1} + \mu_\varepsilon + (-\alpha\delta_{k-1,s} + \delta_{k,s})\theta = \begin{cases} \alpha Y_{k-1} + \mu_\varepsilon & \text{if } k = 1, \dots, s-1, \\ \alpha Y_{k-1} + \mu_\varepsilon + \theta & \text{if } k = s, \\ \alpha Y_{k-1} + \mu_\varepsilon - \alpha\theta & \text{if } k = s+1, \\ \alpha Y_{k-1} + \mu_\varepsilon & \text{if } k \geqslant s+2. \end{cases}
\end{aligned}$$

Hence

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

For all $n \geqslant s+1$, $n \in \mathbb{N}$, we define the function $Q_n : \mathbb{R}^{n+1} \times \mathbb{R}^2 \to \mathbb{R}$, as

$$Q_n(\mathbf{y}_n; \alpha', \theta') := \sum_{k=1}^{n}{}^{(s,s+1)} \big(y_k - \alpha' y_{k-1} - \mu_\varepsilon\big)^2 + \big(y_s - \alpha' y_{s-1} - \mu_\varepsilon - \theta'\big)^2 + \big(y_{s+1} - \alpha' y_s - \mu_\varepsilon + \alpha'\theta'\big)^2, \qquad \mathbf{y}_n \in \mathbb{R}^{n+1}, \ \alpha', \theta' \in \mathbb{R}.$$

By definition, for all $n \geqslant s+1$, a CLS estimator for the parameter $(\alpha, \theta) \in (0,1) \times \mathbb{N}$ is a measurable function $(\widetilde{\alpha}_n, \widetilde{\theta}_n) : S_n \to \mathbb{R}^2$ such that

$$Q_n(\mathbf{y}_n; \widetilde{\alpha}_n(\mathbf{y}_n), \widetilde{\theta}_n(\mathbf{y}_n)) = \inf_{(\alpha',\theta') \in \mathbb{R}^2} Q_n(\mathbf{y}_n; \alpha', \theta') \qquad \forall \ \mathbf{y}_n \in S_n,$$

where S_n is a suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 3.2.1). We note that we do not define the CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_n)$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$. We have for all $(\mathbf{y}_n; \alpha', \theta') \in \mathbb{R}^{n+1} \times \mathbb{R}^2$,

$$\begin{aligned}
\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \theta') &= \sum_{k=1}^{n}{}^{(s,s+1)} 2\big(y_k - \alpha' y_{k-1} - \mu_\varepsilon\big)(-y_{k-1}) + 2\big(y_s - \alpha' y_{s-1} - \mu_\varepsilon - \theta'\big)(-y_{s-1}) \\
&\quad + 2\big(y_{s+1} - \alpha' y_s - \mu_\varepsilon + \alpha'\theta'\big)(-y_s + \theta') \\
&= \sum_{k=1}^{n} 2\big(y_k - \alpha' y_{k-1} - \mu_\varepsilon\big)(-y_{k-1}) - 2\theta'(-y_{s-1}) + 2\alpha'\theta'(-y_s + \theta') + 2\big(y_{s+1} - \alpha' y_s - \mu_\varepsilon\big)\theta',
\end{aligned}$$

and

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

The next lemma is about the existence and uniqueness of the CLS estimator of (α, θ).

3.2.1 Lemma. *There exist subsets $S_n \subset \mathbb{R}^{n+1}$, $n \geqslant s+1$ with the following properties:*

(i) *there exists a unique CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_n) : S_n \to \mathbb{R}^2$,*

(ii) *for all $\mathbf{y}_n \in S_n$, $(\widetilde{\alpha}_n(\mathbf{y}_n), \widetilde{\theta}_n(\mathbf{y}_n))$ is the unique solution of the system of equations*

$$\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \theta') = 0, \qquad \frac{\partial Q_n}{\partial \theta'}(\mathbf{y}_n; \alpha', \theta') = 0, \tag{3.2.2}$$

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

Proof. For any fixed $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $\alpha' \in \mathbb{R}$, the quadratic function $\mathbb{R} \ni \theta' \mapsto Q_n(\mathbf{y}_n; \alpha', \theta')$ can be written in the form

$$Q_n(\mathbf{y}_n; \alpha', \theta') = A_n(\alpha')\big(\theta' - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha')\big)^2 + \widetilde{Q}_n(\mathbf{y}_n; \alpha'),$$

where

$$A_n(\alpha') := 1 + (\alpha')^2,$$

$$t_n(\mathbf{y}_n; \alpha') := (1 + (\alpha')^2) y_s - \alpha'(y_{s-1} + y_{s+1}) - (1 - \alpha')\mu_\varepsilon,$$

$$\widetilde{Q}_n(\mathbf{y}_n; \alpha') := \sum_{k=1}^{n} \big(y_k - \alpha' y_{k-1}\big)^2 - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha')^2.$$

We have $\widetilde{Q}_n(\mathbf{y}_n; \alpha') = R_n(\mathbf{y}_n; \alpha')/A_n(\alpha')$, where $\mathbb{R} \ni \alpha' \mapsto R_n(\mathbf{y}_n; \alpha')$ is a polynomial of order 4 with leading coefficient

$$c_n(\mathbf{y}_n) := \sum_{k=1}^{n} y_{k-1}^2 - y_s^2.$$

Let

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

For $\mathbf{y}_n \in \widetilde{S}_n$, we have $\lim_{|\alpha'| \to \infty} \widetilde{Q}_n(\mathbf{y}_n; \alpha') = \infty$ and the continuous function $\mathbb{R} \ni \alpha' \mapsto \widetilde{Q}_n(\mathbf{y}_n; \alpha')$ attains its infimum. Consequently, for all $n \geqslant s+1$ there exists a CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_n) : \widetilde{S}_n \to \mathbb{R}^2$, where

$$\widetilde{Q}_n(\mathbf{y}_n; \widetilde{\alpha}_n(\mathbf{y}_n)) = \inf_{\alpha' \in \mathbb{R}} \widetilde{Q}_n(\mathbf{y}_n; \alpha') \qquad \forall\, \mathbf{y}_n \in \widetilde{S}_n,$$

$$\widetilde{\theta}_n(\mathbf{y}_n) = A_n(\widetilde{\alpha}_n(\mathbf{y}_n))^{-1} t_n(\mathbf{y}_n; \widetilde{\alpha}_n(\mathbf{y}_n)), \qquad \mathbf{y}_n \in \widetilde{S}_n, \tag{3.2.3}$$

and for all $\mathbf{y}_n \in \widetilde{S}_n$, $(\widetilde{\alpha}_n(\mathbf{y}_n), \widetilde{\theta}_n(\mathbf{y}_n))$ is a solution of the system of equations (3.2.2).

By (2.2.5) and (2.2.6), we get $\mathsf{P}\left(\lim_{n\to\infty} n^{-1} c_n(\mathbf{Y}_n) = \mathsf{E}\widetilde{X}^2\right) = 1$, where $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1). Hence $\mathbf{Y}_n \in \widetilde{S}_n$ holds asymptotically as $n \to \infty$ with probability one.

Now we turn to find sets $S_n \subset \widetilde{S}_n$, $n \geqslant s+1$ such that the system of equations (3.2.2) has a unique solution with respect to (α', θ') for all $\mathbf{y}_n \in S_n$. Let us introduce the (2×2) Hessian matrix

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

and let us denote by $\Delta_{i,n}(\mathbf{y}_n; \alpha', \theta')$ its i-th order leading principal minor, $i = 1, 2$. Further, for all $n \geqslant s+1$, let

$$S_n := \Big\{ \mathbf{y}_n \in \widetilde{S}_n : \Delta_{i,n}(\mathbf{y}_n; \alpha', \theta') > 0,\ i = 1, 2,\ \forall\, (\alpha', \theta') \in \mathbb{R}^2 \Big\}.$$

By Berkovitz [10, Theorem 3.3, Chapter III], the function $\mathbb{R}^2 \ni (\alpha', \theta') \mapsto Q_n(\mathbf{y}_n; \alpha', \theta')$ is strictly convex for all $\mathbf{y}_n \in S_n$. Since it was already proved that the system of equations (3.2.2) has a solution for all $\mathbf{y}_n \in \widetilde{S}_n$, we obtain that this solution is unique for all $\mathbf{y}_n \in S_n$.

Next we check that $\mathbf{Y}_n \in S_n$ holds asymptotically as $n \to \infty$ with probability one. For all $(\alpha', \theta') \in \mathbb{R}^2$,

$$\frac{\partial^2 Q_n}{\partial (\alpha')^2}(\mathbf{Y}_n; \alpha', \theta') = 2\sum_{k=1}^{n} Y_{k-1}^2 + 2\theta'(-Y_s + \theta') - 2Y_s\theta' = 2\left(\sum_{k=1}^{n} Y_{k-1}^2 - 2Y_s\theta' + (\theta')^2\right)$$

$$= 2\left(\sum_{k=1}^{n}{}^{(s+1)} Y_{k-1}^2 + (Y_s - \theta')^2\right) = 2\left(\sum_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s + \theta - \theta')^2\right),$$

and

$$\frac{\partial^2 Q_n}{\partial \alpha' \partial \theta'}(\mathbf{Y}_n; \alpha', \theta') = \frac{\partial^2 Q_n}{\partial \theta' \partial \alpha'}(\mathbf{Y}_n; \alpha', \theta') = 2(Y_{s-1} + Y_{s+1} - 2\alpha' Y_s - \mu_\varepsilon + 2\alpha'\theta')$$

$$= 2(X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu_\varepsilon - 2\alpha'(\theta - \theta')),$$

$$\frac{\partial^2 Q_n}{\partial (\theta')^2}(\mathbf{Y}_n; \alpha', \theta') = 2((\alpha')^2 + 1).$$

Then

$$H_n(\mathbf{Y}_n; \alpha', \theta')$$
$$= 2\begin{bmatrix} \sum_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s + \theta - \theta')^2 & X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu_\varepsilon - 2\alpha'(\theta - \theta') \\ X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu_\varepsilon - 2\alpha'(\theta - \theta') & (\alpha')^2 + 1 \end{bmatrix}$$

has leading principal minors $\Delta_{1,n}(\mathbf{Y}_n; \alpha', \theta') = 2\left(\sum_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s + \theta - \theta')^2\right)$ and

$$\Delta_{2,n}(\mathbf{Y}_n; \alpha', \theta') = \det H_n(\mathbf{Y}_n; \alpha', \theta') = 4((\alpha')^2 + 1)\left(\sum_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s + \theta - \theta')^2\right)$$
$$- 4\Big(X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu_\varepsilon - 2\alpha'(\theta - \theta')\Big)^2.$$

By (2.2.6),

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n}\Delta_{1,n}(\mathbf{Y}_n; \alpha', \theta') = 2\mathsf{E}\widetilde{X}^2, \ \forall \ (\alpha', \theta') \in \mathbb{R}^2\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n}\Delta_{2,n}(\mathbf{Y}_n; \alpha', \theta') = 4((\alpha')^2 + 1)\mathsf{E}\widetilde{X}^2, \ \forall \ (\alpha', \theta') \in \mathbb{R}^2\right) = 1,$$

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

$$\mathsf{P}\big(\lim_{n\to\infty} \Delta_{1,n}(\mathbf{Y}_n; \alpha', \theta') = \infty, \ \forall \ (\alpha', \theta') \in \mathbb{R}^2\big) = 1,$$

$$\mathsf{P}\big(\lim_{n\to\infty} \Delta_{2,n}(\mathbf{Y}_n; \alpha', \theta') = \infty, \ \forall \ (\alpha', \theta') \in \mathbb{R}^2\big) = 1,$$

which yields that $\mathbf{Y}_n \in S_n$ asymptotically as $n \to \infty$ with probability one, since we have already proved that $\mathbf{Y}_n \in \widetilde{S}_n$ asymptotically as $n \to \infty$ with probability one. □

By Lemma 3.2.1, $(\widetilde{\alpha}_n(\mathbf{Y}_n), \widetilde{\theta}_n(\mathbf{Y}_n))$ exists uniquely asymptotically as $n \to \infty$ with probability one. In the sequel we will simply denote it by $(\widetilde{\alpha}_n, \widetilde{\theta}_n)$.

The next result shows that $\widetilde{\alpha}_n$ is a strongly consistent estimator of α, whereas $\widetilde{\theta}_n$ fails to be also a strongly consistent estimator of θ.

3.2.1 Theorem. *For the CLS estimators $(\widetilde{\alpha}_n, \widetilde{\theta}_n)_{n\in\mathbb{N}}$ of the parameter $(\alpha, \theta) \in (0, 1) \times \mathbb{N}$, the sequence $(\widetilde{\alpha}_n)_{n\in\mathbb{N}}$ is strongly consistent for all $(\alpha, \theta) \in (0, 1) \times \mathbb{N}$, i.e.,*

$$\mathsf{P}(\lim_{n\to\infty} \widetilde{\alpha}_n = \alpha) = 1, \qquad \forall\, (\alpha, \theta) \in (0, 1) \times \mathbb{N}., \tag{3.2.4}$$

whereas the sequence $(\widetilde{\theta}_n)_{n\in\mathbb{N}}$ is not strongly consistent for any $(\alpha, \theta) \in (0, 1) \times \mathbb{N}$, namely,

$$\mathsf{P}\left(\lim_{n\to\infty} \widetilde{\theta}_n = Y_s - \frac{\alpha}{1+\alpha^2}(Y_{s-1} + Y_{s+1}) - \frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon\right) = 1, \qquad \forall\, (\alpha, \theta) \in (0, 1) \times \mathbb{N}. \tag{3.2.5}$$

Proof. An easy calculation shows that

$$\widetilde{\alpha}_n = \frac{\sum_{k=1}^n (Y_k - \mu_\varepsilon)Y_{k-1} - \widetilde{\theta}_n(Y_{s-1} + Y_{s+1} - \mu_\varepsilon)}{\sum_{k=1}^n Y_{k-1}^2 - 2\widetilde{\theta}_n Y_s + (\widetilde{\theta}_n)^2}, \tag{3.2.6}$$

$$\widetilde{\theta}_n = Y_s - \frac{\widetilde{\alpha}_n}{1+(\widetilde{\alpha}_n)^2}(Y_{s-1} + Y_{s+1}) - \frac{1-\widetilde{\alpha}_n}{1+(\widetilde{\alpha}_n)^2}\mu_\varepsilon, \tag{3.2.7}$$

hold asymptotically as $n \to \infty$ with probability one. Since $Y_k = X_k + \delta_{k,s}\theta$, $k \in \mathbb{Z}_+$, we get

$$\begin{aligned}
&\sum_{k=1}^n (Y_k - \mu_\varepsilon)Y_{k-1} - \widetilde{\theta}_n(Y_{s-1} + Y_{s+1} - \mu_\varepsilon) \\
&= \sum_{k=1}^{n\ (s,s+1)} (X_k - \mu_\varepsilon)X_{k-1} + (X_s + \theta - \mu_\varepsilon)X_{s-1} + (X_{s+1} - \mu_\varepsilon)(X_s + \theta) \\
&\quad - \widetilde{\theta}_n(X_{s-1} + X_{s+1} - \mu_\varepsilon),
\end{aligned}$$

and

$$\sum_{k=1}^n Y_{k-1}^2 - 2\widetilde{\theta}_n Y_s + (\widetilde{\theta}_n)^2 = \sum_{k=1}^{n\ (s+1)} X_{k-1}^2 + (X_s + \theta)^2 - 2\widetilde{\theta}_n(X_s + \theta) + (\widetilde{\theta}_n)^2,$$

hold asymptotically as $n \to \infty$ with probability one. Hence

$$\widetilde{\alpha}_n = \frac{\sum_{k=1}^n (X_k - \mu_\varepsilon)X_{k-1} + (\theta - \widetilde{\theta}_n)(X_{s-1} + X_{s+1} - \mu_\varepsilon)}{\sum_{k=1}^n X_{k-1}^2 + (\theta - \widetilde{\theta}_n)(\theta - \widetilde{\theta}_n + 2X_s)}, \tag{3.2.8}$$

holds asymptotically as $n \to \infty$ with probability one. We check that in proving (3.2.4) it is enough to verify that

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{(\theta - \widetilde{\theta}_n)(X_{s-1} + X_{s+1} - \mu_\varepsilon)}{n} = 0\right) = 1, \tag{3.2.9}$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{(\theta - \widetilde{\theta}_n)(\theta - \widetilde{\theta}_n + 2X_s)}{n} = 0\right) = 1. \tag{3.2.10}$$

Indeed, using (2.2.5), (2.2.6) and (2.2.7), we get (3.2.9) and (3.2.10) yield that

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

Now we turn to prove (3.2.9) and (3.2.10). By (3.2.7) and using again the decomposition $Y_k = X_k + \delta_{k,s}\theta,\ k\in\mathbb{Z}_+,$ we obtain

$$\widetilde{\theta}_n = X_s + \theta - \frac{\widetilde{\alpha}_n}{1+(\widetilde{\alpha}_n)^2}(X_{s-1}+X_{s+1}) - \frac{1-\widetilde{\alpha}_n}{1+(\widetilde{\alpha}_n)^2}\mu_\varepsilon,$$

and hence

$$|\widetilde{\theta}_n - \theta| \leqslant X_s + \frac{1}{2}(X_{s-1}+X_{s+1}) + \frac{3}{2}\mu_\varepsilon, \tag{3.2.11}$$

i.e., the sequences $(\widetilde{\theta}_n)_{n\in\mathbb{N}}$ and $(\widetilde{\theta}_n-\theta)_{n\in\mathbb{N}}$ are bounded with probability one. This implies (3.2.9) and (3.2.10). By (3.2.7) and (3.2.4), we get (3.2.5). □

3.2.1 Remark. We check that $\mathsf{E}\big(\lim_{n\to\infty}\widetilde{\theta}_n\big)=\theta,\ \forall\,(\alpha,\theta)\in(0,1)\times\mathbb{N},$ and

$$\operatorname{Var}\big(\lim_{n\to\infty}\widetilde{\theta}_n\big) = \frac{\mu_\varepsilon(\alpha+\alpha^3-\alpha^s-\alpha^{s+3}) + \sigma_\varepsilon^2(1+\alpha^2) + (1-\alpha)(\alpha^s+\alpha^{s+3})\mathsf{E}X_0}{(1+\alpha^2)^2}. \tag{3.2.12}$$

Note that, by (3.2.5), with probability one it holds that

$$\begin{aligned}
\lim_{n\to\infty}\widetilde{\theta}_n &= X_s + \theta - \frac{\alpha}{1+\alpha^2}(X_{s-1}+X_{s+1}) - \frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon\\
&= \theta + \frac{X_s - \alpha X_{s-1} - \mu_\varepsilon - \alpha(X_{s+1}-\alpha X_s - \mu_\varepsilon)}{1+\alpha^2} = \theta + \frac{M_s - \alpha M_{s+1}}{1+\alpha^2},
\end{aligned}$$

where $M_k := X_k - \alpha X_{k-1} - \mu_\varepsilon,\ k\in\mathbb{N}.$ Notice that $\mathsf{E}M_k = 0,\ k\in\mathbb{N},$ and $\operatorname{Cov}(M_k,M_\ell) = \delta_{k,\ell}\operatorname{Var}M_k,\ k,\ell\in\mathbb{N},$ where $\operatorname{Var}M_k = \alpha\mu_\varepsilon(1-\alpha^{k-1}) + \alpha^k(1-\alpha)\mathsf{E}X_0 + \sigma_\varepsilon^2,\ k\in\mathbb{N}.$ Indeed, by the recursion $\mathsf{E}X_\ell = \alpha\mathsf{E}X_{\ell-1}+\mu_\varepsilon,\ \ell\in\mathbb{N},$ we get $\mathsf{E}M_k=0,\ k\in\mathbb{N},$ and we get

$$\mathsf{E}X_\ell = \alpha^\ell\mathsf{E}X_0 + (1+\alpha+\cdots+\alpha^{\ell-1})\mu_\varepsilon = \alpha^\ell\mathsf{E}X_0 + \frac{1-\alpha^\ell}{1-\alpha}\mu_\varepsilon, \qquad \ell\in\mathbb{N}, \tag{3.2.13}$$

and hence

$$\begin{aligned}
\operatorname{Var}M_k &= \operatorname{Var}(X_k - \alpha X_{k-1} - \mu_\varepsilon) = \operatorname{Var}\left(\sum_{j=1}^{X_{k-1}}(\xi_{k,j}-\alpha) + (\varepsilon_k - \mu_\varepsilon)\right)\\
&= \alpha(1-\alpha)\mathsf{E}X_{k-1} + \sigma_\varepsilon^2 = \alpha(1-\alpha)\mu_\varepsilon\frac{1-\alpha^{k-1}}{1-\alpha} + \alpha^k(1-\alpha)\mathsf{E}X_0 + \sigma_\varepsilon^2\\
&= \alpha\mu_\varepsilon(1-\alpha^{k-1}) + \alpha^k(1-\alpha)\mathsf{E}X_0 + \sigma_\varepsilon^2, \qquad k\in\mathbb{N}.
\end{aligned} \tag{3.2.14}$$

Hence $\mathsf{E}(\lim_{n\to\infty}\widetilde{\theta}_n)=\theta$ and

$$\begin{aligned}
\operatorname{Var}\big(\lim_{n\to\infty}\widetilde{\theta}_n\big) = \frac{1}{(1+\alpha^2)^2}\Big(&\alpha\mu_\varepsilon(1-\alpha^{s-1}) + \alpha^s(1-\alpha)\mathsf{E}X_0 + \sigma_\varepsilon^2\\
&+ \alpha^2\big(\alpha\mu_\varepsilon(1-\alpha^s) + \alpha^{s+1}(1-\alpha)\mathsf{E}X_0 + \sigma_\varepsilon^2\big)\Big),
\end{aligned}$$

which implies (3.2.12).

We also check that $\widetilde{\theta}_n$ is an asymptotically unbiased estimator of θ as $n \to \infty$ for all $(\alpha, \theta) \in (0,1) \times \mathbb{N}$. By (3.2.5), the sequence $\widetilde{\theta}_n - \theta$, $n \in \mathbb{N}$, converges with probability one, and, by (3.2.11), the dominated convergence theorem yields that

$$\lim_{n\to\infty} \mathsf{E}(\widetilde{\theta}_n - \theta) = \mathsf{E}\big[\lim_{n\to\infty}(\widetilde{\theta}_n - \theta)\big] = 0.$$

Finally, we note that $\lim_{n\to\infty} \widetilde{\theta}_n$ can be negative with positive probability, despite the fact that $\theta \in \mathbb{N}$. $\square$

3.2.1 Definition. *Let $(\zeta_n)_{n\in\mathbb{N}}$, ζ and η be random variables on $(\Omega, \mathcal{A}, \mathsf{P})$ such that η is non-negative and integer-valued. By the expression* "conditionally on the values η, the weak convergence $\zeta_n \xrightarrow{\mathcal{L}} \zeta$ as $n \to \infty$ holds" *we mean that for all non-negative integers $m \in \mathbb{N}$ such that $\mathsf{P}(\eta = m) > 0$, we have*

$$\lim_{n\to\infty} F_{\zeta_n \,|\, \{\eta=m\}}(y) = F_{\zeta \,|\, \{\eta=m\}}(y)$$

for all $y \in \mathbb{R}$ being continuity points of $F_{\zeta \,|\, \{\eta=m\}}$, where $F_{\zeta_n \,|\, \{\eta=m\}}$ and $F_{\zeta \,|\, \{\eta=m\}}$ denote the conditional distribution function of ζ_n and ζ with respect to the event $\{\eta = m\}$, respectively.

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

$$\sqrt{n}(\widetilde{\alpha}_n - \alpha) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2_{\alpha,\varepsilon}) \qquad \text{as } n \to \infty, \tag{3.2.15}$$

where $\sigma^2_{\alpha,\varepsilon}$ is defined in (2.2.9). Furthermore, conditionally on the values Y_{s-1} and Y_{s+1},

$$\sqrt{n}\left(\widetilde{\theta}_n - \lim_{k\to\infty} \widetilde{\theta}_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, c^2_{\alpha,\varepsilon}\right) \qquad \text{as } n \to \infty, \tag{3.2.16}$$

where

$$c^2_{\alpha,\varepsilon} := \frac{\sigma^2_{\alpha,\varepsilon}}{(1+\alpha^2)^4}\big((\alpha^2 - 1)(Y_{s-1} + Y_{s+1}) + (1 + 2\alpha - \alpha^2)\mu_\varepsilon\big)^2.$$

Proof. By (3.2.8), we have $\sqrt{n}(\widetilde{\alpha}_n - \alpha) = \frac{A_n}{B_n}$ holds asymptotically as $n \to \infty$ with probability one, where

$$A_n := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - \alpha X_{k-1} - \mu_\varepsilon) X_{k-1} + \frac{1}{\sqrt{n}}(\theta - \widetilde{\theta}_n)(X_{s-1} + X_{s+1} - \mu_\varepsilon - \alpha(\theta - \widetilde{\theta}_n) - 2\alpha X_s),$$

$$B_n := \frac{1}{n} \sum_{k=1}^{n} X^2_{k-1} + \frac{1}{n}(\theta - \widetilde{\theta}_n)(\theta - \widetilde{\theta}_n + 2X_s).$$

By (2.2.8), we have

$$\frac{\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - \alpha X_{k-1} - \mu_\varepsilon) X_{k-1}}{\frac{1}{n} \sum_{k=1}^{n} X^2_{k-1}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2_{\alpha,\varepsilon}) \qquad \text{as } n \to \infty.$$

By (3.2.11),

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{\sqrt{n}}(\theta-\widetilde{\theta}_n)(X_{s-1}+X_{s+1}-\mu_\varepsilon-\alpha(\theta-\widetilde{\theta}_n)-2\alpha X_s)=0\right)=1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}(\theta-\widetilde{\theta}_n)(\theta-\widetilde{\theta}_n+2X_s)=0\right)=1.$$

Hence Slutsky's lemma yields (3.2.15). Using (3.2.5) and that

$$\widetilde{\theta}_n=Y_s+\frac{-(\widetilde{\alpha}_n-\alpha)(Y_{s-1}+Y_{s+1})+(\widetilde{\alpha}_n-\alpha)\mu_\varepsilon-\alpha(Y_{s-1}+Y_{s+1})-(1-\alpha)\mu_\varepsilon}{1+(\widetilde{\alpha}_n)^2}, \tag{3.2.17}$$

holds asymptotically as $n\to\infty$ with probability one, we get

$$\begin{aligned}
\sqrt{n}\left(\widetilde{\theta}_n-\lim_{k\to\infty}\widetilde{\theta}_k\right)&=\sqrt{n}\left(\widetilde{\theta}_n-\left(Y_s-\frac{\alpha}{1+\alpha^2}(Y_{s-1}+Y_{s+1})-\frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon\right)\right)\\
&=\sqrt{n}\Bigg(\frac{-(\widetilde{\alpha}_n-\alpha)(Y_{s-1}+Y_{s+1})+(\widetilde{\alpha}_n-\alpha)\mu_\varepsilon-\alpha(Y_{s-1}+Y_{s+1})-(1-\alpha)\mu_\varepsilon}{1+(\widetilde{\alpha}_n)^2}\\
&\qquad+\frac{\alpha}{1+\alpha^2}(Y_{s-1}+Y_{s+1})+\frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon\Bigg)\\
&=\sqrt{n}(\widetilde{\alpha}_n-\alpha)\Bigg[\frac{-Y_{s-1}-Y_{s+1}+\mu_\varepsilon}{1+(\widetilde{\alpha}_n)^2}+\alpha(Y_{s-1}+Y_{s+1})\frac{\alpha+\widetilde{\alpha}_n}{(1+(\widetilde{\alpha}_n)^2)(1+\alpha^2)}\\
&\qquad+\mu_\varepsilon\frac{(1-\alpha)(\alpha+\widetilde{\alpha}_n)}{(1+(\widetilde{\alpha}_n)^2)(1+\alpha^2)}\Bigg].
\end{aligned}$$

Using (3.2.15) and (3.2.4), Slutsky's lemma yields (3.2.16) with

$$\begin{aligned}
c^2_{\alpha,\varepsilon}&=\sigma^2_{\alpha,\varepsilon}\left(\frac{-Y_{s-1}-Y_{s+1}+\mu_\varepsilon}{1+\alpha^2}+\frac{2\alpha^2}{(1+\alpha^2)^2}(Y_{s-1}+Y_{s+1})+\frac{2\alpha(1-\alpha)}{(1+\alpha^2)^2}\mu_\varepsilon\right)^2\\
&=\frac{\sigma^2_{\alpha,\varepsilon}}{(1+\alpha^2)^2}\left(\frac{\alpha^2-1}{1+\alpha^2}(Y_{s-1}+Y_{s+1})+\frac{1+2\alpha-\alpha^2}{1+\alpha^2}\mu_\varepsilon\right)^2.
\end{aligned}$$

□

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

We assume $I=1$ and that the relevant time point $s\in\mathbb{N}$ is known. We concentrate on the CLS estimation of α, μ_ε and $\theta:=\theta_1$. For all $n\geqslant s+1$, $n\in\mathbb{N}$, we define the function $Q_n:\mathbb{R}^{n+1}\times\mathbb{R}^3\to\mathbb{R}$, as

$$\begin{aligned}
Q_n(\mathbf{y}_n;\alpha',\mu'_\varepsilon,\theta')&:=\sum_{k=1}^{n}{}^{(s,s+1)}\big(y_k-\alpha' y_{k-1}-\mu'_\varepsilon\big)^2+\big(y_s-\alpha' y_{s-1}-\mu'_\varepsilon-\theta'\big)^2\\
&\qquad+\big(y_{s+1}-\alpha' y_s-\mu'_\varepsilon+\alpha'\theta'\big)^2,\qquad \mathbf{y}_n\in\mathbb{R}^{n+1},\ \alpha',\mu'_\varepsilon,\theta'\in\mathbb{R}.
\end{aligned}$$

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

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

where S_n is suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 3.3.1). We note that we do not define the CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n)$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$. We get for all $(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') \in \mathbb{R}^{n+1} \times \mathbb{R}^3$,

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

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

and

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

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

3.3.1 Lemma. *There exist subsets $S_n \subset \mathbb{R}^{n+1}$, $n \geqslant \max(3, s+1)$ with the following properties:*

(i) *there exists a unique CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n) : S_n \to \mathbb{R}^3$,*

(ii) *for all $\mathbf{y}_n \in S_n$, $(\widehat{\alpha}_n(\mathbf{y}_n), \widehat{\mu}_{\varepsilon,n}(\mathbf{y}_n), \widehat{\theta}_n(\mathbf{y}_n))$ is the unique solution of the system of equations*

$$\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = 0, \quad \frac{\partial Q_n}{\partial \mu'_\varepsilon}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = 0, \quad \frac{\partial Q_n}{\partial \theta'}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = 0, \tag{3.3.1}$$

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

Proof. For any fixed $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $n \geqslant \max(3, s+1)$ and $\alpha' \in \mathbb{R}$, the quadratic function $\mathbb{R}^2 \ni (\mu'_\varepsilon, \theta') \mapsto Q_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta')$ can be written in the form

$$Q_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta')$$

$$= \left(\begin{bmatrix} \mu'_\varepsilon \\ \theta' \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right)^\top A_n(\alpha') \left(\begin{bmatrix} \mu'_\varepsilon \\ \theta' \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right) + \widehat{Q}_n(\mathbf{y}_n; \alpha'),$$

where

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

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

and the matrix

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

is strictly positive definite for all $n \geqslant 3$ and $\alpha' \in \mathbb{R}$. Indeed, the leading principal minors of $A_n(\alpha')$ take the following forms: n,

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

and for all $n \geqslant 3$, the discriminant $4 - 4(n-1)^2$ of the equation $(n-1)x^2 + 2x + n - 1 = 0$ is negative.

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

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

The polynomial $\mathbb{R} \ni \alpha' \mapsto D_n(\alpha')$ is of order 2 with leading coefficient $n - 1$. 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 4 with leading coefficient

$$c_n(\mathbf{y}_n) := (n-1) \sum_{k=1}^n y_{k-1}^2 - \left(\sum_{k=1}^n y_{k-1} \right)^2 + 2 \left(\sum_{k=1}^n y_{k-1} \right) y_s - n y_s^2.$$

Let

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

For $\mathbf{y}_n \in \widehat{S}_n$, we have $\lim_{|\alpha'|\to\infty} \widehat{Q}_n(\mathbf{y}_n;\alpha') = \infty$ and the continuous function $\mathbb{R} \ni \alpha' \mapsto \widehat{Q}_n(\mathbf{y}_n;\alpha')$ attains its infimum. Consequently, for all $n \geqslant \max(3, s+1)$ there exists a CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n) : \widehat{S}_n \to \mathbb{R}^3$, where

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

$$\begin{bmatrix} \widehat{\mu}_{\varepsilon,n}(\mathbf{y}_n) \\ \widehat{\theta}_n(\mathbf{y}_n) \end{bmatrix} = A_n(\widehat{\alpha}_n(\mathbf{y}_n))^{-1} t_n(\mathbf{y}_n;\widehat{\alpha}_n(\mathbf{y}_n)), \qquad \mathbf{y}_n \in \widehat{S}_n, \tag{3.3.2}$$

and for all $\mathbf{y}_n \in \widehat{S}_n$, $(\widehat{\alpha}_n(\mathbf{y}_n), \widehat{\mu}_{\varepsilon,n}(\mathbf{y}_n), \widehat{\theta}_n(\mathbf{y}_n))$ is a solution of the system of equations (3.3.1).

By (2.2.5) and (2.2.6), we get $\mathsf{P}\left(\lim_{n\to\infty} n^{-2} c_n(\mathbf{Y}_n) = \operatorname{Var} \widetilde{X}\right) = 1$, where $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1). Hence $\mathbf{Y}_n \in \widehat{S}_n$ holds asymptotically as $n \to \infty$ with probability one.

Now we turn to find sets $S_n \subset \widehat{S}_n$, $n \geqslant \max(3, s+1)$ such that the system of equations (3.3.1) has a unique solution with respect to $(\alpha', \mu'_\varepsilon, \theta')$ for all $\mathbf{y}_n \in S_n$. Let us introduce the (3×3) Hessian matrix

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

and let us denote by $\Delta_{i,n}(\mathbf{y}_n;\alpha',\mu'_\varepsilon,\theta')$ its i-th order leading principal minor, $i = 1, 2, 3$. Further, for all $n \geqslant \max(3, s+1)$, let

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

By Berkovitz [10, Theorem 3.3, Chapter III], the function $\mathbb{R}^3 \ni (\alpha',\mu'_\varepsilon,\theta') \mapsto Q_n(\mathbf{y}_n;\alpha',\mu'_\varepsilon,\theta')$ is strictly convex for all $\mathbf{y}_n \in S_n$. Since it was already proved that the system of equations (3.3.1) has a solution for all $\mathbf{y}_n \in \widehat{S}_n$, we obtain that this solution is unique for all $\mathbf{y}_n \in S_n$.

Next we check that $\mathbf{Y}_n \in S_n$ holds asymptotically as $n \to \infty$ with probability one. Using also the proof of Lemma 3.2.1, for all $(\alpha',\mu'_\varepsilon,\theta') \in \mathbb{R}^3$, we get

$$\begin{aligned} \frac{\partial^2 Q_n}{\partial(\alpha')^2}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') &= 2\sum_{k=1}^n Y_{k-1}^2 + 2\theta'(-Y_s + \theta') - 2Y_s\theta' \\ &= 2\left(\sum_{k=1}^{n\ (s+1)} X_{k-1}^2 + (X_s + \theta - \theta')^2 \right), \\ \frac{\partial^2 Q_n}{\partial\alpha'\partial\theta'}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = \frac{\partial^2 Q_n}{\partial\theta'\partial\alpha'}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') &= 2(Y_{s-1} + Y_{s+1} - 2\alpha' Y_s - \mu'_\varepsilon + 2\alpha'\theta') \\ &= 2(X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu'_\varepsilon - 2\alpha'(\theta - \theta')), \end{aligned}$$

and

$$\frac{\partial^2 Q_n}{\partial\alpha'\partial\mu'_\varepsilon}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = \frac{\partial^2 Q_n}{\partial\mu'_\varepsilon\partial\alpha'}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = 2\sum_{k=1}^{n} Y_{k-1} - 2\theta' = 2\left(\sum_{k=1}^{n} X_{k-1} + \theta - \theta'\right),$$

$$\frac{\partial^2 Q_n}{\partial\theta'\partial\mu'_\varepsilon}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = \frac{\partial^2 Q_n}{\partial\mu'_\varepsilon\partial\theta'}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = 2(1-\alpha'),$$

$$\frac{\partial^2 Q_n}{\partial(\theta')^2}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = 2((\alpha')^2+1), \qquad \frac{\partial^2 Q_n}{\partial(\mu'_\varepsilon)^2}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = 2n.$$

Then

$$H_n(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = 2\begin{bmatrix} \sum\limits_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s+\theta-\theta')^2 & \sum_{k=1}^n X_{k-1}+\theta-\theta' & a \\ \sum_{k=1}^n X_{k-1}+\theta-\theta' & n & 1-\alpha' \\ a & 1-\alpha' & (\alpha')^2+1 \end{bmatrix},$$

where $a := X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu'_\varepsilon - 2\alpha'(\theta-\theta')$. Then $H_n(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta')$ has leading principal minors $\Delta_{1,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') := 2\left(\sum\limits_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s+\theta-\theta')^2\right)$,

$$\Delta_{2,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') := 4\left(n\left(\sum_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s+\theta-\theta')^2\right) - \left(\sum_{k=1}^{n} X_{k-1}+\theta-\theta'\right)^2\right),$$

and

$$\begin{aligned}\Delta_{3,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') &= \det H_n(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') \\ &= 8\Bigg[n\left(((\alpha')^2+1)\left(\sum_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s+\theta-\theta')^2\right) - a^2\right) \\ &\quad - ((\alpha')^2+1)\left(\sum_{k=1}^{n} X_{k-1}+\theta-\theta'\right)^2 + 2(1-\alpha')a\left(\sum_{k=1}^{n} X_{k-1}+\theta-\theta'\right) \\ &\quad - (1-\alpha')^2\left(\sum_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s+\theta-\theta')^2\right)\Bigg].\end{aligned}$$

By (2.2.5) and (2.2.6), we have

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{2n}\Delta_{1,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = \mathsf{E}\widetilde{X}^2,\ \ \forall\ (\alpha',\mu'_\varepsilon,\theta')\in\mathbb{R}^3\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{4n^2}\Delta_{2,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta') = \operatorname{Var}\widetilde{X},\ \ \forall\ (\alpha',\mu'_\varepsilon,\theta')\in\mathbb{R}^3\right) = 1,$$

and

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{8n^2}\Delta_{3,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta') = ((\alpha')^2+1)\operatorname{Var}\widetilde{X},\ \ \forall\ (\alpha',\mu_\varepsilon',\theta')\in\mathbb{R}^3\right) = 1,$$

for all $(\alpha',\mu_\varepsilon',\theta')\in\mathbb{R}^3$, where $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1). Hence

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

which yields that $\mathbf{Y}_n\in S_n$ asymptotically as $n\to\infty$ with probability one, since we have already proved that $\mathbf{Y}_n\in\widehat{S}_n$ asymptotically as $n\to\infty$ with probability one. □

By Lemma 3.3.1, $(\widehat{\alpha}_n(\mathbf{Y}_n),\widehat{\mu}_{\varepsilon,n}(\mathbf{Y}_n),\widehat{\theta}_n(\mathbf{Y}_n))$ exists uniquely asymptotically as $n\to\infty$ with probability one. In the sequel we will simply denote it by $(\widehat{\alpha}_n,\widehat{\mu}_{\varepsilon,n},\widehat{\theta}_n)$.

The next result shows that $\widehat{\alpha}_n$ and $\widehat{\mu}_{\varepsilon,n}$ are strongly consistent estimators of α and μ_ε, respectively, whereas $\widehat{\theta}_n$ fails to be a strongly consistent estimator of θ.

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

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

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

whereas the sequence $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ is not strongly consistent for any $(\alpha,\mu_\varepsilon,\theta)\in(0,1)\times(0,\infty)\times\mathbb{N}$, namely,

(3.3.5) $$\mathsf{P}\left(\lim_{n\to\infty}\widehat{\theta}_n = Y_s-\frac{\alpha}{1+\alpha^2}(Y_{s-1}+Y_{s+1})-\frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon\right)=1,$$

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

Proof. An easy calculation shows that

$$\left(\sum_{k=1}^n Y_{k-1}^2+(Y_s-\widehat{\theta}_n)^2-Y_s^2\right)\widehat{\alpha}_n+\left(\sum_{k=1}^n Y_{k-1}-\widehat{\theta}_n\right)\widehat{\mu}_{\varepsilon,n}=\sum_{k=1}^n Y_{k-1}Y_k-\widehat{\theta}_n(Y_{s-1}+Y_{s+1})$$

$$\left(\sum_{k=1}^n Y_{k-1}-\widehat{\theta}_n\right)\widehat{\alpha}_n+n\widehat{\mu}_{\varepsilon,n}=\sum_{k=1}^n Y_k-\widehat{\theta}_n,$$

or equivalently

(3.3.6) $$\begin{bmatrix}\sum_{k=1}^n Y_{k-1}^2+(Y_s-\widehat{\theta}_n)^2-Y_s^2 & \sum_{k=1}^n Y_{k-1}-\widehat{\theta}_n\\ \sum_{k=1}^n Y_{k-1}-\widehat{\theta}_n & n\end{bmatrix}\begin{bmatrix}\widehat{\alpha}_n\\ \widehat{\mu}_{\varepsilon,n}\end{bmatrix} = \begin{bmatrix}\sum_{k=1}^n Y_{k-1}Y_k-\widehat{\theta}_n(Y_{s-1}+Y_{s+1})\\ \sum_{k=1}^n Y_k-\widehat{\theta}_n\end{bmatrix}$$

holds asymptotically as $n \to \infty$ with probability one. Let us introduce the notation

$$\begin{aligned} E_n &:= n\left(\sum_{k=1}^{n} Y_{k-1}^2 + (Y_s - \widehat{\theta}_n)^2 - Y_s^2\right) - \left(\sum_{k=1}^{n} Y_{k-1} - \widehat{\theta}_n\right)^2 \\ &= n\left(\sum_{k=1}^{n}{}^{(s+1)} X_{k-1}^2 + (X_s + \theta - \widehat{\theta}_n)^2\right) - \left(\sum_{k=1}^{n} X_{k-1} + \theta - \widehat{\theta}_n\right)^2, \qquad n \geqslant s+1, \ n \in \mathbb{N}. \end{aligned}$$

By (2.2.5) and (2.2.6),

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{E_n}{n^2} = \mathsf{E}\widetilde{X}^2 - (\mathsf{E}\widetilde{X})^2 = \operatorname{Var}\widetilde{X} > 0\right) = 1, \tag{3.3.7}$$

which yields that $\mathsf{P}(\lim_{n\to\infty} E_n = \infty) = 1$. Hence asymptotically as $n \to \infty$ with probability one we get

$$\begin{aligned} \begin{bmatrix} \widehat{\alpha}_n \\ \widehat{\mu}_{\varepsilon,n} \end{bmatrix} &= \frac{1}{E_n} \begin{bmatrix} n & -\sum_{k=1}^n Y_{k-1} + \widehat{\theta}_n \\ -\sum_{k=1}^n Y_{k-1} + \widehat{\theta}_n & \sum_{k=1}^n Y_{k-1}^2 + (Y_s - \widehat{\theta}_n)^2 - Y_s^2 \end{bmatrix} \\ &\qquad \times \begin{bmatrix} \sum_{k=1}^n Y_{k-1}Y_k - \widehat{\theta}_n(Y_{s-1} + Y_{s+1}) \\ \sum_{k=1}^n Y_k - \widehat{\theta}_n \end{bmatrix} \\ &= \frac{1}{E_n} \begin{bmatrix} n & -\sum_{k=1}^n X_{k-1} + (\widehat{\theta}_n - \theta) \\ -\sum_{k=1}^n X_{k-1} + (\widehat{\theta}_n - \theta) & \sum_{k=1}^n X_{k-1}^2 + (X_s + \theta - \widehat{\theta}_n)^2 - X_s^2 \end{bmatrix} \\ &\qquad \times \begin{bmatrix} \sum_{k=1}^n X_{k-1}X_k + (\theta - \widehat{\theta}_n)(X_{s-1} + X_{s+1}) \\ \sum_{k=1}^n X_k - (\widehat{\theta}_n - \theta) \end{bmatrix} \\ &=: \frac{1}{E_n} \begin{bmatrix} V_n^{(1)} \\ V_n^{(2)} \end{bmatrix}, \end{aligned} \tag{3.3.8}$$

where

$$\begin{aligned} V_n^{(1)} :=& n\sum_{k=1}^n X_{k-1}X_k - \left(\sum_{k=1}^n X_{k-1}\right)\left(\sum_{k=1}^n X_k\right) + n(\theta - \widehat{\theta}_n)(X_{s-1} + X_{s+1}) + (\widehat{\theta}_n - \theta)\sum_{k=1}^n X_{k-1} \\ &+ (\widehat{\theta}_n - \theta)\sum_{k=1}^n X_k - (\widehat{\theta}_n - \theta)^2, \end{aligned}$$

and

$$\begin{aligned} V_n^{(2)} :=& \left(\sum_{k=1}^n X_{k-1}^2\right)\left(\sum_{k=1}^n X_k\right) - \left(\sum_{k=1}^n X_{k-1}\right)\left(\sum_{k=1}^n X_{k-1}X_k\right) - (\theta - \widehat{\theta}_n)(X_{s-1} + X_{s+1})\sum_{k=1}^n X_{k-1} \\ &+ (\widehat{\theta}_n - \theta)\sum_{k=1}^n X_{k-1}X_k - (\widehat{\theta}_n - \theta)^2(X_{s-1} + X_{s+1}) - (\widehat{\theta}_n - \theta)\sum_{k=1}^n X_{k-1}^2 \\ &+ (X_s + \theta - \widehat{\theta}_n)^2\sum_{k=1}^n X_k - (\widehat{\theta}_n - \theta)(X_s + \theta - \widehat{\theta}_n)^2 - X_s^2\sum_{k=1}^n X_k + X_s^2(\widehat{\theta}_n - \theta). \end{aligned}$$

Similarly, an easy calculation shows that

$$n\widehat{\mu}_{\varepsilon,n} + (1-\widehat{\alpha}_n)\widehat{\theta}_n = \sum_{k=1}^n Y_k - \widehat{\alpha}_n \sum_{k=1}^n Y_{k-1},$$

$$(1-\widehat{\alpha}_n)\widehat{\mu}_{\varepsilon,n} + (1+(\widehat{\alpha}_n)^2)\widehat{\theta}_n = (1+(\widehat{\alpha}_n)^2)Y_s - \widehat{\alpha}_n(Y_{s-1}+Y_{s+1}),$$

or equivalently

$$\begin{bmatrix} n & 1-\widehat{\alpha}_n \\ 1-\widehat{\alpha}_n & 1+(\widehat{\alpha}_n)^2 \end{bmatrix} \begin{bmatrix} \widehat{\mu}_{\varepsilon,n} \\ \widehat{\theta}_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n Y_k - \widehat{\alpha}_n \sum_{k=1}^n Y_{k-1}) \\ (1+(\widehat{\alpha}_n)^2)Y_s - \widehat{\alpha}_n(Y_{s-1}+Y_{s+1}) \end{bmatrix}$$

holds asymptotically as $n\to\infty$ with probability one. Recalling that $D_n(\widehat{\alpha}_n) = n(1+(\widehat{\alpha}_n)^2) - (1-\widehat{\alpha}_n)^2$, we have asymptotically as $n\to\infty$ with probability one,

$$\begin{aligned}
\begin{bmatrix} \widehat{\mu}_{\varepsilon,n} \\ \widehat{\theta}_n \end{bmatrix} &= \frac{1}{D_n(\widehat{\alpha}_n)} \begin{bmatrix} 1+(\widehat{\alpha}_n)^2 & -(1-\widehat{\alpha}_n) \\ -(1-\widehat{\alpha}_n) & n \end{bmatrix} \begin{bmatrix} \sum_{k=1}^n Y_k - \widehat{\alpha}_n \sum_{k=1}^n Y_{k-1} \\ (1+(\widehat{\alpha}_n)^2)Y_s - \widehat{\alpha}_n(Y_{s-1}+Y_{s+1}) \end{bmatrix} \\
&= \begin{bmatrix} \frac{(1+(\widehat{\alpha}_n)^2)\big(\sum_{k=1}^n Y_k - \widehat{\alpha}_n \sum_{k=1}^n Y_{k-1}\big) - (1-\widehat{\alpha}_n)\big((1+(\widehat{\alpha}_n)^2)Y_s - \widehat{\alpha}_n(Y_{s-1}+Y_{s+1})\big)}{D_n(\widehat{\alpha}_n)} \\ \frac{-(1-\widehat{\alpha}_n)\big(\sum_{k=1}^n Y_k - \widehat{\alpha}_n \sum_{k=1}^n Y_{k-1}\big) + n\big((1+(\widehat{\alpha}_n)^2)Y_s - \widehat{\alpha}_n(Y_{s-1}+Y_{s+1})\big)}{D_n(\widehat{\alpha}_n)} \end{bmatrix}.
\end{aligned}$$

We show that the sequence $(\widehat{\theta}_n - \theta)_{n\in\mathbb{N}}$ is bounded with probability one. Using the decomposition $Y_k = X_k + \delta_{k,s}\theta$, $k\in\mathbb{Z}_+$, we get

$$\begin{bmatrix} \widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon \\ \widehat{\theta}_n - \theta \end{bmatrix} = \frac{1}{D_n(\widehat{\alpha}_n)} \begin{bmatrix} V_n^{(3)} \\ V_n^{(4)} \end{bmatrix}, \tag{3.3.9}$$

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

$$\begin{aligned}
V_n^{(3)} &:= (1+(\widehat{\alpha}_n)^2)\left(\sum_{k=1}^n X_k - \widehat{\alpha}_n \sum_{k=1}^n X_{k-1} + (1-\widehat{\alpha}_n)\theta\right) \\
&\quad - (1-\widehat{\alpha}_n)\Big((1+(\widehat{\alpha}_n)^2)X_s - \widehat{\alpha}_n(X_{s-1}+X_{s+1}) + (1+(\widehat{\alpha}_n)^2)\theta\Big) \\
&\quad - n(1+(\widehat{\alpha}_n)^2)\mu_\varepsilon + (1-\widehat{\alpha}_n)^2\mu_\varepsilon \\
&= (1+(\widehat{\alpha}_n)^2)\left(\sum_{k=1}^n X_k - \widehat{\alpha}_n \sum_{k=1}^n X_{k-1} - n\mu_\varepsilon\right) \\
&\quad - (1-\widehat{\alpha}_n)\Big((1+(\widehat{\alpha}_n)^2)X_s - \widehat{\alpha}_n(X_{s-1}+X_{s+1}) - (1-\widehat{\alpha}_n)\mu_\varepsilon\Big),
\end{aligned}$$

and

$$V_n^{(4)} := -(1-\widehat{\alpha}_n)\left(\sum_{k=1}^n X_k - \widehat{\alpha}_n \sum_{k=1}^n X_{k-1}\right) + n\Big((1+(\widehat{\alpha}_n)^2)X_s - \widehat{\alpha}_n(X_{s-1}+X_{s+1})\Big).$$

By (3.3.9), we have asymptotically as $n \to \infty$ with probability one,

$$|\widehat{\theta}_n - \theta| \leqslant \frac{(1+(\widehat{\alpha}_n)^2)n}{(1+(\widehat{\alpha}_n)^2)n-(1-\widehat{\alpha}_n)^2}\left[\frac{|1-\widehat{\alpha}_n|}{1+(\widehat{\alpha}_n)^2}\frac{\sum_{k=1}^n X_k}{n} + \frac{|\widehat{\alpha}_n(1-\widehat{\alpha}_n)|}{1+(\widehat{\alpha}_n)^2}\frac{\sum_{k=1}^n X_{k-1}}{n} + X_s + \frac{|\widehat{\alpha}_n|}{1+(\widehat{\alpha}_n)^2}(X_{s-1}+X_{s+1})\right]$$

$$\leqslant \frac{1}{1-\frac{(1-\widehat{\alpha}_n)^2}{n(1+(\widehat{\alpha}_n)^2)}}\left[\frac{3}{2}\frac{\sum_{k=1}^n X_k}{n} + \frac{3}{2}\frac{\sum_{k=1}^n X_{k-1}}{n} + X_s + X_{s-1} + X_{s+1}\right],$$

$$\leqslant \frac{1}{1-\frac{(1-\widehat{\alpha}_n)^2}{n(1+(\widehat{\alpha}_n)^2)}}\left[3\frac{\sum_{k=0}^n X_k}{n} + X_s + X_{s-1} + X_{s+1}\right],$$

Using (2.2.5) and that

$$\frac{(1-\widehat{\alpha}_n)^2}{1+(\widehat{\alpha}_n)^2} < 3, \qquad n \in \mathbb{N},$$

we have the sequences $(\widehat{\theta}_n - \theta)_{n\in\mathbb{N}}$ and $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ are bounded with probability one.

Similarly to (3.2.7), one can check that

$$\widehat{\theta}_n = Y_s - \frac{\widehat{\alpha}_n}{1+(\widehat{\alpha}_n)^2}(Y_{s-1}+Y_{s+1}) - \frac{1-\widehat{\alpha}_n}{1+(\widehat{\alpha}_n)^2}\widehat{\mu}_{\varepsilon,n} \tag{3.3.10}$$

holds asymptotically as $n \to \infty$ with probability one.

Using (3.3.7) and (3.3.8), to prove (3.3.3) and (3.3.4), it is enough to check that

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{V_n^{(1)}}{n^2} = \alpha \operatorname{Var}\widetilde{X}\right) = 1 \qquad \text{and} \qquad \mathsf{P}\left(\lim_{n\to\infty}\frac{V_n^{(2)}}{n^2} = \mu_\varepsilon \operatorname{Var}\widetilde{X}\right) = 1,$$

for all $(\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}$. Using that the sequence $(\widehat{\theta}_n - \theta)_{n\in\mathbb{N}}$ is bounded with probability one, by (2.2.5), (2.2.6) and (2.2.7), we get with probability one

$$\lim_{n\to\infty}\frac{V_n^{(1)}}{n^2} = \alpha\mathsf{E}\widetilde{X}^2 + \mu_\varepsilon\mathsf{E}\widetilde{X} - (\mathsf{E}\widetilde{X})^2 = \alpha\operatorname{Var}\widetilde{X} + \mu_\varepsilon\mathsf{E}\widetilde{X} + (\alpha-1)(\mathsf{E}\widetilde{X})^2 = \alpha\operatorname{Var}\widetilde{X},$$

$$\lim_{n\to\infty}\frac{V_n^{(2)}}{n^2} = \mathsf{E}\widetilde{X}^2\mathsf{E}\widetilde{X} - \mathsf{E}\widetilde{X}(\alpha\mathsf{E}\widetilde{X}^2 + \mu_\varepsilon\mathsf{E}\widetilde{X}) = \mu_\varepsilon\operatorname{Var}\widetilde{X} + ((1-\alpha)\mathsf{E}\widetilde{X} - \mu_\varepsilon)\mathsf{E}\widetilde{X}^2 = \mu_\varepsilon\operatorname{Var}\widetilde{X},$$

where the last equality follows by (2.2.3).

Finally, (3.3.5) follows from (3.3.10), (3.3.3) and (3.3.4). □

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

$$\begin{bmatrix}\sqrt{n}(\widehat{\alpha}_n - \alpha)\\ \sqrt{n}(\widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon)\end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, B_{\alpha,\varepsilon}\right) \qquad \text{as } n\to\infty, \tag{3.3.11}$$

where the (2×2)*-matrix* $B_{\alpha,\varepsilon}$ *is defined in* (2.3.2). *Moreover, conditionally on the values* Y_{s-1} *and* Y_{s+1},

$$\sqrt{n}\left(\widehat{\theta}_n - \lim_{k\to\infty}\widehat{\theta}_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, d_{\alpha,\varepsilon}^\top B_{\alpha,\varepsilon} d_{\alpha,\varepsilon}\right) \qquad \text{as } n\to\infty, \tag{3.3.12}$$

where

$$d_{\alpha,\varepsilon} := \frac{1}{(1+\alpha^2)^2}\begin{bmatrix}(\alpha^2-1)(Y_{s-1}+Y_{s+1}) + (2\alpha+1-\alpha^2)\mu_\varepsilon \\ -(1+\alpha^2)(1-\alpha)\end{bmatrix}.$$

Proof. By (3.3.6), with the notation

$$B_n := \begin{bmatrix}\sum_{k=1}^n Y_{k-1}^2 + (Y_s-\widehat{\theta}_n)^2 - Y_s^2 & \sum_{k=1}^n Y_{k-1} - \widehat{\theta}_n \\ \sum_{k=1}^n Y_{k-1} - \widehat{\theta}_n & n\end{bmatrix}, \qquad n\in\mathbb{N},$$

we get

$$\begin{bmatrix}\widehat{\alpha}_n \\ \widehat{\mu}_{\varepsilon,n}\end{bmatrix} = B_n^{-1}\begin{bmatrix}\sum_{k=1}^n Y_{k-1}Y_k - \widehat{\theta}_n(Y_{s-1}+Y_{s+1}) \\ \sum_{k=1}^n Y_k - \widehat{\theta}_n\end{bmatrix},$$

holds asymptotically as $n\to\infty$ with probability one. Hence

$$\begin{aligned}
\begin{bmatrix}\widehat{\alpha}_n - \alpha \\ \widehat{\mu}_{\varepsilon,n}-\mu_\varepsilon\end{bmatrix} &= B_n^{-1}\left(\begin{bmatrix}\sum_{k=1}^n Y_{k-1}Y_k - \widehat{\theta}_n(Y_{s-1}+Y_{s+1}) \\ \sum_{k=1}^n Y_k - \widehat{\theta}_n\end{bmatrix} - B_n\begin{bmatrix}\alpha \\ \mu_\varepsilon\end{bmatrix}\right) \\
&= B_n^{-1}\left(\begin{bmatrix}\sum_{k=1}^n Y_{k-1}Y_k - \widehat{\theta}_n(Y_{s-1}+Y_{s+1}) \\ \sum_{k=1}^n Y_k - \widehat{\theta}_n\end{bmatrix}\right. \\
&\qquad\left. - \begin{bmatrix}\alpha\sum_{k=1}^n Y_{k-1}^2 + \mu_\varepsilon\sum_{k=1}^n Y_{k-1} + \alpha(Y_s-\widehat{\theta}_n)^2 - \alpha Y_s^2 - \mu_\varepsilon\widehat{\theta}_n \\ \alpha\sum_{k=1}^n Y_{k-1} + n\mu_\varepsilon - \alpha\widehat{\theta}_n\end{bmatrix}\right).
\end{aligned}$$

Then

$$\begin{bmatrix}\widehat{\alpha}_n - \alpha \\ \widehat{\mu}_{\varepsilon,n}-\mu_\varepsilon\end{bmatrix} = B_n^{-1}\begin{bmatrix}V_n^{(5)} \\ V_n^{(6)}\end{bmatrix}$$

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

$$V_n^{(5)} := \sum_{k=1}^n (Y_k - \alpha Y_{k-1} - \mu_\varepsilon)Y_{k-1} - \widehat{\theta}_n(Y_{s-1}+Y_{s+1}) + 2\alpha Y_s\widehat{\theta}_n - \alpha(\widehat{\theta}_n)^2 + \mu_\varepsilon\widehat{\theta}_n,$$

$$V_n^{(6)} := \sum_{k=1}^n (Y_k - \alpha Y_{k-1} - \mu_\varepsilon) - (1-\alpha)\widehat{\theta}_n.$$

To prove (3.3.11), it is enough to show that

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{B_n}{n} = \begin{bmatrix}\mathsf{E}\widetilde{X}^2 & \mathsf{E}\widetilde{X} \\ \mathsf{E}\widetilde{X} & 1\end{bmatrix}\right) = 1, \tag{3.3.13}$$

$$\frac{1}{\sqrt{n}}\begin{bmatrix}V_n^{(5)} \\ V_n^{(6)}\end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, A_{\alpha,\varepsilon}\right) \qquad \text{as } n\to\infty, \tag{3.3.14}$$

where $\widetilde{X}$ is a random variable having the unique stationary distribution of the INAR(1) model in (2.1.1) and the (2×2)-matrix $A_{\alpha,\varepsilon}$ is defined in (2.3.3). Using (2.2.5), (2.2.6) and that the sequence $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ is bounded with probability one, we get (3.3.13). Now we turn to prove (3.3.14). An easy calculation shows that

$$\begin{aligned}
V_n^{(5)} &= \sum_{k=1}^{n}{}^{(s,s+1)}(X_k-\alpha X_{k-1}-\mu_\varepsilon)X_{k-1}+(X_s-\alpha X_{s-1}-\mu_\varepsilon+\theta)X_{s-1}\\
&\quad+(X_{s+1}-\alpha X_s-\mu_\varepsilon-\alpha\theta)(X_s+\theta)-\widehat{\theta}_n(X_{s-1}+X_{s+1})+2\alpha(X_s+\theta)\widehat{\theta}_n-\alpha(\widehat{\theta}_n)^2+\mu_\varepsilon\widehat{\theta}_n\\
&=\sum_{k=1}^{n}(X_k-\alpha X_{k-1}-\mu_\varepsilon)X_{k-1}+\theta(X_{s-1}+X_{s+1})-2\alpha\theta X_s-\alpha\theta^2-\theta\mu_\varepsilon\\
&\quad-\widehat{\theta}_n(X_{s-1}+X_{s+1})+2\alpha(X_s+\theta)\widehat{\theta}_n-\alpha(\widehat{\theta}_n)^2+\mu_\varepsilon\widehat{\theta}_n\\
&=\sum_{k=1}^{n}(X_k-\alpha X_{k-1}-\mu_\varepsilon)X_{k-1}+(\theta-\widehat{\theta}_n)(X_{s-1}+X_{s+1}-2\alpha X_s-\mu_\varepsilon-\alpha(\theta-\widehat{\theta}_n)),
\end{aligned}$$

and

$$V_n^{(6)}=\sum_{k=1}^{n}(X_k-\alpha X_{k-1}-\mu_\varepsilon)+(1-\alpha)(\theta-\widehat{\theta}_n).$$

By formula (6.43) in Hall and Heyde [35, Section 6.3],

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

Using that the sequence $(\widehat{\theta}_n-\theta)_{n\in\mathbb{N}}$ is bounded with probability one, by Slutsky's lemma, we get (3.3.14).

Now we turn to prove (3.3.12). Using (3.3.5) and (3.3.10), we have

$$\begin{aligned}
\sqrt{n}(\widehat{\theta}_n-\lim_{k\to\infty}\widehat{\theta}_k)&=\sqrt{n}\left(\widehat{\theta}_n-\left(Y_s-\frac{\alpha}{1+\alpha^2}(Y_{s-1}+Y_{s+1})-\frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon\right)\right)\\
&=\sqrt{n}\left(\left(\frac{\alpha}{1+\alpha^2}-\frac{\widehat{\alpha}_n}{1+(\widehat{\alpha}_n)^2}\right)(Y_{s-1}+Y_{s+1})+\frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon-\frac{1-\widehat{\alpha}_n}{1+(\widehat{\alpha}_n)^2}\widehat{\mu}_{\varepsilon,n}\right)\\
&=\sqrt{n}\left(\frac{(\widehat{\alpha}_n-\alpha)(\alpha\widehat{\alpha}_n-1)}{(1+\alpha^2)(1+(\widehat{\alpha}_n)^2)}(Y_{s-1}+Y_{s+1})+\frac{1-\alpha}{1+\alpha^2}(\mu_\varepsilon-\widehat{\mu}_{\varepsilon,n})+\left(\frac{1-\alpha}{1+\alpha^2}-\frac{1-\widehat{\alpha}_n}{1+(\widehat{\alpha}_n)^2}\right)\widehat{\mu}_{\varepsilon,n}\right)\\
&=\sqrt{n}\left(\frac{(\alpha\widehat{\alpha}_n-1)(Y_{s-1}+Y_{s+1})+(\widehat{\alpha}_n+\alpha+1-\alpha\widehat{\alpha}_n)\widehat{\mu}_{\varepsilon,n}}{(1+\alpha^2)(1+(\widehat{\alpha}_n)^2)}(\widehat{\alpha}_n-\alpha)-\frac{1-\alpha}{1+\alpha^2}(\widehat{\mu}_{\varepsilon,n}-\mu_\varepsilon)\right)
\end{aligned}$$

holds asymptotically as $n\to\infty$ with probability one. Hence

$$\begin{aligned}
&\sqrt{n}(\widehat{\theta}_n-\lim_{k\to\infty}\widehat{\theta}_k)\\
&\qquad=\begin{bmatrix}\frac{(\alpha\widehat{\alpha}_n-1)(Y_{s-1}+Y_{s+1})+(\widehat{\alpha}_n+\alpha+1-\alpha\widehat{\alpha}_n)\widehat{\mu}_{\varepsilon,n}}{(1+\alpha^2)(1+(\widehat{\alpha}_n)^2)} & -\frac{1-\alpha}{1+\alpha^2}\end{bmatrix}\begin{bmatrix}\sqrt{n}(\widehat{\alpha}_n-\alpha)\\ \sqrt{n}(\widehat{\mu}_{\varepsilon,n}-\mu_\varepsilon)\end{bmatrix}
\end{aligned}$$

holds asymptotically as $n \to \infty$ with probability one. Using Slutsky's lemma, by (3.3.3), (3.3.4) and (3.3.11), we have (3.3.12). $\square$

It can be checked that the asymptotic variances of $\sqrt{n}\left(\widetilde{\theta}_n - \lim_{k\to\infty}\widetilde{\theta}_k\right)$ and $\sqrt{n}\left(\widehat{\theta}_n - \lim_{k\to\infty}\widehat{\theta}_k\right)$ are not equal.

3.4 Two not neighbouring outliers, estimation of the mean of the offspring distribution and the outliers' sizes

In this section we assume that $I = 2$ and that the relevant time points $s_1, s_2 \in \mathbb{N}$ are known. We concentrate on the CLS estimation of α, θ_1 and θ_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 $s_1, s_2 \in \mathbb{N}$,

$$
\begin{aligned}
\mathsf{E}(Y_k \,|\, \mathcal{F}_{k-1}^Y) &= \alpha X_{k-1} + \mu_\varepsilon + \delta_{k,s_1}\theta_1 + \delta_{k,s_2}\theta_2 \\
&= \alpha(Y_{k-1} - \delta_{k-1,s_1}\theta_1 - \delta_{k-1,s_2}\theta_2) + \mu_\varepsilon + \delta_{k,s_1}\theta_1 + \delta_{k,s_2}\theta_2 \\
&= \alpha Y_{k-1} + \mu_\varepsilon + (-\alpha\delta_{k-1,s_1} + \delta_{k,s_1})\theta_1 + (-\alpha\delta_{k-1,s_2} + \delta_{k,s_2})\theta_2, \qquad k \in \mathbb{N}.
\end{aligned}
\tag{3.4.1}
$$

In the sequel we also suppose that $s_1 < s_2 - 1$, i.e., the time points s_1 and s_2 are not neighbouring. Then, by (3.4.1),

$$
\mathsf{E}(Y_k \,|\, \mathcal{F}_{k-1}^Y) = \begin{cases}
\alpha Y_{k-1} + \mu_\varepsilon & \text{if } 1 \leqslant k \leqslant s_1 - 1, \\
\alpha Y_{k-1} + \mu_\varepsilon + \theta_1 & \text{if } k = s_1, \\
\alpha Y_{k-1} + \mu_\varepsilon - \alpha\theta_1 & \text{if } k = s_1 + 1, \\
\alpha Y_{k-1} + \mu_\varepsilon & \text{if } s_1 + 2 \leqslant k \leqslant s_2 - 1, \\
\alpha Y_{k-1} + \mu_\varepsilon + \theta_2 & \text{if } k = s_2, \\
\alpha Y_{k-1} + \mu_\varepsilon - \alpha\theta_2 & \text{if } k = s_2 + 1, \\
\alpha Y_{k-1} + \mu_\varepsilon & \text{if } k \geqslant s_2 + 2.
\end{cases}
$$

Hence for all $n \geqslant s_2 + 1$, $n \in \mathbb{N}$,

$$
\begin{aligned}
\sum_{k=1}^n \big(Y_k - \mathsf{E}(Y_k \,|\, \mathcal{F}_{k-1}^Y)\big)^2 = &\sum_{\substack{k=1 \\ k \notin \{s_1, s_1+1, s_2, s_2+1\}}}^n \big(Y_k - \alpha Y_{k-1} - \mu_\varepsilon\big)^2 \\
&+ \big(Y_{s_1} - \alpha Y_{s_1-1} - \mu_\varepsilon - \theta_1\big)^2 + \big(Y_{s_1+1} - \alpha Y_{s_1} - \mu_\varepsilon + \alpha\theta_1\big)^2 \\
&+ \big(Y_{s_2} - \alpha Y_{s_2-1} - \mu_\varepsilon - \theta_2\big)^2 + \big(Y_{s_2+1} - \alpha Y_{s_2} - \mu_\varepsilon + \alpha\theta_2\big)^2.
\end{aligned}
\tag{3.4.2}
$$

For all $n \geqslant s_2 + 1$, $n \in \mathbb{N}$, we define the function $Q_n^\dagger : \mathbb{R}^{n+1} \times \mathbb{R}^3 \to \mathbb{R}$, as

$$
\begin{aligned}
&Q_n^\dagger(\mathbf{y}_n; \alpha', \theta_1', \theta_2') \\
&:= \sum_{\substack{k=1 \\ k \notin \{s_1, s_1+1, s_2, s_2+1\}}}^n \big(y_k - \alpha' y_{k-1} - \mu_\varepsilon\big)^2 + \big(y_{s_1} - \alpha' y_{s_1-1} - \mu_\varepsilon - \theta_1'\big)^2 \\
&\quad + \big(y_{s_1+1} - \alpha' y_{s_1} - \mu_\varepsilon + \alpha'\theta_1'\big)^2 + \big(y_{s_2} - \alpha' y_{s_2-1} - \mu_\varepsilon - \theta_2'\big)^2 + \big(y_{s_2+1} - \alpha' y_{s_2} - \mu_\varepsilon + \alpha'\theta_2'\big)^2,
\end{aligned}
$$

for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $\alpha', \theta_1', \theta_2' \in \mathbb{R}$. By definition, for all $n \geqslant s_2 + 1$, a CLS estimator for the parameter $(\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2$ is a measurable function $(\widetilde{\alpha}_n^\dagger, \widetilde{\theta}_{1,n}^\dagger, \widetilde{\theta}_{2,n}^\dagger) : S_n \to \mathbb{R}^3$ such that

$$Q_n^\dagger(\mathbf{y}_n; \widetilde{\alpha}_n^\dagger(\mathbf{y}_n), \widetilde{\theta}_{1,n}^\dagger(\mathbf{y}_n), \widetilde{\theta}_{2,n}^\dagger(\mathbf{y}_n)) = \inf_{(\alpha', \theta_1', \theta_2') \in \mathbb{R}^3} Q_n^\dagger(\mathbf{y}_n; \alpha', \theta_1', \theta_2') \qquad \forall \ \mathbf{y}_n \in S_n,$$

where S_n is suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 3.4.1). We note that we do not define the CLS estimator $(\widetilde{\alpha}_n^\dagger, \widetilde{\theta}_{1,n}^\dagger, \widetilde{\theta}_{2,n}^\dagger)$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$. For all $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $(\alpha', \theta_1', \theta_2') \in \mathbb{R}^3$,

$$\begin{aligned}
&\frac{\partial Q_n^\dagger}{\partial \alpha'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') \\
&\quad = \sum_{\substack{k=1 \\ k \notin \{s_1, s_1+1, s_2, s_2+1\}}}^{n} \big(y_k - \alpha' y_{k-1} - \mu_\varepsilon\big)(-2y_{k-1}) - 2\big(y_{s_1} - \alpha' y_{s_1-1} - \mu_\varepsilon - \theta_1'\big) y_{s_1-1} \\
&\qquad + 2\big(y_{s_1+1} - \alpha' y_{s_1} - \mu_\varepsilon + \alpha'\theta_1'\big)(-y_{s_1} + \theta_1') - 2\big(y_{s_2} - \alpha' y_{s_2-1} - \mu_\varepsilon - \theta_2'\big) y_{s_2-1} \\
&\qquad + 2\big(y_{s_2+1} - \alpha' y_{s_2} - \mu_\varepsilon + \alpha'\theta_2'\big)(-y_{s_2} + \theta_2'),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial Q_n^\dagger}{\partial \theta_1'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') &= -2(y_{s_1} - \alpha' y_{s_1-1} - \mu_\varepsilon - \theta_1') + 2\alpha'(y_{s_1+1} - \alpha' y_{s_1} - \mu_\varepsilon + \alpha'\theta_1'), \\
\frac{\partial Q_n^\dagger}{\partial \theta_2'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') &= -2(y_{s_2} - \alpha' y_{s_2-1} - \mu_\varepsilon - \theta_2') + 2\alpha'(y_{s_2+1} - \alpha' y_{s_2} - \mu_\varepsilon + \alpha'\theta_2').
\end{aligned}$$

The next lemma is about the existence and uniqueness of the CLS estimator of $(\alpha, \theta_1, \theta_2)$.

3.4.1 Lemma. *There exist subsets* $S_n \subset \mathbb{R}^{n+1}$, $n \geqslant s_2 + 1$ *with the following properties:*

(i) *there exists a unique CLS estimator* $(\widetilde{\alpha}_n^\dagger, \widetilde{\theta}_{1,n}^\dagger, \widetilde{\theta}_{2,n}^\dagger) : S_n \to \mathbb{R}^3$,

(ii) *for all* $\mathbf{y}_n \in S_n$, $(\widetilde{\alpha}_n^\dagger(\mathbf{y}_n), \widetilde{\theta}_{1,n}^\dagger(\mathbf{y}_n), \widetilde{\theta}_{2,n}^\dagger(\mathbf{y}_n))$ *is the unique solution of the system of equations*

$$\begin{aligned}
&\frac{\partial Q_n^\dagger}{\partial \alpha'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 0, \\
&\frac{\partial Q_n^\dagger}{\partial \theta_1'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 0, \qquad (3.4.3) \\
&\frac{\partial Q_n^\dagger}{\partial \theta_2'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 0,
\end{aligned}$$

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

Proof. For any fixed $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $\alpha' \in \mathbb{R}$, the quadratic function $\mathbb{R}^2 \ni (\theta_1', \theta_2') \mapsto Q_n^\dagger(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$ can be written in the form

$$Q_n^\dagger(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$$
$$= \left(\begin{bmatrix} \theta_1' \\ \theta_2' \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right)^\top A_n(\alpha') \left(\begin{bmatrix} \theta_1' \\ \theta_2' \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right) + \widetilde{Q}_n^\dagger(\mathbf{y}_n; \alpha'),$$

where

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

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

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

Then $\widetilde{Q}_n^\dagger(\mathbf{y}_n; \alpha') = R_n(\mathbf{y}_n; \alpha')/D_n(\alpha')$, where $D_n(\alpha') := (1+(\alpha')^2)^2$ and $\mathbb{R} \ni \alpha' \mapsto R_n(\mathbf{y}_n; \alpha')$ is a polynomial of order 6 with leading coefficient

$$c_n(\mathbf{y}_n) := \sum_{k=1}^n y_{k-1}^2 - (y_{s_1}^2 + y_{s_2}^2).$$

Let

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

For $\mathbf{y}_n \in \widetilde{S}_n^\dagger$, we have $\lim_{|\alpha'| \to \infty} \widetilde{Q}_n^\dagger(\mathbf{y}_n; \alpha') = \infty$ and the continuous function $\mathbb{R} \ni \alpha' \mapsto \widetilde{Q}_n^\dagger(\mathbf{y}_n; \alpha')$ attains its infimum. Consequently, for all $n \geqslant s_2 + 1$ there exists a CLS estimator $(\widetilde{\alpha}_n^\dagger, \widetilde{\theta}_{1,n}^\dagger, \widetilde{\theta}_{2,n}^\dagger) : \widetilde{S}_n^\dagger \to \mathbb{R}^3$, where

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

(3.4.4)
$$\begin{bmatrix} \widetilde{\theta}_{1,n}^\dagger(\mathbf{y}_n) \\ \widetilde{\theta}_{2,n}^\dagger(\mathbf{y}_n) \end{bmatrix} = A_n(\widetilde{\alpha}_n^\dagger(\mathbf{y}_n))^{-1} t_n(\mathbf{y}_n; \widetilde{\alpha}_n^\dagger(\mathbf{y}_n)), \qquad \mathbf{y}_n \in \widetilde{S}_n^\dagger,$$

and for all $\mathbf{y}_n \in \widetilde{S}_n^\dagger$, $(\widetilde{\alpha}_n^\dagger(\mathbf{y}_n), \widetilde{\theta}_{1,n}^\dagger(\mathbf{y}_n), \widetilde{\theta}_{2,n}^\dagger(\mathbf{y}_n))$ is a solution of the system of equations (3.4.3).

By (2.2.5) and (2.2.6), we get $\mathsf{P}\left(\lim_{n\to\infty} n^{-1} c_n(\mathbf{Y}_n) = \mathsf{E}\widetilde{X}^2\right) = 1$, where $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1). Hence $\mathbf{Y}_n \in \widetilde{S}_n^\dagger$ holds asymptotically as $n \to \infty$ with probability one.

Now we turn to find sets $S_n \subset \widetilde{S}_n^\dagger$, $n \geqslant s_2 + 1$ such that the system of equations (3.4.3) has a unique solution with respect to $(\alpha', \theta_1', \theta_2')$ for all $\mathbf{y}_n \in S_n$. Let us introduce the (3×3)

Hessian matrix

$$H_n(\mathbf{y}_n;\alpha',\theta_1',\theta_2') := \begin{bmatrix} \frac{\partial^2 Q_n^\dagger}{\partial(\alpha')^2} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\alpha'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\alpha'} \\ \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\theta_1'} & \frac{\partial^2 Q_n^\dagger}{\partial(\theta_1')^2} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\theta_1'} \\ \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\theta_2'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\theta_2'} & \frac{\partial^2 Q_n^\dagger}{\partial(\theta_2')^2} \end{bmatrix}(\mathbf{y}_n;\alpha',\theta_1',\theta_2'),$$

and let us denote by $\Delta_{i,n}(\mathbf{y}_n;\alpha',\theta_1',\theta_2')$ its i-th order leading principal minor, $i=1,2,3$. Further, for all $n\in\mathbb{N}$, let

$$S_n := \Big\{\mathbf{y}_n \in \widetilde{S}_n^\dagger : \Delta_{i,n}(\mathbf{y}_n;\alpha',\theta_1',\theta_2') > 0,\ \ i=1,2,3,\, \forall\,(\alpha',\theta_1',\theta_2')\in\mathbb{R}^3\Big\}.$$

By Berkovitz [10, Theorem 3.3, Chapter III], the function $\mathbb{R}^3 \ni (\alpha',\theta_1',\theta_2') \mapsto Q_n^\dagger(\mathbf{y}_n;\alpha',\theta_1',\theta_2')$ is strictly convex for all $\mathbf{y}_n\in S_n$. Since it was already proved that the system of equations (3.4.3) has a solution for all $\mathbf{y}_n\in\widetilde{S}_n^\dagger$, we obtain that this solution is unique for all $\mathbf{y}_n\in S_n$.

Next we check that $\mathbf{Y}_n\in S_n$ holds asymptotically as $n\to\infty$ with probability one. For all $(\alpha',\theta_1',\theta_2')\in\mathbb{R}^3$,

$$\begin{aligned}
&\frac{\partial^2 Q_n^\dagger}{\partial(\alpha')^2}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') \\
&\qquad = 2\sum_{\substack{k=1\\ k\notin\{s_1,s_1+1,s_2,s_2+1\}}}^{n} Y_{k-1}^2 + 2Y_{s_1-1}^2 + 2(Y_{s_1}-\theta_1')^2 + 2Y_{s_2-1}^2 + 2(Y_{s_2}-\theta_2')^2 \\
&\qquad = 2\sum_{\substack{k=1\\ k\notin\{s_1+1,s_2+1\}}}^{n} X_{k-1}^2 + 2(X_{s_1}+\theta_1-\theta_1')^2 + 2(X_{s_2}+\theta_2-\theta_2')^2,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\alpha'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') &= \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\theta_1'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 2(Y_{s_1-1}+Y_{s_1+1}-2\alpha' Y_{s_1}-\mu_\varepsilon+2\alpha'\theta_1') \\
&= 2(X_{s_1-1}+X_{s_1+1}-2\alpha' X_{s_1}-\mu_\varepsilon-2\alpha'(\theta_1-\theta_1')),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\alpha'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') &= \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\theta_2'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 2(Y_{s_2-1}+Y_{s_2+1}-2\alpha' Y_{s_2}-\mu_\varepsilon+2\alpha'\theta_2') \\
&= 2(X_{s_2-1}+X_{s_2+1}-2\alpha' X_{s_2}-\mu_\varepsilon-2\alpha'(\theta_2-\theta_2')),
\end{aligned}$$

and

$$\frac{\partial^2 Q_n^\dagger}{\partial(\theta_1')^2}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \frac{\partial^2 Q_n^\dagger}{\partial(\theta_2')^2}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 2((\alpha')^2+1),$$

$$\frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\theta_2'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\theta_1'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 0.$$

Then $H_n(\mathbf{Y}_n;\alpha',\theta_1',\theta_2')$ has the following leading principal minors

$$\Delta_{1,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \sum_{\substack{k=1\\ k\notin\{s_1+1,s_2+1\}}}^{n} 2X_{k-1}^2 + 2(X_{s_1}+\theta_1-\theta_1')^2 + 2(X_{s_2}+\theta_2-\theta_2')^2,$$

$$\Delta_{2,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 4\Bigg(((\alpha')^2+1)\Bigg(\sum_{\substack{k=1\\ k\notin\{s_1+1,s_2+1\}}}^{n} X_{k-1}^2 + (X_{s_1}+\theta_1-\theta_1')^2 + (X_{s_2}+\theta_2-\theta_2')^2\Bigg)$$
$$- \big(X_{s_1-1}+X_{s_1+1}-2\alpha'X_{s_1}-\mu_\varepsilon-2\alpha'(\theta_1-\theta_1')\big)^2\Bigg),$$

and

$$\Delta_{3,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \det H_n(\mathbf{Y}_n;\alpha',\theta_1',\theta_2')$$
$$= 8\Bigg(((\alpha')^2+1)^2\Bigg(\sum_{\substack{k=1\\ k\notin\{s_1+1,s_2+1\}}}^{n} X_{k-1}^2 + (X_{s_1}+\theta_1-\theta_1')^2 + (X_{s_2}+\theta_2-\theta_2')^2\Bigg)$$
$$- ((\alpha')^2+1)\big(X_{s_1-1}+X_{s_1+1}-2\alpha'X_{s_1}-\mu_\varepsilon-2\alpha'(\theta_1-\theta_1')\big)^2$$
$$- ((\alpha')^2+1)\big(X_{s_2-1}+X_{s_2+1}-2\alpha'X_{s_2}-\mu_\varepsilon-2\alpha'(\theta_2-\theta_2')\big)^2\Bigg).$$

By (2.2.6),

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\Delta_{1,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 2\mathsf{E}\widetilde{X}^2,\ \ \forall\ (\alpha',\theta_1',\theta_2')\in\mathbb{R}^3\right) = 1,$$
$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\Delta_{2,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 4((\alpha')^2+1)\mathsf{E}\widetilde{X}^2,\ \ \forall\ (\alpha',\theta_1',\theta_2')\in\mathbb{R}^3\right) = 1,$$

and

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\Delta_{3,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 8((\alpha')^2+1)^2\mathsf{E}\widetilde{X}^2,\ \ \forall\ (\alpha',\theta_1',\theta_2')\in\mathbb{R}^3\right) = 1,$$

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

$$\mathsf{P}\big(\lim_{n\to\infty}\Delta_{i,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \infty,\ \ \forall\ (\alpha',\theta_1',\theta_2')\in\mathbb{R}^3\big) = 1, \qquad i=1,2,3,$$

which yields that $\mathbf{Y}_n \in S_n$ asymptotically as $n\to\infty$ with probability one, since we have already proved that $\mathbf{Y}_n \in \widetilde{S}_n^\dagger$ asymptotically as $n\to\infty$ with probability one. □

By Lemma 3.4.1, $(\widetilde{\alpha}_n^\dagger(\mathbf{Y}_n),\widetilde{\theta}_{1,n}^\dagger(\mathbf{Y}_n),\widetilde{\theta}_{2,n}^\dagger(\mathbf{Y}_n))$ exists uniquely asymptotically as $n\to\infty$ with probability one. In the sequel we will simply denote it by $(\widetilde{\alpha}_n^\dagger,\widetilde{\theta}_{1,n}^\dagger,\widetilde{\theta}_{2,n}^\dagger)$.

An easy calculation shows that

$$\widetilde{\alpha}_n^\dagger = \frac{\sum_{k=1}^n (Y_k - \mu_\varepsilon) Y_{k-1} - \widetilde{\theta}_{1,n}^\dagger (Y_{s_1-1} + Y_{s_1+1} - \mu_\varepsilon) - \widetilde{\theta}_{2,n}^\dagger (Y_{s_2-1} + Y_{s_2+1} - \mu_\varepsilon)}{\sum_{k=1}^n Y_{k-1}^2 - 2\widetilde{\theta}_{1,n}^\dagger Y_{s_1} + (\widetilde{\theta}_{1,n}^\dagger)^2 - 2\widetilde{\theta}_{2,n}^\dagger Y_{s_2} + (\widetilde{\theta}_{2,n}^\dagger)^2}, \tag{3.4.5}$$

$$\widetilde{\theta}_{1,n}^\dagger = Y_{s_1} - \frac{\widetilde{\alpha}_n^\dagger}{1 + (\widetilde{\alpha}_n^\dagger)^2} (Y_{s_1-1} + Y_{s_1+1}) - \frac{1 - \widetilde{\alpha}_n^\dagger}{1 + (\widetilde{\alpha}_n^\dagger)^2} \mu_\varepsilon, \tag{3.4.6}$$

$$\widetilde{\theta}_{2,n}^\dagger = Y_{s_2} - \frac{\widetilde{\alpha}_n^\dagger}{1 + (\widetilde{\alpha}_n^\dagger)^2} (Y_{s_2-1} + Y_{s_2+1}) - \frac{1 - \widetilde{\alpha}_n^\dagger}{1 + (\widetilde{\alpha}_n^\dagger)^2} \mu_\varepsilon, \tag{3.4.7}$$

hold asymptotically as $n \to \infty$ with probability one.

The next result shows that $\widetilde{\alpha}_n^\dagger$ is a strongly consistent estimator of α, whereas $\widetilde{\theta}_{1,n}^\dagger$ and $\widetilde{\theta}_{2,n}^\dagger$ fail to be strongly consistent estimators of θ_1 and θ_2, respectively.

3.4.1 Theorem. *For the CLS estimators $(\widetilde{\alpha}_n^\dagger, \widetilde{\theta}_{1,n}^\dagger, \widetilde{\theta}_{2,n}^\dagger)_{n\in\mathbb{N}}$ of the parameter $(\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2$, the sequence $(\widetilde{\alpha}_n^\dagger)_{n\in\mathbb{N}}$ is strongly consistent for all $(\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2$, i.e.,*

$$\mathsf{P}(\lim_{n\to\infty} \widetilde{\alpha}_n^\dagger = \alpha) = 1, \qquad \forall\, (\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2, \tag{3.4.8}$$

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

$$\mathsf{P}\left(\lim_{n\to\infty} \widetilde{\theta}_{1,n}^\dagger = Y_{s_1} - \frac{\alpha}{1+\alpha^2}(Y_{s_1-1} + Y_{s_1+1}) - \frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon\right) = 1, \tag{3.4.9}$$

$$\mathsf{P}\left(\lim_{n\to\infty} \widetilde{\theta}_{2,n}^\dagger = Y_{s_2} - \frac{\alpha}{1+\alpha^2}(Y_{s_2-1} + Y_{s_2+1}) - \frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon\right) = 1, \tag{3.4.10}$$

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

Proof. Similarly to (3.2.11), we obtain

$$|\widetilde{\theta}_{1,n}^\dagger - \theta_1| \leqslant X_{s_1} + \frac{1}{2}(X_{s_1-1} + X_{s_1+1}) + \frac{3}{2}\mu_\varepsilon, \tag{3.4.11}$$

$$|\widetilde{\theta}_{2,n}^\dagger - \theta_2| \leqslant X_{s_2} + \frac{1}{2}(X_{s_2-1} + X_{s_2+1}) + \frac{3}{2}\mu_\varepsilon, \tag{3.4.12}$$

which yield that the sequences $(\widetilde{\theta}_{1,n}^\dagger - \theta_1)_{n\in\mathbb{N}}$ and $(\widetilde{\theta}_{2,n}^\dagger - \theta_2)_{n\in\mathbb{N}}$ are bounded with probability one. Using (3.4.5), (3.4.11) and (3.4.12), by the same arguments as in the proof of Theorem 3.2.1, one can derive (3.4.8). Then (3.4.8), (3.4.6) and (3.4.7) yield (3.4.9) and (3.4.10). □

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

$$\sqrt{n}(\widetilde{\alpha}_n^\dagger - \alpha) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, \sigma_{\alpha,\varepsilon}^2) \qquad \text{as } n \to \infty, \tag{3.4.13}$$

where $\sigma^2_{\alpha,\,\varepsilon}$ *is defined in* (2.2.9). *Moreover, conditionally on the values* Y_{s_1-1}, Y_{s_2-1} *and* Y_{s_1+1}, Y_{s_2+1},

$$\begin{bmatrix}\sqrt{n}\big(\widetilde{\theta}^{\dagger}_{1,n} - \lim_{k\to\infty}\widetilde{\theta}^{\dagger}_{1,k}\big)\\ \sqrt{n}\big(\widetilde{\theta}^{\dagger}_{2,n} - \lim_{k\to\infty}\widetilde{\theta}^{\dagger}_{2,k}\big)\end{bmatrix} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, e_{\alpha,\varepsilon}\sigma^2_{\alpha,\,\varepsilon}e^{\top}_{\alpha,\varepsilon}\right) \qquad \text{as } n\to\infty, \tag{3.4.14}$$

where

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

Proof. Using (3.4.5), (3.4.11) and (3.4.12), by the very same arguments as in the proof of (3.2.15), one can obtain (3.4.13). Now we turn to prove (3.4.14). Using the notation

$$B^{\dagger}_n := \begin{bmatrix}1+(\widetilde{\alpha}^{\dagger}_n)^2 & 0\\ 0 & 1+(\widetilde{\alpha}^{\dagger}_n)^2\end{bmatrix},$$

by (3.4.6) and (3.4.7), we have

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

holds asymptotically as $n\to\infty$ with probability one. Theorem 3.4.1 yields that

$$\mathsf{P}\left(\lim_{n\to\infty} B^{\dagger}_n = \begin{bmatrix}1+\alpha^2 & 0\\ 0 & 1+\alpha^2\end{bmatrix} =: B^{\dagger}\right) = 1.$$

By (3.4.9) and (3.4.10), we have

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

and hence

$$
\begin{bmatrix} \sqrt{n}\big(\widetilde{\theta}_{1,n}^{\dagger} - \lim_{k\to\infty} \widetilde{\theta}_{1,k}^{\dagger}\big) \\ \sqrt{n}\big(\widetilde{\theta}_{2,n}^{\dagger} - \lim_{k\to\infty} \widetilde{\theta}_{2,k}^{\dagger}\big) \end{bmatrix}
$$

$$
= \sqrt{n}(B_n^{\dagger})^{-1} \begin{bmatrix} (\widetilde{\alpha}_n^{\dagger} - \alpha)\big((\widetilde{\alpha}_n^{\dagger} + \alpha)Y_{s_1} - Y_{s_1-1} - Y_{s_1+1} + \mu_\varepsilon\big) \\ (\widetilde{\alpha}_n^{\dagger} - \alpha)\big((\widetilde{\alpha}_n^{\dagger} + \alpha)Y_{s_2} - Y_{s_2-1} - Y_{s_2+1} + \mu_\varepsilon\big) \end{bmatrix}
$$

$$
+ \sqrt{n}(B_n^{\dagger})^{-1}\big(B^{\dagger} - B_n^{\dagger}\big)(B^{\dagger})^{-1} \begin{bmatrix} (1+\alpha^2)Y_{s_1} - \alpha(Y_{s_1-1} + Y_{s_1+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s_2} - \alpha(Y_{s_2-1} + Y_{s_2+1}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}.
$$

Then

$$
\begin{bmatrix} \sqrt{n}\big(\widetilde{\theta}_{1,n}^{\dagger} - \lim_{k\to\infty} \widetilde{\theta}_{1,k}^{\dagger}\big) \\ \sqrt{n}\big(\widetilde{\theta}_{2,n}^{\dagger} - \lim_{k\to\infty} \widetilde{\theta}_{2,k}^{\dagger}\big) \end{bmatrix} = \sqrt{n}(\widetilde{\alpha}_n^{\dagger} - \alpha) \begin{bmatrix} K_n^{\dagger} \\ L_n^{\dagger} \end{bmatrix} \tag{3.4.15}
$$

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

$$
\begin{bmatrix} K_n^{\dagger} \\ L_n^{\dagger} \end{bmatrix} := (B_n^{\dagger})^{-1} \begin{bmatrix} -(\widetilde{\alpha}_n^{\dagger} + \alpha) & 0 \\ 0 & -(\widetilde{\alpha}_n^{\dagger} + \alpha) \end{bmatrix} (B^{\dagger})^{-1} \begin{bmatrix} (1+\alpha^2)Y_{s_1} - \alpha(Y_{s_1-1} + Y_{s_1+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s_2} - \alpha(Y_{s_2-1} + Y_{s_2+1}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}
$$

$$
+ (B_n^{\dagger})^{-1} \begin{bmatrix} (\widetilde{\alpha}_n^{\dagger} + \alpha)Y_{s_1} - Y_{s_1-1} - Y_{s_1+1} + \mu_\varepsilon \\ (\widetilde{\alpha}_n^{\dagger} + \alpha)Y_{s_2} - Y_{s_2-1} - Y_{s_2+1} + \mu_\varepsilon \end{bmatrix}.
$$

By (3.4.8), we have $\begin{bmatrix} K_n^{\dagger} & L_n^{\dagger} \end{bmatrix}^{\top}$ converges almost surely as $n \to \infty$ to

$$
(B^{\dagger})^{-1} \begin{bmatrix} 2\alpha Y_{s_1} - Y_{s_1-1} - Y_{s_1+1} + \mu_\varepsilon \\ 2\alpha Y_{s_2} - Y_{s_2-1} - Y_{s_2+1} + \mu_\varepsilon \end{bmatrix}
$$

$$
+ (B^{\dagger})^{-1} \begin{bmatrix} -2\alpha & 0 \\ 0 & -2\alpha \end{bmatrix} (B^{\dagger})^{-1} \begin{bmatrix} (1+\alpha^2)Y_{s_1} - \alpha(Y_{s_1-1} + Y_{s_1+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s_2} - \alpha(Y_{s_2-1} + Y_{s_2+1}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}
$$

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

By (3.4.15), (3.4.13) and Slutsky's lemma, we have (3.4.14). □

3.5 Two neighbouring outliers, estimation of the mean of the offspring distribution and the outliers' sizes

In this section we assume that $I = 2$ and that the relevant time points $s_1, s_2 \in \mathbb{N}$ are known. We also suppose that $s_1 := s$ and $s_2 := s + 1$, i.e., the time points s_1 and s_2 are neighbouring. We concentrate on the CLS estimation of α, θ_1 and θ_2. Then, by (3.4.1),

$$\mathsf{E}(Y_k \mid \mathcal{F}_{k-1}^Y) = \begin{cases} \alpha Y_{k-1} + \mu_\varepsilon & \text{if } 1 \leqslant k \leqslant s_1 - 1 = s - 1, \\ \alpha Y_{k-1} + \mu_\varepsilon + \theta_1 & \text{if } k = s_1 = s, \\ \alpha Y_{k-1} + \mu_\varepsilon - \alpha\theta_1 + \theta_2 & \text{if } k = s + 1 = s_1 + 1 = s_2, \\ \alpha Y_{k-1} + \mu_\varepsilon - \alpha\theta_2 & \text{if } k = s + 2 = s_1 + 2 = s_2 + 1, \\ \alpha Y_{k-1} + \mu_\varepsilon & \text{if } k \geqslant s + 2 = s_2 + 2. \end{cases}$$

Hence

(3.5.1)

$$\sum_{k=1}^{n} \big(Y_k - \mathsf{E}(Y_k \mid \mathcal{F}_{k-1}^Y)\big)^2$$

$$= \sum_{\substack{k=1 \\ k\notin\{s,s+1,s+2\}}}^{n} \big(Y_k - \alpha Y_{k-1} - \mu_\varepsilon\big)^2 + \big(Y_s - \alpha Y_{s-1} - \mu_\varepsilon - \theta_1\big)^2 + \big(Y_{s+1} - \alpha Y_s - \mu_\varepsilon + \alpha\theta_1 - \theta_2\big)^2$$

$$+ \big(Y_{s+2} - \alpha Y_{s+1} - \mu_\varepsilon + \alpha\theta_2\big)^2, \qquad n \geqslant s + 2, \ n \in \mathbb{N}.$$

For all $n \geqslant s + 2$, $n \in \mathbb{N}$, we define the function $Q_n^{\dagger\dagger} : \mathbb{R}^{n+1} \times \mathbb{R}^3 \to \mathbb{R}$, as

$$Q_n^{\dagger\dagger}(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$$

$$:= \sum_{\substack{k=1 \\ k\notin\{s,s+1,s+2\}}}^{n} \big(y_k - \alpha' y_{k-1} - \mu_\varepsilon\big)^2 + \big(y_s - \alpha' y_{s-1} - \mu_\varepsilon - \theta_1'\big)^2 + \big(y_{s+1} - \alpha' y_s - \mu_\varepsilon + \alpha'\theta_1' - \theta_2'\big)^2$$

$$+ \big(y_{s+2} - \alpha' y_{s+1} - \mu_\varepsilon + \alpha'\theta_2'\big)^2, \qquad \mathbf{y}_n \in \mathbb{R}^{n+1}, \ \alpha', \theta_1', \theta_2' \in \mathbb{R}.$$

By definition, for all $n \geqslant s + 2$, a CLS estimator for the parameter $(\alpha, \theta_1, \theta_2) \in (0, 1) \times \mathbb{N}^2$ is a measurable function $(\widetilde{\alpha}_n^{\dagger\dagger}, \widetilde{\theta}_{1,n}^{\dagger\dagger}, \widetilde{\theta}_{2,n}^{\dagger\dagger}) : S_n \to \mathbb{R}^3$ such that

$$Q_n^{\dagger\dagger}(\mathbf{y}_n; \widetilde{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n), \widetilde{\theta}_{1,n}^{\dagger\dagger}(\mathbf{y}_n), \widetilde{\theta}_{2,n}^{\dagger\dagger}(\mathbf{y}_n)) = \inf_{(\alpha',\theta_1',\theta_2')\in\mathbb{R}^3} Q_n^{\dagger\dagger}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') \qquad \forall \ \mathbf{y}_n \in S_n,$$

where S_n is suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 3.5.1). We note that we do not define the CLS estimator $(\widetilde{\alpha}_n^{\dagger\dagger}, \widetilde{\theta}_{1,n}^{\dagger\dagger}, \widetilde{\theta}_{2,n}^{\dagger\dagger})$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$. We have

$$\frac{\partial Q_n^{\dagger\dagger}}{\partial \alpha'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$$

$$= \sum_{\substack{k=1 \\ k\notin\{s,s+1,s+2\}}}^{n} \big(y_k - \alpha' y_{k-1} - \mu_\varepsilon\big)(-2y_{k-1}) - 2\big(y_s - \alpha' y_{s-1} - \mu_\varepsilon - \theta_1'\big)y_{s-1}$$

$$+ 2\big(y_{s+1} - \alpha' y_s - \mu_\varepsilon + \alpha'\theta_1' - \theta_2'\big)(-y_s + \theta_1') + 2\big(y_{s+2} - \alpha' y_{s+1} - \mu_\varepsilon + \alpha'\theta_2'\big)(-y_{s+1} + \theta_2'),$$

and

$$\frac{\partial Q_n^{\dagger\dagger}}{\partial \theta_1'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = -2(y_s - \alpha' y_{s-1} - \mu_\varepsilon - \theta_1') + 2\alpha'(y_{s+1} - \alpha' y_s - \mu_\varepsilon + \alpha'\theta_1' - \theta_2'),$$

$$\frac{\partial Q_n^{\dagger\dagger}}{\partial \theta_2'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = -2(y_{s+1} - \alpha' y_s - \mu_\varepsilon + \alpha'\theta_1' - \theta_2') + 2\alpha'(y_{s+2} - \alpha' y_{s+1} - \mu_\varepsilon + \alpha'\theta_2').$$

The next lemma is about the existence and uniqueness of the CLS estimator of $(\alpha, \theta_1, \theta_2)$.

3.5.1 Lemma. *There exist subsets* $S_n \subset \mathbb{R}^{n+1}$, $n \geqslant s+2$ *with the following properties:*

(i) *there exists a unique CLS estimator* $(\widetilde{\alpha}_n^{\dagger\dagger}, \widetilde{\theta}_{1,n}^{\dagger\dagger}, \widetilde{\theta}_{2,n}^{\dagger\dagger}) : S_n \to \mathbb{R}^3$,

(ii) *for all* $\mathbf{y}_n \in S_n$, $(\widetilde{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n), \widetilde{\theta}_{1,n}^{\dagger\dagger}(\mathbf{y}_n), \widetilde{\theta}_{2,n}^{\dagger\dagger}(\mathbf{y}_n))$ *is the unique solution of the system of equations*

$$\begin{aligned} &\frac{\partial Q_n^{\dagger\dagger}}{\partial \alpha'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 0, \\ &\frac{\partial Q_n^{\dagger\dagger}}{\partial \theta_1'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 0, \\ &\frac{\partial Q_n^{\dagger\dagger}}{\partial \theta_2'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 0, \end{aligned} \tag{3.5.2}$$

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

Proof. For any fixed $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $\alpha' \in \mathbb{R}$, the quadratic function $\mathbb{R}^2 \ni (\theta_1', \theta_2') \mapsto Q_n^{\dagger\dagger}(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$ can be written in the form

$$\begin{aligned} &Q_n^{\dagger\dagger}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') \\ &= \left(\begin{bmatrix} \theta_1' \\ \theta_2' \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right)^\top A_n(\alpha') \left(\begin{bmatrix} \theta_1' \\ \theta_2' \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right) + \widetilde{Q}_n^{\dagger\dagger}(\mathbf{y}_n; \alpha'), \end{aligned}$$

where

$$t_n(\mathbf{y}_n; \alpha') := \begin{bmatrix} (1+(\alpha')^2) y_s - \alpha'(y_{s-1} + y_{s+1}) - (1-\alpha')\mu_\varepsilon \\ (1+(\alpha')^2) y_{s+1} - \alpha'(y_s + y_{s+2}) - (1-\alpha')\mu_\varepsilon \end{bmatrix},$$

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

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

Then $\widetilde{Q}_n^{\dagger\dagger}(\mathbf{y}_n; \alpha') = R_n(\mathbf{y}_n; \alpha') / D_n(\alpha')$, where $D_n(\alpha') := (1+(\alpha')^2)^2 - (\alpha')^2 = (\alpha')^4 + (\alpha')^2 + 1 > 0$ and $\mathbb{R} \ni \alpha' \mapsto R_n(\mathbf{y}_n; \alpha')$ is a polynomial of order 6 with leading coefficient

$$c_n(\mathbf{y}_n) := \sum_{k=1}^n y_{k-1}^2 - (y_s^2 + y_{s+1}^2).$$

Let

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

For $\mathbf{y}_n \in \widetilde{S}_n^{\dagger\dagger}$, we have $\lim_{|\alpha'|\to\infty} \widetilde{Q}_n^{\dagger\dagger}(\mathbf{y}_n;\alpha') = \infty$ and the continuous function $\mathbb{R} \ni \alpha' \mapsto \widetilde{Q}_n^{\dagger\dagger}(\mathbf{y}_n;\alpha')$ attains its infimum. Consequently, for all $n \geqslant s+2$ there exists a CLS estimator $(\widetilde{\alpha}_n^{\dagger\dagger}, \widetilde{\theta}_{1,n}^{\dagger\dagger}, \widetilde{\theta}_{2,n}^{\dagger\dagger}) : \widetilde{S}_n^{\dagger\dagger} \to \mathbb{R}^3$, where

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

$$\begin{bmatrix} \widetilde{\theta}_{1,n}^{\dagger\dagger}(\mathbf{y}_n) \\ \widetilde{\theta}_{2,n}^{\dagger\dagger}(\mathbf{y}_n) \end{bmatrix} = A_n(\widetilde{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n))^{-1} t_n(\mathbf{y}_n;\widetilde{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n)), \qquad \mathbf{y}_n \in \widetilde{S}_n^{\dagger\dagger}, \tag{3.5.3}$$

and for all $\mathbf{y}_n \in \widetilde{S}_n^{\dagger\dagger}$, $(\widetilde{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n), \widetilde{\theta}_{1,n}^{\dagger\dagger}(\mathbf{y}_n), \widetilde{\theta}_{2,n}^{\dagger\dagger}(\mathbf{y}_n))$ is a solution of the system of equations (3.5.2).

By (2.2.5) and (2.2.6), we get $\mathsf{P}\left(\lim_{n\to\infty} n^{-1} c_n(\mathbf{Y}_n) = \mathsf{E}\widetilde{X}^2\right) = 1$, where $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1). Hence $\mathbf{Y}_n \in \widetilde{S}_n^{\dagger\dagger}$ holds asymptotically as $n \to \infty$ with probability one.

Now we turn to find sets $S_n \subset \widetilde{S}_n^{\dagger\dagger}$, $n \geqslant s+2$ such that the system of equations (3.5.2) has a unique solution with respect to $(\alpha', \theta_1', \theta_2')$ for all $\mathbf{y}_n \in S_n$. Let us introduce the (3×3) Hessian matrix

$$H_n(\mathbf{y}_n;\alpha',\theta_1',\theta_2') := \begin{bmatrix} \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\alpha')^2} & \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta_1'\partial\alpha'} & \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta_2'\partial\alpha'} \\ \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\alpha'\partial\theta_1'} & \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\theta_1')^2} & \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta_2'\partial\theta_1'} \\ \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\alpha'\partial\theta_2'} & \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta_1'\partial\theta_2'} & \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\theta_2')^2} \end{bmatrix} (\mathbf{y}_n;\alpha',\theta_1',\theta_2'),$$

and let us denote by $\Delta_{i,n}(\mathbf{y}_n;\alpha',\theta_1',\theta_2')$ its i-th order leading principal minor, $i = 1,2,3$. Further, for all $n \geqslant s+2$, let

$$S_n := \Big\{ \mathbf{y}_n \in \widetilde{S}^{\dagger\dagger} : \Delta_{i,n}(\mathbf{y}_n;\alpha',\theta_1',\theta_2') > 0,\ i = 1,2,3,\ \forall\,(\alpha',\theta_1',\theta_2') \in \mathbb{R}^3 \Big\}.$$

By Berkovitz [10, Theorem 3.3, Chapter III], the function $\mathbb{R}^3 \ni (\alpha',\theta_1',\theta_2') \mapsto Q_n^{\dagger\dagger}(\mathbf{y}_n;\alpha',\theta_1',\theta_2')$ is strictly convex for all $\mathbf{y}_n \in S_n$. Since it was already proved that the system of equations (3.5.2) has a solution for all $\mathbf{y}_n \in \widetilde{S}_n^{\dagger\dagger}$, we obtain that this solution is unique for all $\mathbf{y}_n \in S_n$.

Next we check that $\mathbf{Y}_n \in S_n$ holds asymptotically as $n \to \infty$ with probability one. For all $(\alpha',\theta_1',\theta_2') \in \mathbb{R}^3$,

$$\begin{aligned} \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\alpha')^2}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') &= 2 \sum_{\substack{k=1 \\ k\notin\{s,s+1,s+2\}}}^{n} Y_{k-1}^2 + 2Y_{s-1}^2 + 2(Y_s - \theta_1')^2 + 2(Y_{s+1} - \theta_2')^2 \\ &= 2 \sum_{\substack{k=1 \\ k\notin\{s+1,s+2\}}}^{n} X_{k-1}^2 + 2(X_s + \theta_1 - \theta_1')^2 + 2(X_{s+1} + \theta_2 - \theta_2')^2, \end{aligned}$$

and

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta_1'\partial\alpha'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\alpha'\partial\theta_1'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2')$$

$$= 2\big(Y_{s-1} + Y_{s+1} - 2\alpha' Y_s - \mu_\varepsilon + 2\alpha'\theta_1' - \theta_2'\big)$$

$$= 2\big(X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu_\varepsilon - 2\alpha'(\theta_1 - \theta_1') + (\theta_2 - \theta_2')\big),$$

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta_2'\partial\alpha'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\alpha'\partial\theta_2'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2')$$

$$= 2\big(Y_s + Y_{s+2} - 2\alpha' Y_{s+1} - \mu_\varepsilon - \theta_1' + 2\alpha'\theta_2'\big)$$

$$= 2\big(X_s + X_{s+2} - 2\alpha' X_{s+1} - \mu_\varepsilon + (\theta_1 - \theta_1') - 2\alpha'(\theta_2 - \theta_2')\big),$$

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\theta_1')^2}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\theta_2')^2}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 2((\alpha')^2 + 1),$$

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta_1'\partial\theta_2'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta_2'\partial\theta_1'}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = -2\alpha'.$$

Then $H_n(\mathbf{Y}_n;\alpha',\theta_1',\theta_2')$ has the following leading principal minors

$$\Delta_{1,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 2\sum_{\substack{k=1\\ k\notin\{s+1,s+2\}}}^{n} X_{k-1}^2 + 2(X_s + \theta_1 - \theta_1')^2 + 2(X_{s+1} + \theta_2 - \theta_2')^2,$$

$$\Delta_{2,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 4\Bigg(((\alpha')^2+1)\Bigg(\sum_{\substack{k=1\\ k\notin\{s+1,s+2\}}}^{n} X_{k-1}^2 + (X_s + \theta_1 - \theta_1')^2 + (X_{s+1} + \theta_2 - \theta_2')^2\Bigg)$$

$$- \big(X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu_\varepsilon - 2\alpha'(\theta_1 - \theta_1') + (\theta_2 - \theta_2')\big)^2\Bigg),$$

and

$$\Delta_{3,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \det H_n(\mathbf{Y}_n;\alpha',\theta_1',\theta_2')$$

$$= 8\Bigg[((\alpha')^4 + (\alpha')^2 + 1)\Bigg(\sum_{\substack{k=1\\ k\notin\{s+1,s+2\}}}^{n} X_{k-1}^2 + (X_s + \theta_1 - \theta_1')^2 + (X_{s+1} + \theta_2 - \theta_2')^2\Bigg)$$

$$- 2\alpha' ab - ((\alpha')^2+1)b^2 - ((\alpha')^2+1)a^2\Bigg],$$

where

$$a := X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu_\varepsilon - 2\alpha'(\theta_1 - \theta_1') + (\theta_2 - \theta_2'),$$

$$b := X_s + X_{s+2} - 2\alpha' X_{s+1} - \mu_\varepsilon + (\theta_1 - \theta_1') - 2\alpha'(\theta_2 - \theta_2').$$

By (2.2.6),

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\Delta_{1,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 2\mathsf{E}\widetilde{X}^2,\ \ \forall\ (\alpha',\theta_1',\theta_2')\in\mathbb{R}^3\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\Delta_{2,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 4((\alpha')^2+1)\mathsf{E}\widetilde{X}^2,\ \ \forall\ (\alpha',\theta_1',\theta_2')\in\mathbb{R}^3\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\Delta_{3,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = 8((\alpha')^4+(\alpha')^2+1)\mathsf{E}\widetilde{X}^2,\ \ \forall\ (\alpha',\theta_1',\theta_2')\in\mathbb{R}^3\right) = 1,$$

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

$$\mathsf{P}\big(\lim_{n\to\infty}\Delta_{i,n}(\mathbf{Y}_n;\alpha',\theta_1',\theta_2') = \infty,\ \ \forall\ (\alpha',\theta_1',\theta_2')\in\mathbb{R}^3\big) = 1, \qquad i=1,2,3,$$

which yields that $\mathbf{Y}_n\in S_n$ asymptotically as $n\to\infty$ with probability one, since we have already proved that $\mathbf{Y}_n\in\widetilde{S}_n^{\dagger\dagger}$ asymptotically as $n\to\infty$ with probability one. □

By Lemma 3.5.1, $(\widetilde{\alpha}_n^{\dagger\dagger}(\mathbf{Y}_n),\widetilde{\theta}_{1,n}^{\dagger\dagger}(\mathbf{Y}_n),\widetilde{\theta}_{2,n}^{\dagger\dagger}(\mathbf{Y}_n))$ exists uniquely asymptotically as $n\to\infty$ with probability one. In the sequel we will simply denote it by $(\widetilde{\alpha}_n^{\dagger\dagger},\widetilde{\theta}_{1,n}^{\dagger\dagger},\widetilde{\theta}_{2,n}^{\dagger\dagger})$.

An easy calculation shows that

$$(3.5.4)\quad \widetilde{\alpha}_n^{\dagger\dagger} = \frac{\sum_{k=1}^n (Y_k-\mu_\varepsilon)Y_{k-1} - \widetilde{\theta}_{1,n}^{\dagger\dagger}(Y_{s-1}+Y_{s+1}-\mu_\varepsilon) - \widetilde{\theta}_{2,n}^{\dagger\dagger}(Y_s+Y_{s+2}-\mu_\varepsilon) + \widetilde{\theta}_{1,n}^{\dagger\dagger}\widetilde{\theta}_{2,n}^{\dagger\dagger}}{\sum_{k=1}^n Y_{k-1}^2 - 2\widetilde{\theta}_{1,n}^{\dagger\dagger}Y_s + (\widetilde{\theta}_{1,n}^{\dagger\dagger})^2 - 2\widetilde{\theta}_{2,n}^{\dagger\dagger}Y_{s+1} + (\widetilde{\theta}_{2,n}^{\dagger\dagger})^2},$$

and

$$(3.5.5)\quad \begin{bmatrix} 1+(\widetilde{\alpha}_n^{\dagger\dagger})^2 & -\widetilde{\alpha}_n^{\dagger\dagger} \\ -\widetilde{\alpha}_n^{\dagger\dagger} & 1+(\widetilde{\alpha}_n^{\dagger\dagger})^2 \end{bmatrix} \begin{bmatrix} \widetilde{\theta}_{1,n}^{\dagger\dagger} \\ \widetilde{\theta}_{2,n}^{\dagger\dagger} \end{bmatrix} = \begin{bmatrix} Y_s - \widetilde{\alpha}_n^{\dagger\dagger}Y_{s-1} - \mu_\varepsilon - \widetilde{\alpha}_n^{\dagger\dagger}(Y_{s+1}-\widetilde{\alpha}_n^{\dagger\dagger}Y_s-\mu_\varepsilon) \\ Y_{s+1} - \widetilde{\alpha}_n^{\dagger\dagger}Y_s - \mu_\varepsilon - \widetilde{\alpha}_n^{\dagger\dagger}(Y_{s+2}-\widetilde{\alpha}_n^{\dagger\dagger}Y_{s+1}-\mu_\varepsilon) \end{bmatrix}$$

hold asymptotically as $n\to\infty$ with probability one. Recalling that $D_n(\widetilde{\alpha}_n^{\dagger\dagger}) = (\widetilde{\alpha}_n^{\dagger\dagger})^4 + (\widetilde{\alpha}_n^{\dagger\dagger})^2 + 1 > 0$, we have

$$(3.5.6)\quad \begin{aligned} \widetilde{\theta}_{1,n}^{\dagger\dagger} = \frac{1}{D_n(\widetilde{\alpha}_n^{\dagger\dagger})}\Big((1+(\widetilde{\alpha}_n^{\dagger\dagger})^2)\Big[Y_s - \widetilde{\alpha}_n^{\dagger\dagger}Y_{s-1} - \mu_\varepsilon - \widetilde{\alpha}_n^{\dagger\dagger}(Y_{s+1}-\widetilde{\alpha}_n^{\dagger\dagger}Y_s-\mu_\varepsilon)\Big] \\ + \widetilde{\alpha}_n^{\dagger\dagger}\Big[Y_{s+1} - \widetilde{\alpha}_n^{\dagger\dagger}Y_s - \mu_\varepsilon - \widetilde{\alpha}_n^{\dagger\dagger}(Y_{s+2}-\widetilde{\alpha}_n^{\dagger\dagger}Y_{s+1}-\mu_\varepsilon)\Big]\Big), \end{aligned}$$

and

$$(3.5.7)\quad \begin{aligned} \widetilde{\theta}_{2,n}^{\dagger\dagger} = \frac{1}{D_n(\widetilde{\alpha}_n^{\dagger\dagger})}\Big(\widetilde{\alpha}_n^{\dagger\dagger}\Big[Y_s - \widetilde{\alpha}_n^{\dagger\dagger}Y_{s-1} - \mu_\varepsilon - \widetilde{\alpha}_n^{\dagger\dagger}(Y_{s+1}-\widetilde{\alpha}_n^{\dagger\dagger}Y_s-\mu_\varepsilon)\Big] \\ + (1+(\widetilde{\alpha}_n^{\dagger\dagger})^2)\Big[Y_{s+1} - \widetilde{\alpha}_n^{\dagger\dagger}Y_s - \mu_\varepsilon - \widetilde{\alpha}_n^{\dagger\dagger}(Y_{s+2}-\widetilde{\alpha}_n^{\dagger\dagger}Y_{s+1}-\mu_\varepsilon)\Big]\Big) \end{aligned}$$

hold asymptotically as $n\to\infty$ with probability one.

The next result shows that $\widetilde{\alpha}_n^{\dagger\dagger}$ is a strongly consistent estimator of α, whereas $\widetilde{\theta}_{1,n}^{\dagger\dagger}$ and $\widetilde{\theta}_{2,n}^{\dagger\dagger}$ fail to be strongly consistent estimators of θ_1 and θ_2, respectively.

3.5.1 Theorem. *For the CLS estimators* $(\widetilde{\alpha}_n^{\dagger\dagger}, \widetilde{\theta}_{1,n}^{\dagger\dagger}, \widetilde{\theta}_{2,n}^{\dagger\dagger})_{n\in\mathbb{N}}$ *of the parameter* $(\alpha, \theta_1, \theta_2) \in (0,1)\times\mathbb{N}^2$, *the sequence* $(\widetilde{\alpha}_n^{\dagger\dagger})_{n\in\mathbb{N}}$ *is strongly consistent for all* $(\alpha, \theta_1, \theta_2) \in (0,1)\times\mathbb{N}^2$, *i.e.,*

$$\mathsf{P}(\lim_{n\to\infty} \widetilde{\alpha}_n^{\dagger\dagger} = \alpha) = 1, \qquad \forall\, (\alpha, \theta_1, \theta_2) \in (0,1)\times\mathbb{N}^2, \tag{3.5.8}$$

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

$$\mathsf{P}\left(\lim_{n\to\infty}\begin{bmatrix}\widetilde{\theta}_{1,n}^{\dagger\dagger}\\ \widetilde{\theta}_{2,n}^{\dagger\dagger}\end{bmatrix} = \begin{bmatrix}Y_s\\ Y_{s+1}\end{bmatrix} + \begin{bmatrix}\frac{-\alpha(1+\alpha^2)Y_{s-1}-\alpha^2Y_{s+2}-(1-\alpha^3)\mu_\varepsilon}{1+\alpha^2+\alpha^4}\\ \frac{-\alpha^2Y_{s-1}-\alpha(1+\alpha^2)Y_{s+2}-(1-\alpha^3)\mu_\varepsilon}{1+\alpha^2+\alpha^4}\end{bmatrix}\right) = 1 \tag{3.5.9}$$

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

Proof. Using that for all $p_i \in \mathbb{R}$, $i = 0, 1, \ldots, 4$,

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

by (3.5.6) and (3.5.7), we get the sequences $(\widetilde{\theta}_{1,n}^{\dagger\dagger})_{n\in\mathbb{N}}$ and $(\widetilde{\theta}_{2,n}^{\dagger\dagger})_{n\in\mathbb{N}}$ are bounded with probability one. Hence using (3.5.4), by the same arguments as in the proof of Theorem 3.2.1, one can derive (3.5.8). Then (3.5.8), (3.5.6) and (3.5.7) yield (3.5.9). □

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

$$\sqrt{n}(\widetilde{\alpha}_n^{\dagger\dagger} - \alpha) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, \sigma_{\alpha,\,\varepsilon}^2) \qquad \text{as } n\to\infty, \tag{3.5.10}$$

where $\sigma_{\alpha,\,\varepsilon}^2$ *is defined in* (2.2.9). *Moreover, conditionally on the values* Y_{s-1} *and* Y_{s+2},

$$\begin{bmatrix}\sqrt{n}\big(\widetilde{\theta}_{1,n}^{\dagger\dagger} - \lim_{k\to\infty}\widetilde{\theta}_{1,k}^{\dagger\dagger}\big)\\ \sqrt{n}\big(\widetilde{\theta}_{2,n}^{\dagger\dagger} - \lim_{k\to\infty}\widetilde{\theta}_{2,k}^{\dagger\dagger}\big)\end{bmatrix} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, f_{\alpha,\varepsilon}\sigma_{\alpha,\,\varepsilon}^2 f_{\alpha,\varepsilon}^\top\right) \qquad \text{as } n\to\infty, \tag{3.5.11}$$

where $f_{\alpha,\varepsilon}$ *defined by*

$$\frac{1}{(1+\alpha^2+\alpha^4)^2}\begin{bmatrix}(\alpha^2-1)(\alpha^4+3\alpha^2+1)Y_{s-1} + 2\alpha(\alpha^4-1)Y_{s+2} + \alpha(2-\alpha)(1+\alpha+\alpha^2)^2\mu_\varepsilon\\ 2\alpha(\alpha^4-1)Y_{s-1} + (\alpha^2-1)(\alpha^4+3\alpha^2+1)Y_{s+2} + \alpha(2-\alpha)(1+\alpha+\alpha^2)^2\mu_\varepsilon\end{bmatrix}.$$

Proof. Using (3.5.4) and that the sequences $(\widetilde{\theta}_{1,n}^{\dagger\dagger})_{n\in\mathbb{N}}$ and $(\widetilde{\theta}_{2,n}^{\dagger\dagger})_{n\in\mathbb{N}}$ are bounded with probability one, by the very same arguments as in the proof of (3.2.15), one can obtain (3.5.10). Now we turn to prove (3.5.11). Using the notation

$$B_n^{\dagger\dagger} := \begin{bmatrix}1 + (\widetilde{\alpha}_n^{\dagger\dagger})^2 & -\widetilde{\alpha}_n^{\dagger\dagger}\\ -\widetilde{\alpha}_n^{\dagger\dagger} & 1 + (\widetilde{\alpha}_n^{\dagger\dagger})^2\end{bmatrix},$$

by (3.5.5), we have

$$\begin{bmatrix}\widetilde{\theta}_{1,n}^{\dagger\dagger}\\ \widetilde{\theta}_{2,n}^{\dagger\dagger}\end{bmatrix} = (B_n^{\dagger\dagger})^{-1}\begin{bmatrix}(1 + (\widetilde{\alpha}_n^{\dagger\dagger})^2)Y_s - \widetilde{\alpha}_n^{\dagger\dagger}(Y_{s-1} + Y_{s+1}) - (1 - \widetilde{\alpha}_n^{\dagger\dagger})\mu_\varepsilon\\ (1 + (\widetilde{\alpha}_n^{\dagger\dagger})^2)Y_{s+1} - \widetilde{\alpha}_n^{\dagger\dagger}(Y_s + Y_{s+2}) - (1 - \widetilde{\alpha}_n^{\dagger\dagger})\mu_\varepsilon\end{bmatrix}$$

holds asymptotically as $n \to \infty$ with probability one. Theorem 3.5.1 yields that

$$\mathsf{P}\left(\lim_{n\to\infty} B_n^{\dagger\dagger} = \begin{bmatrix} 1+\alpha^2 & -\alpha \\ -\alpha & 1+\alpha^2 \end{bmatrix} =: B^{\dagger\dagger}\right) = 1.$$

By (3.5.9), we have

$$\begin{bmatrix} \sqrt{n}\big(\widetilde{\theta}_{1,n}^{\dagger\dagger} - \lim_{k\to\infty}\widetilde{\theta}_{1,k}^{\dagger\dagger}\big) \\ \sqrt{n}\big(\widetilde{\theta}_{2,n}^{\dagger\dagger} - \lim_{k\to\infty}\widetilde{\theta}_{2,k}^{\dagger\dagger}\big) \end{bmatrix} =$$

$$= \sqrt{n}(B_n^{\dagger\dagger})^{-1}\left(\begin{bmatrix} (1+(\widetilde{\alpha}_n^{\dagger\dagger})^2)Y_s - \widetilde{\alpha}_n^{\dagger\dagger}(Y_{s-1}+Y_{s+1}) - (1-\widetilde{\alpha}_n^{\dagger\dagger})\mu_\varepsilon \\ (1+(\widetilde{\alpha}_n^{\dagger\dagger})^2)Y_{s+1} - \widetilde{\alpha}_n^{\dagger\dagger}(Y_s+Y_{s+2}) - (1-\widetilde{\alpha}_n^{\dagger\dagger})\mu_\varepsilon \end{bmatrix}\right.$$

$$\left. - B_n^{\dagger\dagger}(B^{\dagger\dagger})^{-1}\begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1}+Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s+Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}\right),$$

and hence

$$\begin{bmatrix} \sqrt{n}\big(\widetilde{\theta}_{1,n}^{\dagger\dagger} - \lim_{k\to\infty}\widetilde{\theta}_{1,k}^{\dagger\dagger}\big) \\ \sqrt{n}\big(\widetilde{\theta}_{2,n}^{\dagger\dagger} - \lim_{k\to\infty}\widetilde{\theta}_{2,k}^{\dagger\dagger}\big) \end{bmatrix} =$$

$$= \sqrt{n}(B_n^{\dagger\dagger})^{-1}\left(\begin{bmatrix} (1+(\widetilde{\alpha}_n^{\dagger\dagger})^2)Y_s - \widetilde{\alpha}_n^{\dagger\dagger}(Y_{s-1}+Y_{s+1}) - (1-\widetilde{\alpha}_n^{\dagger\dagger})\mu_\varepsilon \\ (1+(\widetilde{\alpha}_n^{\dagger\dagger})^2)Y_{s+1} - \widetilde{\alpha}_n^{\dagger\dagger}(Y_s+Y_{s+2}) - (1-\widetilde{\alpha}_n^{\dagger\dagger})\mu_\varepsilon \end{bmatrix}\right.$$

$$\left. - \begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1}+Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s+Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}\right)$$

$$+ \sqrt{n}\left((B_n^{\dagger\dagger})^{-1} - (B^{\dagger\dagger})^{-1}\right)\begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1}+Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s+Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}$$

$$= \sqrt{n}(B_n^{\dagger\dagger})^{-1}\begin{bmatrix} (\widetilde{\alpha}_n^{\dagger\dagger}-\alpha)\big((\widetilde{\alpha}_n^{\dagger\dagger}+\alpha)Y_s - Y_{s-1} - Y_{s+1} + \mu_\varepsilon\big) \\ (\widetilde{\alpha}_n^{\dagger\dagger}-\alpha)\big((\widetilde{\alpha}_n^{\dagger\dagger}+\alpha)Y_{s+1} - Y_s - Y_{s+2} + \mu_\varepsilon\big) \end{bmatrix}$$

$$+ \sqrt{n}(B_n^{\dagger\dagger})^{-1}\big(B^{\dagger\dagger} - B_n^{\dagger\dagger}\big)(B^{\dagger\dagger})^{-1}\begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1}+Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s+Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}.$$

Then

$$\begin{bmatrix} \sqrt{n}\big(\widetilde{\theta}_{1,n}^{\dagger\dagger} - \lim_{k\to\infty}\widetilde{\theta}_{1,k}^{\dagger\dagger}\big) \\ \sqrt{n}\big(\widetilde{\theta}_{2,n}^{\dagger\dagger} - \lim_{k\to\infty}\widetilde{\theta}_{2,k}^{\dagger\dagger}\big) \end{bmatrix} = \sqrt{n}(\widetilde{\alpha}_n^{\dagger\dagger}-\alpha)\begin{bmatrix} K_n^{\dagger\dagger} \\ L_n^{\dagger\dagger} \end{bmatrix} \tag{3.5.12}$$

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

$$\begin{bmatrix} K_n^{\dagger\dagger} \\ L_n^{\dagger\dagger} \end{bmatrix} := (B_n^{\dagger\dagger})^{-1}\begin{bmatrix} -(\widetilde{\alpha}_n^{\dagger\dagger}+\alpha) & 1 \\ 1 & -(\widetilde{\alpha}_n^{\dagger\dagger}+\alpha) \end{bmatrix}(B^{\dagger\dagger})^{-1}\begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1}+Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s+Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}$$

$$+ (B_n^{\dagger\dagger})^{-1}\begin{bmatrix} (\widetilde{\alpha}_n^{\dagger\dagger}+\alpha)Y_s - Y_{s-1} - Y_{s+1} + \mu_\varepsilon \\ (\widetilde{\alpha}_n^{\dagger\dagger}+\alpha)Y_{s+1} - Y_s - Y_{s+2} + \mu_\varepsilon \end{bmatrix}.$$

By (3.5.8), we have $\begin{bmatrix} K_n^{\dagger\dagger} & L_n^{\dagger\dagger} \end{bmatrix}^\top$ converges almost surely as $n \to \infty$ to

$$
\begin{aligned}
&(B^{\dagger\dagger})^{-1} \begin{bmatrix} 2\alpha Y_s - Y_{s-1} - Y_{s+1} + \mu_\varepsilon \\ 2\alpha Y_{s+1} - Y_s - Y_{s+2} + \mu_\varepsilon \end{bmatrix} \\
&\qquad + (B^{\dagger\dagger})^{-1} \begin{bmatrix} -2\alpha & 1 \\ 1 & -2\alpha \end{bmatrix} (B^{\dagger\dagger})^{-1} \begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1} + Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s + Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix} \\
&= \frac{1}{1+\alpha^2+\alpha^4} \begin{bmatrix} 1+\alpha^2 & \alpha \\ \alpha & 1+\alpha^2 \end{bmatrix} \left(\begin{bmatrix} 2\alpha Y_s - Y_{s-1} - Y_{s+1} + \mu_\varepsilon \\ 2\alpha Y_{s+1} - Y_s - Y_{s+2} + \mu_\varepsilon \end{bmatrix} \right. \\
&\left. + \frac{1}{(1+\alpha^2+\alpha^4)^2} \begin{bmatrix} -2\alpha^5 - 4\alpha^3 & 1 - \alpha^2 - 3\alpha^4 \\ 1 - \alpha^2 - 3\alpha^4 & -2\alpha^5 - 4\alpha^3 \end{bmatrix} \begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1} + Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s + Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix} \right),
\end{aligned}
$$

which is equal to $f_{\alpha,\varepsilon}$, by an easy, but tedious calculation. Hence, by (3.5.12), (3.5.10) and Slutsky's lemma, we have (3.5.11). □

3.6 Two not neighbouring outliers, estimation of the mean of the offspring and innovation distributions and the outliers' sizes

In this section we assume that $I = 2$ and that the relevant time points $s_1, s_2 \in \mathbb{N}$ are known. We also suppose that $s_1 < s_2 - 1$, i.e., the time points s_1 and s_2 are not neighbouring. We concentrate on the CLS estimation of α, μ_ε, θ_1 and θ_2.

Motivated by (3.4.2), for all $n \geqslant s_2+1$, $n \in \mathbb{N}$, we define the function $Q_n^\dagger : \mathbb{R}^{n+1} \times \mathbb{R}^4 \to \mathbb{R}$, as

$$
\begin{aligned}
&Q_n^\dagger(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') \\
&:= \sum_{\substack{k=1 \\ k \notin \{s_1, s_1+1, s_2, s_2+1\}}}^{n} \big(y_k - \alpha' y_{k-1} - \mu_\varepsilon'\big)^2 + \big(y_{s_1} - \alpha' y_{s_1-1} - \mu_\varepsilon' - \theta_1'\big)^2 \\
&\quad + \big(y_{s_1+1} - \alpha' y_{s_1} - \mu_\varepsilon' + \alpha'\theta_1'\big)^2 + \big(y_{s_2} - \alpha' y_{s_2-1} - \mu_\varepsilon' - \theta_2'\big)^2 + \big(y_{s_2+1} - \alpha' y_{s_2} - \mu_\varepsilon' + \alpha'\theta_2'\big)^2,
\end{aligned}
$$

for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $\alpha', \mu_\varepsilon', \theta_1', \theta_2' \in \mathbb{R}$. By definition, for all $n \geqslant s_2 + 1$, 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

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

such that

$$
\begin{aligned}
&Q_n^\dagger(\mathbf{y}_n; \widehat{\alpha}_n^\dagger(\mathbf{y}_n), \widehat{\mu}_{\varepsilon,n}^\dagger(\mathbf{y}_n), \widehat{\theta}_{1,n}^\dagger(\mathbf{y}_n), \widehat{\theta}_{2,n}^\dagger(\mathbf{y}_n)) \\
&\qquad = \inf_{(\alpha', \mu_\varepsilon', \theta_1', \theta_2') \in \mathbb{R}^4} Q_n^\dagger(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') \qquad \forall\, \mathbf{y}_n \in S_n,
\end{aligned}
$$

where S_n is suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 3.6.1). We note that we do not define the CLS estimator $(\widehat{\alpha}_n^\dagger, \widehat{\mu}_{\varepsilon,n}^\dagger, \widehat{\theta}_{1,n}^\dagger, \widehat{\theta}_{2,n}^\dagger)$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$.

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

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

(i) *there exists a unique CLS estimator* $(\widehat{\alpha}_n^\dagger, \widehat{\mu}_{\varepsilon,n}^\dagger, \widehat{\theta}_{1,n}^\dagger, \widehat{\theta}_{2,n}^\dagger) : S_n \to \mathbb{R}^4$,

(ii) *for all* $\mathbf{y}_n \in S_n$, $(\widehat{\alpha}_n^\dagger(\mathbf{y}_n), \widehat{\mu}_{\varepsilon,n}^\dagger(\mathbf{y}_n), \widehat{\theta}_{1,n}^\dagger(\mathbf{y}_n), \widehat{\theta}_{2,n}^\dagger(\mathbf{y}_n))$ *is the unique solution of the system of equations*

$$
\begin{aligned}
&\frac{\partial Q_n^\dagger}{\partial \alpha'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 0, \qquad \frac{\partial Q_n^\dagger}{\partial \mu_\varepsilon'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 0,\\
&\frac{\partial Q_n^\dagger}{\partial \theta_1'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 0, \qquad \frac{\partial Q_n^\dagger}{\partial \theta_2'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 0,
\end{aligned}
\tag{3.6.1}
$$

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

Proof. For any fixed $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $n \geqslant \max(5, s_2+1)$ 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

$$
\begin{aligned}
&Q_n^\dagger(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2')\\
&= \left(\begin{bmatrix} \mu_\varepsilon' \\ \theta_1' \\ \theta_2' \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right)^\top A_n(\alpha') \left(\begin{bmatrix} \mu_\varepsilon' \\ \theta_1' \\ \theta_2' \end{bmatrix} - A_n(\alpha')^{-1} t_n(\mathbf{y}_n; \alpha') \right) + \widehat{Q}_n^\dagger(\mathbf{y}_n; \alpha'),
\end{aligned}
$$

where

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

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

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 $n \geqslant 5$ and $\alpha' \in \mathbb{R}$. Indeed, the leading principal minors of $A_n(\alpha')$ take the following forms: n,

$$
\begin{aligned}
&n(1+(\alpha')^2) - (1-\alpha')^2 = (n-1)(\alpha')^2 + 2\alpha' + n - 1,\\
&D_n(\alpha') := (1+(\alpha')^2)\big((n-2)(\alpha')^2 + 4\alpha' + n - 2\big),
\end{aligned}
$$

and for all $n \geqslant 5$, the discriminant $16 - 4(n-2)^2$ of the equation $(n-2)x^2 + 4x + n - 2 = 0$ is negative.

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

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

The polynomial $\mathbb{R} \ni \alpha' \mapsto D_n(\alpha')$ is of order 4 with leading coefficient $n-2$. We have $\widehat{Q}_n^\dagger(\mathbf{y}_n;\alpha') = R_n(\mathbf{y}_n;\alpha')/D_n(\alpha')$, where $\mathbb{R} \ni \alpha' \mapsto R_n(\mathbf{y}_n;\alpha')$ is a polynomial of order 6 with leading coefficient

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

Let

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

For $\mathbf{y}_n \in \widehat{S}_n^\dagger$, we have $\lim_{|\alpha'|\to\infty} \widehat{Q}_n^\dagger(\mathbf{y}_n;\alpha') = \infty$ and the continuous function $\mathbb{R} \ni \alpha' \mapsto \widehat{Q}_n^\dagger(\mathbf{y}_n;\alpha')$ attains its infimum. Consequently, for all $n \geqslant \max(5, s_2+1)$ there exists a CLS estimator $(\widehat{\alpha}_n^\dagger, \widehat{\mu}_{\varepsilon,n}^\dagger, \widehat{\theta}_{1,n}^\dagger, \widehat{\theta}_{2,n}^\dagger) : \widehat{S}_n^\dagger \to \mathbb{R}^4$, where

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

$$\text{(3.6.2)} \qquad \begin{bmatrix} \widehat{\mu}_{\varepsilon,n}^\dagger(\mathbf{y}_n) \\ \widehat{\theta}_{1,n}^\dagger(\mathbf{y}_n) \\ \widehat{\theta}_{2,n}^\dagger(\mathbf{y}_n) \end{bmatrix} = A_n(\widehat{\alpha}_n^\dagger(\mathbf{y}_n))^{-1} t_n(\mathbf{y}_n;\widehat{\alpha}_n^\dagger(\mathbf{y}_n)), \qquad \mathbf{y}_n \in \widehat{S}_n^\dagger,$$

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 (3.6.1).

By (2.2.5) and (2.2.6), we get $\mathsf{P}\left(\lim_{n\to\infty} n^{-2}c_n(\mathbf{Y}_n) = \operatorname{Var}\widetilde{X}\right) = 1$, where $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1). Hence $\mathbf{Y}_n \in \widehat{S}_n^\dagger$ holds asymptotically as $n \to \infty$ with probability one.

Now we turn to find sets $S_n \subset \widehat{S}_n^\dagger$, $n \geqslant \max(5, s_2+1)$ such that the system of equations (3.6.1) has a unique solution with respect to $(\alpha', \mu_\varepsilon', \theta_1', \theta_2')$ for all $\mathbf{y}_n \in S_n$. Let us introduce the (4×4) Hessian matrix

$$H_n(\mathbf{y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') := \begin{bmatrix} \frac{\partial^2 Q_n^\dagger}{\partial(\alpha')^2} & \frac{\partial^2 Q_n^\dagger}{\partial\mu_\varepsilon'\partial\alpha'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\alpha'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\alpha'} \\ \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\mu_\varepsilon'} & \frac{\partial^2 Q_n^\dagger}{\partial(\mu_\varepsilon')^2} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\mu_\varepsilon'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\mu_\varepsilon'} \\ \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\theta_1'} & \frac{\partial^2 Q_n^\dagger}{\partial\mu_\varepsilon'\partial\theta_1'} & \frac{\partial^2 Q_n^\dagger}{\partial(\theta_1')^2} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\theta_1'} \\ \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\theta_2'} & \frac{\partial^2 Q_n^\dagger}{\partial\mu_\varepsilon'\partial\theta_2'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\theta_2'} & \frac{\partial^2 Q_n^\dagger}{\partial(\theta_2')^2} \end{bmatrix}(\mathbf{y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2'),$$

and let us denote by $\Delta_{i,n}(\mathbf{y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)$ its i-th order leading principal minor, $i=1,2,3,4$. Further, for all $n \geqslant \max(5, s_2+1)$, let

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

By Berkovitz [10, Theorem 3.3, Chapter III], the function $\mathbb{R}^4 \ni (\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2) \mapsto Q^\dagger_n(\mathbf{y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)$ is strictly convex for all $\mathbf{y}_n \in S_n$. Since it was already proved that the system of equations (3.6.1) has a solution for all $\mathbf{y}_n \in \widehat{S}^\dagger_n$, we obtain that this solution is unique for all $\mathbf{y}_n \in S_n$.

For all $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $(\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2) \in \mathbb{R}^4$, we have

$$\begin{aligned}
&\frac{\partial Q^\dagger_n}{\partial \alpha'}(\mathbf{y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)\\
&\quad= \sum_{\substack{k=1\\ k\notin\{s_1,s_1+1,s_2,s_2+1\}}}^{n} \big(y_k-\alpha' y_{k-1}-\mu'_\varepsilon\big)(-2y_{k-1}) - 2\big(y_{s_1}-\alpha' y_{s_1-1}-\mu'_\varepsilon-\theta'_1\big)y_{s_1-1}\\
&\qquad + 2\big(y_{s_1+1}-\alpha' y_{s_1}-\mu'_\varepsilon+\alpha'\theta'_1\big)(-y_{s_1}+\theta'_1) - 2\big(y_{s_2}-\alpha' y_{s_2-1}-\mu'_\varepsilon-\theta'_2\big)y_{s_2-1}\\
&\qquad + 2\big(y_{s_2+1}-\alpha' y_{s_2}-\mu'_\varepsilon+\alpha'\theta'_2\big)(-y_{s_2}+\theta'_2),
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial Q^\dagger_n}{\partial \mu'_\varepsilon}(\mathbf{y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)\\
&\quad= \sum_{\substack{k=1\\ k\notin\{s_1,s_1+1,s_2,s_2+1\}}}^{n} (-2)\big(y_k-\alpha' y_{k-1}-\mu'_\varepsilon\big) - 2\big(y_{s_1}-\alpha' y_{s_1-1}-\mu'_\varepsilon-\theta'_1\big)\\
&\qquad - 2\big(y_{s_1+1}-\alpha' y_{s_1}-\mu'_\varepsilon+\alpha'\theta'_1\big) - 2\big(y_{s_2}-\alpha' y_{s_2-1}-\mu'_\varepsilon-\theta'_2\big)\\
&\qquad - 2\big(y_{s_2+1}-\alpha' y_{s_2}-\mu'_\varepsilon+\alpha'\theta'_2\big),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial Q^\dagger_n}{\partial \theta'_i}(\mathbf{y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)\\
&\quad= -2(y_{s_i}-\alpha' y_{s_i-1}-\mu'_\varepsilon-\theta'_i) + 2\alpha'(y_{s_i+1}-\alpha' y_{s_i}-\mu'_\varepsilon+\alpha'\theta'_i), \qquad i=1,2.
\end{aligned} \tag{3.6.3}$$

We also get for all $(\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2) \in \mathbb{R}^4$,

$$\begin{aligned}
&\frac{\partial^2 Q^\dagger_n}{\partial (\alpha')^2}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)\\
&\qquad= 2\sum_{\substack{k=1\\ k\notin\{s_1,s_1+1,s_2,s_2+1\}}}^{n} Y_{k-1}^2 + 2Y_{s_1-1}^2 + 2(Y_{s_1}-\theta'_1)^2 + 2Y_{s_2-1}^2 + 2(Y_{s_2}-\theta'_2)^2\\
&\qquad= 2\sum_{\substack{k=1\\ k\notin\{s_1+1,s_2+1\}}}^{n} X_{k-1}^2 + 2(X_{s_1}+\theta_1-\theta'_1)^2 + 2(X_{s_2}+\theta_2-\theta'_2)^2,
\end{aligned}$$

and

$$\frac{\partial^2 Q_n^\dagger}{\partial \mu_\varepsilon' \partial \alpha'}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = \frac{\partial^2 Q_n^\dagger}{\partial \alpha' \partial \mu_\varepsilon'}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2')$$

$$= 2\sum_{k=1}^n Y_{k-1} - 2\theta_1' - 2\theta_2' = 2\sum_{k=1}^n X_{k-1} + 2(\theta_1 - \theta_1') + 2(\theta_2 - \theta_2'),$$

$$\frac{\partial^2 Q_n^\dagger}{\partial \theta_i' \partial \alpha'}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = \frac{\partial^2 Q_n^\dagger}{\partial \alpha' \partial \theta_i'}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2')$$

$$= 2(Y_{s_i-1} + Y_{s_i+1} - 2\alpha' Y_{s_i} - \mu_\varepsilon' + 2\alpha'\theta_2')$$

$$= 2(X_{s_i-1} + X_{s_i+1} - 2\alpha' X_{s_i} - \mu_\varepsilon' - 2\alpha'(\theta_2 - \theta_2')), \qquad i = 1, 2,$$

$$\frac{\partial^2 Q_n^\dagger}{\partial (\mu_\varepsilon')^2}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 2n,$$

$$\frac{\partial^2 Q_n^\dagger}{\partial (\theta_1')^2}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = \frac{\partial^2 Q_n^\dagger}{\partial (\theta_2')^2}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 2((\alpha')^2 + 1),$$

$$\frac{\partial^2 Q_n^\dagger}{\partial \theta_1' \partial \theta_2'}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = \frac{\partial^2 Q_n^\dagger}{\partial \theta_2' \partial \theta_1'}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 0,$$

$$\frac{\partial^2 Q_n^\dagger}{\partial \theta_i' \partial \mu_\varepsilon'}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = \frac{\partial^2 Q_n^\dagger}{\partial \mu_\varepsilon' \partial \theta_i'}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 2(1 - \alpha'), \qquad i = 1, 2.$$

The matrix $H_n(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2')$ has the following leading principal minors

$$\Delta_{1,n}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 2 \sum_{\substack{k=1 \\ k \notin \{s_1+1, s_2+1\}}}^n X_{k-1}^2 + 2(X_{s_1} + \theta_1 - \theta_1')^2 + 2(X_{s_2} + \theta_2 - \theta_2')^2,$$

$$\Delta_{2,n}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 4n \left(\sum_{\substack{k=1 \\ k \notin \{s_1+1, s_2+1\}}}^n X_{k-1}^2 + (X_{s_1} + \theta_1 - \theta_1')^2 + (X_{s_2} + \theta_2 - \theta_2')^2 \right)$$
$$- 4\left(\sum_{k=1}^n X_{k-1} + (\theta_1 - \theta_1') + (\theta_2 - \theta_2') \right)^2,$$

$$\Delta_{3,n}(\mathbf{Y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 8\Big(((\alpha')^2 + 1)n - (1 - \alpha')^2 \Big)$$
$$\times \left(\sum_{\substack{k=1 \\ k \notin \{s_1+1, s_2+1\}}}^n X_{k-1}^2 + (X_{s_1} + \theta_1 - \theta_1')^2 + (X_{s_2} + \theta_2 - \theta_2')^2 \right)$$
$$+ 16(1 - \alpha')L \left(\sum_{k=1}^n X_{k-1} + (\theta_1 - \theta_1') + (\theta_2 - \theta_2') \right) - 8nL^2$$
$$- 8((\alpha')^2 + 1) \left(\sum_{k=1}^n X_{k-1} + (\theta_1 - \theta_1') + (\theta_2 - \theta_2') \right)^2$$

and

$$\Delta_{4,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2) = \det H_n(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2),$$

where $L := X_{s_1-1} + X_{s_1+1} - 2\alpha' X_{s_1} - \mu'_\varepsilon - 2\alpha'(\theta_1 - \theta'_1)$. By (2.2.5) and (2.2.6), we get the following events have probability one

$$\left\{\lim_{n\to\infty}\frac{1}{n}\Delta_{1,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2) = 2\mathsf{E}\widetilde{X}^2,\ \forall\,(\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)\in\mathbb{R}^4\right\},$$

$$\begin{aligned}\Bigg\{\lim_{n\to\infty}\frac{1}{n^2}\Delta_{2,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2) &= 4(\mathsf{E}\widetilde{X}^2 - (\mathsf{E}\widetilde{X})^2)\\ &= 4\operatorname{Var}\widetilde{X},\ \forall\,(\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)\in\mathbb{R}^4\Bigg\},\end{aligned}$$

$$\left\{\lim_{n\to\infty}\frac{1}{n^2}\Delta_{3,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2) = 8((\alpha')^2+1)\operatorname{Var}\widetilde{X},\ \forall\,(\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)\in\mathbb{R}^4\right\},$$

$$\left\{\lim_{n\to\infty}\frac{1}{n^2}\Delta_{4,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2) = 16((\alpha')^2+1)^2\operatorname{Var}\widetilde{X},\ \forall\,(\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)\in\mathbb{R}^4\right\},$$

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

$$\mathsf{P}\Big(\lim_{n\to\infty}\Delta_{i,n}(\mathbf{Y}_n;\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2) = \infty,\quad \forall\,(\alpha',\mu'_\varepsilon,\theta'_1,\theta'_2)\in\mathbb{R}^4,\ \ i=1,2,3,4\Big) = 1,$$

which yields that $\mathbf{Y}_n \in S_n$ asymptotically as $n\to\infty$ with probability one, since we have already proved that $\mathbf{Y}_n \in \widehat{S}_n^\dagger$ asymptotically as $n\to\infty$ with probability one. □

By Lemma 3.6.1, $(\widehat{\alpha}_n^\dagger(\mathbf{Y}_n), \widehat{\mu}_{\varepsilon,n}^\dagger(\mathbf{Y}_n), \widehat{\theta}_{1,n}^\dagger(\mathbf{Y}_n), \widehat{\theta}_{2,n}^\dagger(\mathbf{Y}_n))$ exists uniquely asymptotically as $n\to\infty$ with probability one. In the sequel we will simply denote it by $(\widehat{\alpha}_n^\dagger, \widehat{\mu}_{\varepsilon,n}^\dagger, \widehat{\theta}_{1,n}^\dagger, \widehat{\theta}_{2,n}^\dagger)$.

The next result shows that $\widehat{\alpha}_n^\dagger$ is a strongly consistent estimator of α, $\widehat{\mu}_{\varepsilon,n}^\dagger$ is a strongly consistent estimator of μ_ε, whereas $\widehat{\theta}_{1,n}^\dagger$ and $\widehat{\theta}_{2,n}^\dagger$ fail to be strongly consistent estimators of θ_1 and θ_2, respectively.

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

(3.6.4) $$\mathsf{P}(\lim_{n\to\infty}\widehat{\alpha}_n^\dagger = \alpha) = 1,\qquad \forall\,(\alpha,\mu_\varepsilon,\theta_1,\theta_2)\in(0,1)\times(0,\infty)\times\mathbb{N}^2,$$

(3.6.5) $$\mathsf{P}(\lim_{n\to\infty}\widehat{\mu}_{\varepsilon,n}^\dagger = \mu_\varepsilon) = 1,\qquad \forall\,(\alpha,\mu_\varepsilon,\theta_1,\theta_2)\in(0,1)\times(0,\infty)\times\mathbb{N}^2,$$

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

(3.6.6) $$\mathsf{P}\left(\lim_{n\to\infty}\widehat{\theta}_{i,n}^\dagger = Y_{s_i} - \frac{\alpha}{1+\alpha^2}(Y_{s_i-1}+Y_{s_i+1}) - \frac{1-\alpha}{1+\alpha^2}\mu_\varepsilon\right) = 1,\qquad i=1,2,$$

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

Proof. 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. By (3.6.1), (3.6.3) and Lemma 3.6.1, we get

$$\widehat{\theta}_{i,n}^\dagger = Y_{s_i} - \frac{\widehat{\alpha}_n^\dagger}{1+(\widehat{\alpha}_n^\dagger)^2}(Y_{s_i-1}+Y_{s_i+1}) - \frac{1-\widehat{\alpha}_n^\dagger}{1+(\widehat{\alpha}_n^\dagger)^2}\widehat{\mu}_{\varepsilon,n}^\dagger \qquad i=1,2. \tag{3.6.7}$$

By (3.6.2) and the explicit form of the inverse matrix $A_n(\alpha')^{-1}$, we obtain

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

where

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

Using (2.2.5) and that for all $p_i \in \mathbb{R}$, $i=0,\dots,4$,

$$\sup_{x\in\mathbb{R},\, n\geqslant 5} \frac{n(p_4x^4+p_3x^3+p_2x^2+p_1x+p_0)}{(1+x^2)((n-2)x^2+4x+n-2)} < \infty,$$

one can think it over that $H_n/D_n(\widehat{\alpha}_n^\dagger)$, $n\in\mathbb{N}$, and $J_n/D_n(\widehat{\alpha}_n^\dagger)$, $n\in\mathbb{N}$, are bounded with probability one, which yields also 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.

Again by Lemma 3.6.1 and equations (3.6.1) we get that

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

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

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

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

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

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

Here we emphasize that the matrix

$$\begin{bmatrix} a_n & b_n \\ b_n & n \end{bmatrix}$$

is invertible asymptotically as $n \to \infty$ with probability one, since using (2.2.5), (2.2.6) and 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 we get

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{a_n}{n} = \mathsf{E}\widetilde{X}^2\right) = 1, \qquad \mathsf{P}\left(\lim_{n\to\infty} \frac{b_n}{n} = \mathsf{E}\widetilde{X}\right) = 1, \tag{3.6.8}$$

and hence

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n^2}(na_n - b_n^2) = \mathsf{E}\widetilde{X}^2 - (\mathsf{E}\widetilde{X})^2 = \operatorname{Var}\widetilde{X}\right) = 1.$$

This yields that

$$\mathsf{P}\left(\lim_{n\to\infty}(na_n - b_n^2) = \infty\right) = 1.$$

Further

$$\begin{bmatrix} \widehat{\alpha}_n^{\dagger} - \alpha \\ \widehat{\mu}_{\varepsilon,n}^{\dagger} - \mu_\varepsilon \end{bmatrix} = \begin{bmatrix} a_n & b_n \\ b_n & n \end{bmatrix}^{-1} \begin{bmatrix} e_n \\ f_n \end{bmatrix} \tag{3.6.9}$$

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

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

$$f_n := \sum_{k=1}^{n}(X_k - \alpha X_{k-1} - \mu_\varepsilon) + (1-\alpha)(\theta_1 - \widehat{\theta}_{1,n}^{\dagger} + \theta_2 - \widehat{\theta}_{2,n}^{\dagger}).$$

Then, using again (2.2.3), (2.2.5), (2.2.6), (2.2.7) and that the sequences $(\widehat{\theta}^{\dagger}_{1,n} - \theta_1)_{n\in\mathbb{N}}$ and $(\widehat{\theta}^{\dagger}_{2,n} - \theta_2)_{n\in\mathbb{N}}$ are bounded with probability one, we get

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

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

Hence, by (3.6.9), we obtain

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

which yields (3.6.4) and (3.6.5). Then (3.6.4), (3.6.5) and (3.6.7) imply (3.6.6). □

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

$$\begin{bmatrix}\sqrt{n}(\widehat{\alpha}^{\dagger}_n - \alpha)\\ \sqrt{n}(\widehat{\mu}^{\dagger}_{\varepsilon,n} - \mu_\varepsilon)\end{bmatrix} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, B_{\alpha,\varepsilon}\right) \qquad \text{as } n\to\infty, \tag{3.6.10}$$

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

$$\begin{bmatrix}\sqrt{n}\big(\widehat{\theta}^{\dagger}_{1,n} - \lim_{k\to\infty}\widehat{\theta}^{\dagger}_{1,k}\big)\\ \sqrt{n}\big(\widehat{\theta}^{\dagger}_{2,n} - \lim_{k\to\infty}\widehat{\theta}^{\dagger}_{2,k}\big)\end{bmatrix} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, C_{\alpha,\varepsilon}B_{\alpha,\varepsilon}C_{\alpha,\varepsilon}^\top\right) \qquad \text{as } n\to\infty, \tag{3.6.11}$$

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. By (2.3.2) and (3.6.9), to prove (3.6.10) it is enough to show that

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\begin{bmatrix}a_n & b_n\\ b_n & n\end{bmatrix} = \begin{bmatrix}\mathsf{E}\widetilde{X}^2 & \mathsf{E}\widetilde{X}\\ \mathsf{E}\widetilde{X} & 1\end{bmatrix}\right) = 1, \tag{3.6.12}$$

$$\frac{1}{\sqrt{n}}\begin{bmatrix}e_n\\ f_n\end{bmatrix} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, A_{\alpha,\varepsilon}\right) \qquad \text{as } n\to\infty, \tag{3.6.13}$$

where $\widetilde{X}$ is a random variable having the unique stationary distribution of the INAR(1) model in (2.1.1) and the (2×2)-matrix $A_{\alpha,\varepsilon}$ is defined in (2.3.3). By (3.6.8), we have (3.6.12). By formula (6.43) in Hall and Heyde [35, Section 6.3],

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

Hence using that the sequences $(\widehat{\theta}^\dagger_{1,n}-\theta_1)_{n\in\mathbb{N}}$ and $(\widehat{\theta}^\dagger_{2,n}-\theta_2)_{n\in\mathbb{N}}$ are bounded with probability one, by Slutsky's lemma (see, e.g., Lemma 2.8 in van der Vaart [63]), we get (3.6.13).

Now we turn to prove (3.6.11). Using the notation

$$B^\dagger_n := \begin{bmatrix} 1+(\widehat{\alpha}^\dagger_n)^2 & 0 \\ 0 & 1+(\widehat{\alpha}^\dagger_n)^2 \end{bmatrix},$$

by (3.6.7), we have

$$\begin{bmatrix} \widehat{\theta}^\dagger_{1,n} \\ \widehat{\theta}^\dagger_{2,n} \end{bmatrix} = (B^\dagger_n)^{-1} \begin{bmatrix} (1+(\widehat{\alpha}^\dagger_n)^2)Y_{s_1} - \widehat{\alpha}^\dagger_n(Y_{s_1-1}+Y_{s_1+1}) - (1-\widehat{\alpha}^\dagger_n)\widehat{\mu}^\dagger_{\varepsilon,n} \\ (1+(\widehat{\alpha}^\dagger_n)^2)Y_{s_2} - \widehat{\alpha}^\dagger_n(Y_{s_2-1}+Y_{s_2+1}) - (1-\widehat{\alpha}^\dagger_n)\widehat{\mu}^\dagger_{\varepsilon,n} \end{bmatrix}$$

holds asymptotically as $n\to\infty$ with probability one. Theorem 3.6.1 yields that

$$\mathsf{P}\left(\lim_{n\to\infty} B^\dagger_n = \begin{bmatrix} 1+\alpha^2 & 0 \\ 0 & 1+\alpha^2 \end{bmatrix} =: B^\dagger\right) = 1.$$

By (3.6.6), we have

$$\begin{aligned}
&\begin{bmatrix} \sqrt{n}\big(\widehat{\theta}^\dagger_{1,n} - \lim_{k\to\infty}\widehat{\theta}^\dagger_{1,k}\big) \\ \sqrt{n}\big(\widehat{\theta}^\dagger_{2,n} - \lim_{k\to\infty}\widehat{\theta}^\dagger_{2,k}\big) \end{bmatrix} \\
&\quad= \sqrt{n}(B^\dagger_n)^{-1}\left(\begin{bmatrix} (1+(\widehat{\alpha}^\dagger_n)^2)Y_{s_1} - \widehat{\alpha}^\dagger_n(Y_{s_1-1}+Y_{s_1+1}) - (1-\widehat{\alpha}^\dagger_n)\widehat{\mu}^\dagger_{\varepsilon,n} \\ (1+(\widehat{\alpha}^\dagger_n)^2)Y_{s_2} - \widehat{\alpha}^\dagger_n(Y_{s_2-1}+Y_{s_2+1}) - (1-\widehat{\alpha}^\dagger_n)\widehat{\mu}^\dagger_{\varepsilon,n} \end{bmatrix}\right. \\
&\qquad\qquad \left. - B^\dagger_n (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) \\
&\quad= \sqrt{n}(B^\dagger_n)^{-1}\left(\begin{bmatrix} (1+(\widehat{\alpha}^\dagger_n)^2)Y_{s_1} - \widehat{\alpha}^\dagger_n(Y_{s_1-1}+Y_{s_1+1}) - (1-\widehat{\alpha}^\dagger_n)\widehat{\mu}^\dagger_{\varepsilon,n} \\ (1+(\widehat{\alpha}^\dagger_n)^2)Y_{s_2} - \widehat{\alpha}^\dagger_n(Y_{s_2-1}+Y_{s_2+1}) - (1-\widehat{\alpha}^\dagger_n)\widehat{\mu}^\dagger_{\varepsilon,n} \end{bmatrix}\right. \\
&\qquad\qquad \left. - \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) \\
&\qquad + \sqrt{n}\left((B^\dagger_n)^{-1} - (B^\dagger)^{-1}\right) \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} \\
&\quad= \sqrt{n}(B^\dagger_n)^{-1} \begin{bmatrix} (\widehat{\alpha}^\dagger_n+\alpha)Y_{s_1} - (Y_{s_1-1}+Y_{s_1+1}) + \widehat{\mu}^\dagger_{\varepsilon,n} & \alpha-1 \\ (\widehat{\alpha}^\dagger_n+\alpha)Y_{s_2} - (Y_{s_2-1}+Y_{s_2+1}) + \widehat{\mu}^\dagger_{\varepsilon,n} & \alpha-1 \end{bmatrix} \begin{bmatrix} \widehat{\alpha}^\dagger_n - \alpha \\ \widehat{\mu}^\dagger_{\varepsilon,n} - \mu_\varepsilon \end{bmatrix} \\
&\qquad + \sqrt{n}(B^\dagger_n)^{-1}\big(B^\dagger - B^\dagger_n\big)(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

$$\begin{bmatrix} \sqrt{n}\big(\widehat{\theta}^{\dagger}_{1,n} - \lim_{k\to\infty} \widehat{\theta}^{\dagger}_{1,k}\big) \\ \sqrt{n}\big(\widehat{\theta}^{\dagger}_{2,n} - \lim_{k\to\infty} \widehat{\theta}^{\dagger}_{2,k}\big) \end{bmatrix} = C_{n,\alpha,\varepsilon} \begin{bmatrix} \sqrt{n}(\widehat{\alpha}^{\dagger}_n - \alpha) \\ \sqrt{n}(\widehat{\mu}^{\dagger}_{\varepsilon,n} - \mu_\varepsilon) \end{bmatrix} \tag{3.6.14}$$

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

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

By (3.6.4) and (3.6.5), we have $C_{n,\alpha,\varepsilon}$ converges almost surely as $n \to \infty$ to

$$\begin{aligned} &(B^{\dagger})^{-1} \begin{bmatrix} 2\alpha Y_{s_1} - Y_{s_1-1} - Y_{s_1+1} + \mu_\varepsilon & \alpha - 1 \\ 2\alpha Y_{s_2} - Y_{s_2-1} - Y_{s_2+1} + \mu_\varepsilon & \alpha - 1 \end{bmatrix} \\ &+ (B^{\dagger})^{-1} \begin{bmatrix} -2\alpha & 0 \\ 0 & -2\alpha \end{bmatrix} (B^{\dagger})^{-1} \begin{bmatrix} (1+\alpha^2)Y_{s_1} - \alpha(Y_{s_1-1} + Y_{s_1+1}) - (1-\alpha)\mu_\varepsilon & 0 \\ (1+\alpha^2)Y_{s_2} - \alpha(Y_{s_2-1} + Y_{s_2+1}) - (1-\alpha)\mu_\varepsilon & 0 \end{bmatrix} \\ &= \frac{1}{(1+\alpha^2)^2} \begin{bmatrix} (\alpha^2-1)(Y_{s_1-1} + Y_{s_1+1}) + (1+2\alpha-\alpha^2)\mu_\varepsilon & (\alpha-1)(1+\alpha^2) \\ (\alpha^2-1)(Y_{s_2-1} + Y_{s_2+1}) + (1+2\alpha-\alpha^2)\mu_\varepsilon & (\alpha-1)(1+\alpha^2) \end{bmatrix} \\ &= C_{\alpha,\varepsilon}. \end{aligned}$$

By (3.6.14), (3.6.10) and Slutsky's lemma, we have (3.6.11). □

3.7 Two neighbouring outliers, estimation of the mean of the offspring and innovation distributions and the outliers' sizes

In this section we assume that $I = 2$ and that the relevant time points $s_1, s_2 \in \mathbb{N}$ are known. We also suppose that $s_1 := s$ and $s_2 := s+1$, i.e., the time points s_1 and s_2 are neighbouring. We concentrate on the CLS estimation of α, μ_ε, θ_1 and θ_2.

Motivated by (3.5.1), for all $n \geqslant s+2$, $n \in \mathbb{N}$, we define the function $Q^{\dagger\dagger}_n : \mathbb{R}^{n+1} \times \mathbb{R}^4 \to \mathbb{R}$, as

$$\begin{aligned} &Q^{\dagger\dagger}_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \\ &:= \sum_{\substack{k=1 \\ k\notin\{s,s+1,s+2\}}}^{n} \big(y_k - \alpha' y_{k-1} - \mu'_\varepsilon\big)^2 + \big(y_s - \alpha' y_{s-1} - \mu'_\varepsilon - \theta'_1\big)^2 + \big(y_{s+1} - \alpha' y_s - \mu'_\varepsilon + \alpha'\theta'_1 - \theta'_2\big)^2 \\ &\quad + \big(y_{s+2} - \alpha' y_{s+1} - \mu'_\varepsilon + \alpha'\theta'_2\big)^2, \qquad \mathbf{y}_n \in \mathbb{R}^{n+1},\ \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2 \in \mathbb{R}. \end{aligned}$$

By definition, for all $n \geqslant s+2$, 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 $(\widehat{\alpha}_n^{\dagger\dagger}, \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}, \widehat{\theta}_{1,n}^{\dagger\dagger}, \widehat{\theta}_{2,n}^{\dagger\dagger}) : S_n \to \mathbb{R}^4$ such that

$$
\begin{aligned}
Q_n^{\dagger\dagger}(\mathbf{y}_n;\, &\widehat{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n), \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}(\mathbf{y}_n), \widehat{\theta}_{1,n}^{\dagger\dagger}(\mathbf{y}_n), \widehat{\theta}_{2,n}^{\dagger\dagger}(\mathbf{y}_n)) \\
&= \inf_{(\alpha', \mu_\varepsilon', \theta_1', \theta_2') \in \mathbb{R}^4} Q_n^{\dagger\dagger}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') \qquad \forall\ \mathbf{y}_n \in S_n,
\end{aligned}
$$

where S_n is suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 3.7.1). We note that we do not define the CLS estimator $(\widehat{\alpha}_n^{\dagger\dagger}, \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}, \widehat{\theta}_{1,n}^{\dagger\dagger}, \widehat{\theta}_{2,n}^{\dagger\dagger})$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$. For all $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $(\alpha', \mu_\varepsilon', \theta_1', \theta_2') \in \mathbb{R}^4$,

$$
\begin{aligned}
&\frac{\partial Q_n^{\dagger\dagger}}{\partial \alpha'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') \\
&\quad = \sum_{\substack{k=1 \\ k \notin \{s, s+1, s+2\}}}^{n} \big(y_k - \alpha' y_{k-1} - \mu_\varepsilon'\big)(-2y_{k-1}) - 2\big(y_s - \alpha' y_{s-1} - \mu_\varepsilon' - \theta_1'\big) y_{s-1} \\
&\qquad + 2\big(y_{s+1} - \alpha' y_s - \mu_\varepsilon' + \alpha'\theta_1' - \theta_2'\big)(-y_s + \theta_1') + 2\big(y_{s+2} - \alpha' y_{s+1} - \mu_\varepsilon' + \alpha'\theta_2'\big)(-y_{s+1} + \theta_2'),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\partial Q_n^{\dagger\dagger}}{\partial \mu_\varepsilon'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') \\
&\quad = \sum_{\substack{k=1 \\ k \notin \{s, s+1, s+2\}}}^{n} (-2)\big(y_k - \alpha' y_{k-1} - \mu_\varepsilon'\big) - 2\big(y_s - \alpha' y_{s-1} - \mu_\varepsilon' - \theta_1'\big) \\
&\qquad - 2\big(y_{s+1} - \alpha' y_s - \mu_\varepsilon' + \alpha'\theta_1' - \theta_2'\big) - 2\big(y_{s+2} - \alpha' y_{s+1} - \mu_\varepsilon' + \alpha'\theta_2'\big), \\
&\frac{\partial Q_n^{\dagger\dagger}}{\partial \theta_1'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') \\
&\quad = -2(y_s - \alpha' y_{s-1} - \mu_\varepsilon' - \theta_1') + 2\alpha'(y_{s+1} - \alpha' y_s - \mu_\varepsilon' + \alpha'\theta_1' - \theta_2'), \\
&\frac{\partial Q_n^{\dagger\dagger}}{\partial \theta_2'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') \\
&\quad = -2(y_{s+1} - \alpha' y_s - \mu_\varepsilon' + \alpha'\theta_1' - \theta_2') + 2\alpha'(y_{s+2} - \alpha' y_{s+1} - \mu_\varepsilon' + \alpha'\theta_2').
\end{aligned}
$$

The next lemma is about the existence and uniqueness of the CLS estimator of $(\alpha, \mu_\varepsilon, \theta_1, \theta_2)$.

3.7.1 Lemma. *There exist subsets $S_n \subset \mathbb{R}^{n+1}$, $n \geqslant \max(3, s+2)$ with the following properties:*

(i) *there exists a unique CLS estimator $(\widehat{\alpha}_n^{\dagger\dagger}, \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}, \widehat{\theta}_{1,n}^{\dagger\dagger}, \widehat{\theta}_{2,n}^{\dagger\dagger}) : S_n \to \mathbb{R}^4$,*

(ii) *for all $\mathbf{y}_n \in S_n$, $(\widehat{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n), \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}(\mathbf{y}_n), \widehat{\theta}_{1,n}^{\dagger\dagger}(\mathbf{y}_n), \widehat{\theta}_{2,n}^{\dagger\dagger}(\mathbf{y}_n))$ is the unique solution of the system of equations*

$$
\begin{aligned}
&\frac{\partial Q_n^{\dagger\dagger}}{\partial \alpha'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 0, \qquad \frac{\partial Q_n^{\dagger\dagger}}{\partial \mu_\varepsilon'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 0, \\
&\frac{\partial Q_n^{\dagger\dagger}}{\partial \theta_1'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 0, \qquad \frac{\partial Q_n^{\dagger\dagger}}{\partial \theta_2'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta_1', \theta_2') = 0,
\end{aligned}
\tag{3.7.1}
$$

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

Proof. For any fixed $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $n \geqslant \max(3, s+2)$ and $\alpha' \in \mathbb{R}$, the quadratic function $\mathbb{R}^3 \ni (\mu'_\varepsilon, \theta'_1, \theta'_2) \mapsto Q_n^{\dagger\dagger}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$ can be written in the form

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

where

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

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

and the matrix

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

is strictly positive definite for all $n \geqslant 3$ and $\alpha' \in \mathbb{R}$. Indeed, the leading principal minors of $A_n(\alpha')$ take the following forms: n,

$$
\begin{aligned}
&n(1 + (\alpha')^2) - (1 - \alpha')^2 = (n-1)(\alpha')^2 + 2\alpha' + n - 1, \\
&D_n(\alpha') := n(1 + (\alpha')^2)^2 - 2(1 - \alpha')^2(1 + (\alpha')^2) - 2\alpha'(1 - \alpha')^2 - n(\alpha')^2 \\
&\qquad = n(1 + \alpha' + (\alpha')^2)(1 - \alpha' + (\alpha')^2) - 2(1 - \alpha')^2(1 + \alpha' + (\alpha')^2) \\
&\qquad = (1 + \alpha' + (\alpha')^2)\big((n-2)(\alpha')^2 - (n-4)\alpha' + n - 2\big),
\end{aligned}
$$

and for all $n \geqslant 3$, the discriminant $(n-4)^2 - 4(n-2)^2 = -3n^2 + 8n$ of the equation $(n-2)x^2 - (n-4)x + n - 2 = 0$ is negative. The inverse matrix $A_n(\alpha')^{-1}$ takes the form

$$
\frac{1}{D_n(\alpha')} \begin{bmatrix} 1 + (\alpha')^2 + (\alpha')^4 & -(1 - \alpha')(1 + \alpha' + (\alpha')^2) & -(1 - \alpha')(1 + \alpha' + (\alpha')^2) \\ -(1 - \alpha')(1 + \alpha' + (\alpha')^2) & n(1 + (\alpha')^2) - (1 - \alpha')^2 & (1 - \alpha')^2 + n\alpha' \\ -(1 - \alpha')(1 + \alpha' + (\alpha')^2) & (1 - \alpha')^2 + n\alpha' & n(1 + (\alpha')^2) - (1 - \alpha')^2 \end{bmatrix}.
$$

The polynomial $\mathbb{R} \ni \alpha' \mapsto D_n(\alpha')$ is of order 4 with leading coefficient $n - 2$. We have $\widehat{Q}_n^{\dagger\dagger}(\mathbf{y}_n; \alpha') = R_n(\mathbf{y}_n; \alpha') / D_n(\alpha')$, where $\mathbb{R} \ni \alpha' \mapsto R_n(\mathbf{y}_n; \alpha')$ is a polynomial of order 6 with leading coefficient

$$c_n(\mathbf{y}_n) := (n-2)\sum_{k=1}^n y_{k-1}^2 - \left(\sum_{k=1}^n y_{k-1}\right)^2 - (n-1)(y_s^2 + y_{s+1}^2)$$

$$+ 2(y_s + y_{s+1})\sum_{k=1}^n y_{k-1} - 2y_s y_{s+1}.$$

Let

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

For $\mathbf{y}_n \in \widehat{S}_n^{\dagger\dagger}$, we have $\lim_{|\alpha'|\to\infty} \widehat{Q}_n^{\dagger\dagger}(\mathbf{y}_n;\alpha') = \infty$ and the continuous function $\mathbb{R} \ni \alpha' \mapsto \widehat{Q}_n^{\dagger\dagger}(\mathbf{y}_n;\alpha')$ attains its infimum. Consequently, for all $n \geqslant \max(3, s+2)$ there exists a CLS estimator $(\widehat{\alpha}_n^{\dagger\dagger}, \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}, \widehat{\theta}_{1,n}^{\dagger\dagger}, \widehat{\theta}_{2,n}^{\dagger\dagger}) : \widehat{S}_n^{\dagger\dagger} \to \mathbb{R}^4$, where

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

$$\begin{bmatrix} \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}(\mathbf{y}_n) \\ \widehat{\theta}_{1,n}^{\dagger\dagger}(\mathbf{y}_n) \\ \widehat{\theta}_{2,n}^{\dagger\dagger}(\mathbf{y}_n) \end{bmatrix} = A_n(\widehat{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n))^{-1} t_n(\mathbf{y}_n;\widehat{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n)), \qquad \mathbf{y}_n \in \widehat{S}_n^{\dagger\dagger}, \tag{3.7.2}$$

and for all $\mathbf{y}_n \in \widehat{S}_n^{\dagger\dagger}$, $(\widehat{\alpha}_n^{\dagger\dagger}(\mathbf{y}_n), \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}(\mathbf{y}_n), \widehat{\theta}_{1,n}^{\dagger\dagger}(\mathbf{y}_n), \widehat{\theta}_{2,n}^{\dagger\dagger}(\mathbf{y}_n))$ is a solution of the system of equations (3.7.1).

By (2.2.5) and (2.2.6), we get $\mathsf{P}\left(\lim_{n\to\infty} n^{-2} c_n(\mathbf{Y}_n) = \operatorname{Var}\widetilde{X}\right) = 1$, where $\widetilde{X}$ denotes a random variable with the unique stationary distribution of the INAR(1) model in (2.1.1). Hence $\mathbf{Y}_n \in \widehat{S}_n^{\dagger\dagger}$ holds asymptotically as $n \to \infty$ with probability one.

Now we turn to find sets $S_n \subset \widehat{S}_n^{\dagger\dagger}$, $n \geqslant \max(3, s+2)$ such that the system of equations (3.7.1) has a unique solution with respect to $(\alpha', \mu_\varepsilon', \theta_1', \theta_2')$ for all $\mathbf{y}_n \in S_n$. Let us introduce the (4×4) Hessian matrix

$$H_n(\mathbf{y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') := \begin{bmatrix} \frac{\partial^2 Q_n^\dagger}{\partial(\alpha')^2} & \frac{\partial^2 Q_n^\dagger}{\partial\mu_\varepsilon'\partial\alpha'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\alpha'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\alpha'} \\ \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\mu_\varepsilon'} & \frac{\partial^2 Q_n^\dagger}{\partial(\mu_\varepsilon')^2} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\mu_\varepsilon'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\mu_\varepsilon'} \\ \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\theta_1'} & \frac{\partial^2 Q_n^\dagger}{\partial\mu_\varepsilon'\partial\theta_1'} & \frac{\partial^2 Q_n^\dagger}{\partial(\theta_1')^2} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_2'\partial\theta_1'} \\ \frac{\partial^2 Q_n^\dagger}{\partial\alpha'\partial\theta_2'} & \frac{\partial^2 Q_n^\dagger}{\partial\mu_\varepsilon'\partial\theta_2'} & \frac{\partial^2 Q_n^\dagger}{\partial\theta_1'\partial\theta_2'} & \frac{\partial^2 Q_n^\dagger}{\partial(\theta_2')^2} \end{bmatrix} (\mathbf{y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2'),$$

and let us denote by $\Delta_{i,n}(\mathbf{y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2')$ its i-th order leading principal minor, $i = 1,2,3,4$. Further, for all $n \geqslant \max(3, s+2)$, let

$$S_n := \left\{\mathbf{y}_n \in \widehat{S}_n^{\dagger\dagger} : \Delta_{i,n}(\mathbf{y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') > 0,\ i = 1,2,3,4, \forall\,(\alpha',\mu_\varepsilon',\theta_1',\theta_2') \in \mathbb{R}^4\right\}.$$

By Berkovitz [10, Theorem 3.3, Chapter III], the function $\mathbb{R}^4 \ni (\alpha',\mu_\varepsilon',\theta_1',\theta_2') \mapsto Q_n^{\dagger\dagger}(\mathbf{y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2')$ is strictly convex for all $\mathbf{y}_n \in S_n$. Since it was already proved that the system of equations (3.7.1) has a solution for all $\mathbf{y}_n \in \widehat{S}_n^{\dagger\dagger}$, we obtain that this solution is unique for all $\mathbf{y}_n \in S_n$.

Next we check that $\mathbf{Y}_n \in S_n$ holds asymptotically as $n \to \infty$ with probability one. For all $(\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \in \mathbb{R}^4$,

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\alpha')^2}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$$

$$= 2 \sum_{\substack{k=1 \\ k \notin \{s, s+1, s+2\}}}^{n} Y_{k-1}^2 + 2Y_{s-1}^2 + 2(Y_s - \theta'_1)^2 + 2(Y_{s+1} - \theta'_2)^2$$

$$= 2 \sum_{\substack{k=1 \\ k \notin \{s+1, s+2\}}}^{n} X_{k-1}^2 + 2(X_s + \theta_1 - \theta'_1)^2 + 2(X_{s+1} + \theta_2 - \theta'_2)^2,$$

and

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\mu'_\varepsilon \partial\alpha'}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\alpha' \partial\mu'_\varepsilon}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$$

$$= 2\sum_{k=1}^{n} Y_{k-1} - 2\theta'_1 - 2\theta'_2 = 2\sum_{k=1}^{n} X_{k-1} + 2(\theta_1 - \theta'_1) + 2(\theta_2 - \theta'_2),$$

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta'_1 \partial\alpha'}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\alpha' \partial\theta'_1}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$$

$$= 2(Y_{s-1} + Y_{s+1} - 2\alpha' Y_s - \mu'_\varepsilon + 2\alpha'\theta'_1 - \theta'_2)$$

$$= 2(X_{s-1} + X_{s+1} - 2\alpha' X_s - \mu'_\varepsilon - 2\alpha'(\theta_1 - \theta'_1) + (\theta_2 - \theta'_2)),$$

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta'_2 \partial\alpha'}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\alpha' \partial\theta'_2}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$$

$$= 2(Y_s + Y_{s+2} - 2\alpha' Y_{s+1} - \mu'_\varepsilon - \theta'_1 + 2\alpha'\theta'_2)$$

$$= 2(X_s + X_{s+2} - 2\alpha' X_{s+1} - \mu'_\varepsilon + (\theta_1 - \theta'_1) - 2\alpha'(\theta_2 - \theta'_2)),$$

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\mu'_\varepsilon)^2}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 2n,$$

and

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\theta'_1)^2}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial(\theta'_2)^2}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 2((\alpha')^2 + 1),$$

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta'_1 \partial\theta'_2}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta'_2 \partial\theta'_1}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = -2\alpha',$$

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta'_1 \partial\mu'_\varepsilon}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\mu'_\varepsilon \partial\theta'_1}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 2(1 - \alpha'),$$

$$\frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\theta'_2 \partial\mu'_\varepsilon}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = \frac{\partial^2 Q_n^{\dagger\dagger}}{\partial\mu'_\varepsilon \partial\theta'_2}(\mathbf{Y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 2(1 - \alpha').$$

Then $H_n(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2')$ has the following leading principal minors

$$\Delta_{1,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') = 2\sum_{\substack{k=1\\ k\notin\{s+1,s+2\}}}^{n} X_{k-1}^2 + 2(X_s+\theta_1-\theta_1')^2 + 2(X_{s+1}+\theta_2-\theta_2')^2,$$

$$\begin{aligned}\Delta_{2,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') &= 4n\left(\sum_{\substack{k=1\\ k\notin\{s+1,s+2\}}}^{n} X_{k-1}^2 + (X_s+\theta_1-\theta_1')^2 + (X_{s+1}+\theta_2-\theta_2')^2\right)\\ &\quad - 4\left(\sum_{k=1}^{n} X_{k-1} + (\theta_1-\theta_1') + (\theta_2-\theta_2')\right)^2,\end{aligned}$$

and

$$\begin{aligned}&\Delta_{3,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2')\\ &\quad = 8\big(((\alpha')^2+1)n-(1-\alpha')^2\big)\left(\sum_{\substack{k=1\\ k\notin\{s+1,s+2\}}}^{n} X_{k-1}^2 + (X_s+\theta_1-\theta_1')^2 + (X_{s+1}+\theta_2-\theta_2')^2\right)\\ &\qquad + 16L\left(\sum_{k=1}^{n} X_{k-1} + (\theta_1-\theta_1') + (\theta_2-\theta_2')\right) - 8nL^2\\ &\qquad - 8((\alpha')^2+1)\left(\sum_{k=1}^{n} X_{k-1} + (\theta_1-\theta_1') + (\theta_2-\theta_2')\right)^2,\end{aligned}$$

$$\Delta_{4,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') = \det H_n(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2'),$$

where $L := X_{s-1}+X_{s+1}-2\alpha' X_s-\mu_\varepsilon'-2\alpha'(\theta_1-\theta_1')+\theta_2-\theta_2'$. By (2.2.5) and (2.2.6), we get

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\Delta_{1,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') = 2\mathsf{E}\widetilde{X}^2,\ \ \forall\ (\alpha',\mu_\varepsilon',\theta_1',\theta_2')\in\mathbb{R}^4\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n^2}\Delta_{2,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') = 4(\mathsf{E}\widetilde{X}^2-(\mathsf{E}\widetilde{X})^2) = 4\operatorname{Var}\widetilde{X},\ \ \forall\ (\alpha',\mu_\varepsilon',\theta_1',\theta_2')\in\mathbb{R}^4\right) = 1,$$

and

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n^2}\Delta_{3,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') = 8((\alpha')^2+1)\operatorname{Var}\widetilde{X},\ \ \forall\ (\alpha',\mu_\varepsilon',\theta_1',\theta_2')\in\mathbb{R}^4\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n^2}\Delta_{4,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') = 16((\alpha')^4+(\alpha')^2+1)\operatorname{Var}\widetilde{X},\ \ \forall\ (\alpha',\mu_\varepsilon',\theta_1',\theta_2')\in\mathbb{R}^4\right) = 1,$$

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

$$\mathsf{P}\big(\lim_{n\to\infty}\Delta_{i,n}(\mathbf{Y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2') = \infty,\ \ \forall\ (\alpha',\mu_\varepsilon',\theta_1',\theta_2')\in\mathbb{R}^4\big) = 1, \qquad i=1,2,3,4,$$

which yields that $\mathbf{Y}_n \in S_n$ asymptotically as $n \to \infty$ with probability one, since we have already proved that $\mathbf{Y}_n \in \widehat{S}_n^{\dagger\dagger}$ asymptotically as $n \to \infty$ with probability one. $\Box$

By Lemma 3.7.1, $(\widehat{\alpha}_n^{\dagger\dagger}(\mathbf{Y}_n), \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}(\mathbf{Y}_n), \widehat{\theta}_{1,n}^{\dagger\dagger}(\mathbf{Y}_n), \widehat{\theta}_{2,n}^{\dagger\dagger}(\mathbf{Y}_n))$ exists uniquely asymptotically as $n \to \infty$ with probability one. In the sequel we will simply denote it by $(\widehat{\alpha}_n^{\dagger\dagger}, \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}, \widehat{\theta}_{1,n}^{\dagger\dagger}, \widehat{\theta}_{2,n}^{\dagger\dagger})$.

The next result shows that $\widehat{\alpha}_n^{\dagger\dagger}$ is a strongly consistent estimator of α, $\widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}$ is a strongly consistent estimator of μ_ε, whereas $\widehat{\theta}_{1,n}^{\dagger\dagger}$ and $\widehat{\theta}_{2,n}^{\dagger\dagger}$ fail to be strongly consistent estimators of θ_1 and θ_2, respectively.

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

$$\mathsf{P}(\lim_{n\to\infty} \widehat{\alpha}_n^{\dagger\dagger} = \alpha) = 1, \qquad \forall\, (\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0,1) \times (0,\infty) \times \mathbb{N}^2, \tag{3.7.3}$$

$$\mathsf{P}(\lim_{n\to\infty} \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} = \mu_\varepsilon) = 1, \qquad \forall\, (\alpha, \mu_\varepsilon, \theta_1, \theta_2) \in (0,1) \times (0,\infty) \times \mathbb{N}^2, \tag{3.7.4}$$

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

$$\mathsf{P}\left(\lim_{n\to\infty} \begin{bmatrix} \widetilde{\theta}_{1,n}^{\dagger\dagger} \\ \widetilde{\theta}_{2,n}^{\dagger\dagger} \end{bmatrix} = \begin{bmatrix} Y_s \\ Y_{s+1} \end{bmatrix} + \begin{bmatrix} \frac{-\alpha(1+\alpha^2)Y_{s-1}-\alpha^2 Y_{s+2}-(1-\alpha^3)\mu_\varepsilon}{1+\alpha^2+\alpha^4} \\ \frac{-\alpha^2 Y_{s-1}-\alpha(1+\alpha^2)Y_{s+2}-(1-\alpha^3)\mu_\varepsilon}{1+\alpha^2+\alpha^4} \end{bmatrix}\right) = 1 \tag{3.7.5}$$

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

Proof. An easy calculation shows that

$$\left(\sum_{\substack{k=1\\ k\notin\{s+1,s+2\}}}^{n} Y_{k-1}^2 + (Y_s - \widehat{\theta}_{1,n}^{\dagger\dagger})^2 + (Y_{s+1} - \widehat{\theta}_{2,n}^{\dagger\dagger})^2\right) \widehat{\alpha}_n^{\dagger\dagger} + \left(\sum_{k=1}^n Y_{k-1} - \widehat{\theta}_{1,n}^{\dagger\dagger} - \widehat{\theta}_{2,n}^{\dagger\dagger}\right) \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}$$

$$= \sum_{k=1}^n Y_{k-1}Y_k - \widehat{\theta}_{1,n}^{\dagger\dagger}(Y_{s-1} + Y_{s+1}) - \widehat{\theta}_{2,n}^{\dagger\dagger}(Y_s + Y_{s+2}) + \widehat{\theta}_{1,n}^{\dagger\dagger}\widehat{\theta}_{2,n}^{\dagger\dagger},$$

$$\left(\sum_{k=1}^n Y_{k-1} - \widehat{\theta}_{1,n}^{\dagger\dagger} - \widehat{\theta}_{2,n}^{\dagger\dagger}\right) \widehat{\alpha}_n^{\dagger\dagger} + n\widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} = \sum_{k=1}^n Y_k - \widehat{\theta}_{1,n}^{\dagger\dagger} - \widehat{\theta}_{2,n}^{\dagger\dagger},$$

hold asymptotically as $n \to \infty$ with probability one, and hence

$$\begin{bmatrix} \widehat{\alpha}_n^{\dagger\dagger} \\ \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} \end{bmatrix} = \begin{bmatrix} a_n & b_n \\ b_n & n \end{bmatrix}^{-1} \begin{bmatrix} k_n \\ \ell_n \end{bmatrix}, \tag{3.7.6}$$

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

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

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

and

$$k_n := \sum_{k=1}^{n} X_{k-1}X_k + (\theta_1 - \widehat{\theta}_{1,n}^{\dagger\dagger})(X_{s-1} + X_{s+1}) + (\theta_2 - \widehat{\theta}_{2,n}^{\dagger\dagger})(X_s + X_{s+2}) + (\theta_1 - \widehat{\theta}_{1,n}^{\dagger\dagger})(\theta_2 - \widehat{\theta}_{2,n}^{\dagger\dagger}),$$

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

Furthermore,

$$\begin{bmatrix} \widehat{\alpha}_n^{\dagger\dagger} - \alpha \\ \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} - \mu_\varepsilon \end{bmatrix} = \begin{bmatrix} a_n & b_n \\ b_n & n \end{bmatrix}^{-1} \begin{bmatrix} c_n \\ d_n \end{bmatrix} \tag{3.7.7}$$

hold asymptotically as $n \to \infty$ with probability one, where

$$\begin{aligned} c_n := \sum_{k=1}^{n} X_{k-1}(X_k - \alpha X_{k-1} - \mu_\varepsilon) &+ (\theta_1 - \widehat{\theta}_{1,n}^{\dagger\dagger})\big(X_{s-1} + X_{s+1} - 2\alpha X_s - \mu_\varepsilon - \alpha(\theta_1 - \widehat{\theta}_{1,n}^{\dagger\dagger})\big) \\ &+ (\theta_2 - \widehat{\theta}_{2,n}^{\dagger\dagger})\big(X_s + X_{s+2} - 2\alpha X_{s+1} - \mu_\varepsilon - \alpha(\theta_2 - \widehat{\theta}_{2,n}^{\dagger\dagger})\big) + (\theta_1 - \widehat{\theta}_{1,n}^{\dagger\dagger})(\theta_2 - \widehat{\theta}_{2,n}^{\dagger\dagger}), \\ d_n := \sum_{k=1}^{n}(X_k - \alpha X_{k-1} - \mu_\varepsilon) &+ (1-\alpha)(\theta_1 - \widehat{\theta}_{1,n}^{\dagger\dagger} + \theta_2 - \widehat{\theta}_{2,n}^{\dagger\dagger}). \end{aligned}$$

We show that the sequences $(\widehat{\theta}_{1,n}^{\dagger\dagger} - \theta_1)_{n\in\mathbb{N}}$ and $(\widehat{\theta}_{2,n}^{\dagger\dagger} - \theta_2)_{n\in\mathbb{N}}$ are bounded with probability one. An easy calculation shows that

$$\begin{aligned} n\widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} + (1 - \widehat{\alpha}_n^{\dagger\dagger})\widehat{\theta}_{1,n}^{\dagger\dagger} + (1 - \widehat{\alpha}_n^{\dagger\dagger})\widehat{\theta}_{2,n}^{\dagger\dagger} &= \sum_{k=1}^{n}(Y_k - \widehat{\alpha}_n^{\dagger\dagger} Y_{k-1}), \\ (1 - \widehat{\alpha}_n^{\dagger\dagger})\widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} + (1 + (\widehat{\alpha}_n^{\dagger\dagger})^2)\widehat{\theta}_{1,n}^{\dagger\dagger} - \widehat{\alpha}_n^{\dagger\dagger}\widehat{\theta}_{2,n}^{\dagger\dagger} &= (1 + (\widehat{\alpha}_n^{\dagger\dagger})^2)Y_s - \widehat{\alpha}_n^{\dagger\dagger}(Y_{s-1} + Y_{s+1}), \\ (1 - \widehat{\alpha}_n^{\dagger\dagger})\widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} - \widehat{\alpha}_n^{\dagger\dagger}\widehat{\theta}_{1,n}^{\dagger\dagger} + (1 + (\widehat{\alpha}_n^{\dagger\dagger})^2)\widehat{\theta}_{2,n}^{\dagger\dagger} &= (1 + (\widehat{\alpha}_n^{\dagger\dagger})^2)Y_{s+1} - \widehat{\alpha}_n^{\dagger\dagger}(Y_s + Y_{s+2}), \end{aligned}$$

hold asymptotically as $n \to \infty$ with probability one, or equivalently

$$\begin{bmatrix} n & 1 - \widehat{\alpha}_n^{\dagger\dagger} & 1 - \widehat{\alpha}_n^{\dagger\dagger} \\ 1 - \widehat{\alpha}_n^{\dagger\dagger} & 1 + (\widehat{\alpha}_n^{\dagger\dagger})^2 & -\widehat{\alpha}_n^{\dagger\dagger} \\ 1 - \widehat{\alpha}_n^{\dagger\dagger} & -\widehat{\alpha}_n^{\dagger\dagger} & 1 + (\widehat{\alpha}_n^{\dagger\dagger})^2 \end{bmatrix} \begin{bmatrix} \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} \\ \widehat{\theta}_{1,n}^{\dagger\dagger} \\ \widehat{\theta}_{2,n}^{\dagger\dagger} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n}(Y_k - \widehat{\alpha}_n^{\dagger\dagger} Y_{k-1}) \\ (1 + (\widehat{\alpha}_n^{\dagger\dagger})^2)Y_s - \widehat{\alpha}_n^{\dagger\dagger}(Y_{s-1} + Y_{s+1}) \\ (1 + (\widehat{\alpha}_n^{\dagger\dagger})^2)Y_{s+1} - \widehat{\alpha}_n^{\dagger\dagger}(Y_s + Y_{s+2}) \end{bmatrix} \tag{3.7.8}$$

holds asymptotically as $n \to \infty$ with probability one. Since for all $n \geqslant 3$

$$D_n(\widehat{\alpha}_n^{\dagger\dagger}) = (1 + \widehat{\alpha}_n^{\dagger\dagger} + (\widehat{\alpha}_n^{\dagger\dagger})^2)\big((n-2)(\widehat{\alpha}_n^{\dagger\dagger})^2 - (n-4)\widehat{\alpha}_n^{\dagger\dagger} + n - 2\big) > 0,$$

we get asymptotically as $n \to \infty$ with probability one we have

(3.7.9)

$$\begin{bmatrix} \widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} \\ \widehat{\theta}^{\dagger\dagger}_{1,n} \\ \widehat{\theta}^{\dagger\dagger}_{2,n} \end{bmatrix} = \frac{1}{D_n(\widehat{\alpha}^{\dagger\dagger}_n)} \begin{bmatrix} 1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2 + (\widehat{\alpha}^{\dagger\dagger}_n)^4 & u_n & u_n \\ u_n & w_n & v_n \\ u_n & v_n & w_n \end{bmatrix} \begin{bmatrix} \sum_{k=1}^n (Y_k - \widehat{\alpha}^{\dagger\dagger}_n Y_{k-1}) \\ (1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2) Y_s - \widehat{\alpha}^{\dagger\dagger}_n (Y_{s-1} + Y_{s+1}) \\ (1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2) Y_{s+1} - \widehat{\alpha}^{\dagger\dagger}_n (Y_s + Y_{s+2}) \end{bmatrix}$$

$$=: \frac{1}{D_n(\widehat{\alpha}^{\dagger\dagger}_n)} \begin{bmatrix} G_n \\ H_n \\ J_n \end{bmatrix},$$

where

$$\begin{aligned} u_n &:= -(1 - \widehat{\alpha}^{\dagger\dagger}_n)(1 + \widehat{\alpha}^{\dagger\dagger}_n + (\widehat{\alpha}^{\dagger\dagger}_n)^2), \\ v_n &:= (1 - \widehat{\alpha}^{\dagger\dagger}_n)^2 + n\widehat{\alpha}^{\dagger\dagger}_n, \\ w_n &:= n(1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2) - (1 - \widehat{\alpha}^{\dagger\dagger}_n)^2, \end{aligned}$$

and

$$\begin{aligned} G_n &:= -(1 - \widehat{\alpha}^{\dagger\dagger}_n)(1 + \widehat{\alpha}^{\dagger\dagger}_n + (\widehat{\alpha}^{\dagger\dagger}_n)^2)\Big((1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2)(Y_s + Y_{s+1}) - \widehat{\alpha}^{\dagger\dagger}_n (Y_{s-1} + Y_{s+1} + Y_s + Y_{s+2})\Big) \\ &\quad + (1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2 + (\widehat{\alpha}^{\dagger\dagger}_n)^4) \sum_{k=1}^n (Y_k - \widehat{\alpha}^{\dagger\dagger}_n Y_{k-1}), \\ H_n &:= \big(n(1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2) - (1 - \widehat{\alpha}^{\dagger\dagger}_n)^2\big)\Big((1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2) Y_s - \widehat{\alpha}^{\dagger\dagger}_n (Y_{s-1} + Y_{s+1})\Big) \\ &\quad + ((1 - \widehat{\alpha}^{\dagger\dagger}_n)^2 + n\widehat{\alpha}^{\dagger\dagger}_n)\Big((1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2) Y_{s+1} - \widehat{\alpha}^{\dagger\dagger}_n (Y_s + Y_{s+2})\Big) \\ &\quad - (1 - \widehat{\alpha}^{\dagger\dagger}_n)(1 + \widehat{\alpha}^{\dagger\dagger}_n + (\widehat{\alpha}^{\dagger\dagger}_n)^2) \sum_{k=1}^n (Y_k - \widehat{\alpha}^{\dagger\dagger}_n Y_{k-1}), \\ J_n &:= \big((1 - \widehat{\alpha}^{\dagger\dagger}_n)^2 + n\widehat{\alpha}^{\dagger\dagger}_n\big)\Big((1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2) Y_s - \widehat{\alpha}^{\dagger\dagger}_n (Y_{s-1} + Y_{s+1})\Big) \\ &\quad + \big(n(1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2) - (1 - \widehat{\alpha}^{\dagger\dagger}_n)^2\big)\Big((1 + (\widehat{\alpha}^{\dagger\dagger}_n)^2) Y_{s+1} - \widehat{\alpha}^{\dagger\dagger}_n (Y_s + Y_{s+2})\Big) \\ &\quad - (1 - \widehat{\alpha}^{\dagger\dagger}_n)(1 + \widehat{\alpha}^{\dagger\dagger}_n + (\widehat{\alpha}^{\dagger\dagger}_n)^2) \sum_{k=1}^n (Y_k - \widehat{\alpha}^{\dagger\dagger}_n Y_{k-1}). \end{aligned}$$

Using (2.2.5) and that for all $p_i \in \mathbb{R}$, $i = 0, \ldots, 4$,

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

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

By the same arguments as in the proof of Theorem 3.6.1, one can derive (3.7.3) and (3.7.4). Indeed, using (2.2.3), (2.2.5), (2.2.6) and (2.2.7), we get

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{a_n}{n}=\mathsf{E}\widetilde{X}^2\right)=1,\qquad \mathsf{P}\left(\lim_{n\to\infty}\frac{b_n}{n}=\mathsf{E}\widetilde{X}\right)=1,$$

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

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{c_n}{n}=\alpha\mathsf{E}\widetilde{X}^2+\mu_\varepsilon\mathsf{E}\widetilde{X}-\alpha\mathsf{E}\widetilde{X}^2-\mu_\varepsilon\mathsf{E}\widetilde{X}=0\right)=1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{d_n}{n}=\mathsf{E}\widetilde{X}-\alpha\mathsf{E}\widetilde{X}-\mu_\varepsilon=0\right)=1.$$

Hence, by (3.7.7), we obtain

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

which yields (3.7.3) and (3.7.4). Then (3.7.3), (3.7.4) and (3.7.9) imply (3.7.5). Indeed,

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{D_n(\widehat{\alpha}_n^{\dagger\dagger})}{n}=1+\alpha^2+\alpha^4\right)=1,\qquad \forall\ (\alpha,\mu_\varepsilon,\theta_1,\theta_2)\in(0,1)\times(0,\infty)\times\mathbb{N}^2,$$

and $\frac{H_n}{n}$ converges almost surely as $n\to\infty$ to

$$\begin{aligned}&(1+\alpha^2)\Big((1+\alpha^2)Y_s-\alpha(Y_{s-1}+Y_{s+1})\Big)+\alpha\Big((1+\alpha^2)Y_{s+1}-\alpha(Y_s+Y_{s+2})\Big)\\ &\qquad-(1-\alpha)(1+\alpha+\alpha^2)(1-\alpha)\mathsf{E}\widetilde{X}\\ &=-\alpha(1+\alpha^2)Y_{s-1}+(1+\alpha^2+\alpha^4)Y_s-\alpha^2Y_{s+2}-(1-\alpha^3)\mu_\varepsilon,\end{aligned}$$

and $\frac{J_n}{n}$ converges almost surely as $n\to\infty$ to

$$\begin{aligned}&\alpha\Big((1+\alpha^2)Y_s-\alpha(Y_{s-1}+Y_{s+1})\Big)+(1+\alpha^2)\Big((1+\alpha^2)Y_{s+1}-\alpha(Y_s+Y_{s+2})\Big)\\ &\qquad-(1-\alpha)(1+\alpha+\alpha^2)(1-\alpha)\mathsf{E}\widetilde{X}\\ &=-\alpha^2Y_{s-1}+(1+\alpha^2+\alpha^4)Y_{s+1}-\alpha(1+\alpha^2)Y_{s+2}-(1-\alpha^3)\mu_\varepsilon.\end{aligned}$$

□

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

$$\begin{bmatrix}\sqrt{n}(\widehat{\alpha}_n^{\dagger\dagger}-\alpha)\\ \sqrt{n}(\widehat{\mu}_{\varepsilon,n}^{\dagger\dagger}-\mu_\varepsilon)\end{bmatrix}\xrightarrow{\mathcal{L}}\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},B_{\alpha,\varepsilon}\right)\qquad \text{as } n\to\infty, \tag{3.7.10}$$

where $B_{\alpha,\varepsilon}$ *is defined in* (2.3.2). *Moreover, conditionally on the values* Y_{s-1} *and* Y_{s+2},

$$\begin{bmatrix}\sqrt{n}\big(\widehat{\theta}_{1,n}^{\dagger\dagger}-\lim_{k\to\infty}\widehat{\theta}_{1,k}^{\dagger\dagger}\big)\\ \sqrt{n}\big(\widehat{\theta}_{2,n}^{\dagger\dagger}-\lim_{k\to\infty}\widehat{\theta}_{2,k}^{\dagger\dagger}\big)\end{bmatrix}\xrightarrow{\mathcal{L}}\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},D_{\alpha,\varepsilon}B_{\alpha,\varepsilon}D_{\alpha,\varepsilon}^\top\right)\qquad \text{as } n\to\infty, \tag{3.7.11}$$

where the (2×2)*-matrix* $D_{\alpha,\varepsilon}$ *is defined by*

$$D_{\alpha,\varepsilon} := \begin{bmatrix} \frac{(\alpha^2-1)(\alpha^4+3\alpha^2+1)Y_{s-1}+2\alpha(\alpha^4-1)Y_{s+2}+\alpha(2-\alpha)(1+\alpha+\alpha^2)^2\mu_\varepsilon}{(1+\alpha^2+\alpha^4)^2} & \frac{\alpha^3-1}{1+\alpha^2+\alpha^4} \\ \frac{2\alpha(\alpha^4-1)Y_{s-1}+(\alpha^2-1)(\alpha^4+3\alpha^2+1)Y_{s+2}+\alpha(2-\alpha)(1+\alpha+\alpha^2)^2\mu_\varepsilon}{(1+\alpha^2+\alpha^4)^2} & \frac{\alpha^3-1}{1+\alpha^2+\alpha^4} \end{bmatrix}.$$

Proof. Using (3.7.7) and that the sequences $(\widehat{\theta}^{\dagger\dagger}_{1,n}-\theta_1)_{n\in\mathbb{N}}$ and $(\widehat{\theta}^{\dagger\dagger}_{2,n}-\theta_2)_{n\in\mathbb{N}}$ are bounded with probability one, by the very same arguments as in the proof of (3.3.11), one can obtain (3.7.10). Now we turn to prove (3.7.11). Using the notation

$$B^{\dagger\dagger}_n := \begin{bmatrix} 1+(\widehat{\alpha}^{\dagger\dagger}_n)^2 & -\widehat{\alpha}^{\dagger\dagger}_n \\ -\widehat{\alpha}^{\dagger\dagger}_n & 1+(\widehat{\alpha}^{\dagger\dagger}_n)^2 \end{bmatrix},$$

by (3.7.8), we have

$$\begin{bmatrix} \widehat{\theta}^{\dagger\dagger}_{1,n} \\ \widehat{\theta}^{\dagger\dagger}_{2,n} \end{bmatrix} = (B^{\dagger\dagger}_n)^{-1} \begin{bmatrix} (1+(\widehat{\alpha}^{\dagger\dagger}_n)^2)Y_s - \widehat{\alpha}^{\dagger\dagger}_n(Y_{s-1}+Y_{s+1}) - (1-\widehat{\alpha}^{\dagger\dagger}_n)\widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} \\ (1+(\widehat{\alpha}^{\dagger\dagger}_n)^2)Y_{s+1} - \widehat{\alpha}^{\dagger\dagger}_n(Y_s+Y_{s+2}) - (1-\widehat{\alpha}^{\dagger\dagger}_n)\widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} \end{bmatrix}$$

holds asymptotically as $n\to\infty$ with probability one. Theorem 3.7.1 yields that

$$\mathsf{P}\left(\lim_{n\to\infty} B^{\dagger\dagger}_n = \begin{bmatrix} 1+\alpha^2 & -\alpha \\ -\alpha & 1+\alpha^2 \end{bmatrix} =: B^{\dagger\dagger}\right) = 1.$$

Again by Theorem 3.7.1, we have

$$\begin{aligned}
&\begin{bmatrix} \sqrt{n}\big(\widehat{\theta}^{\dagger\dagger}_{1,n} - \lim_{k\to\infty}\widehat{\theta}^{\dagger\dagger}_{1,k}\big) \\ \sqrt{n}\big(\widehat{\theta}^{\dagger\dagger}_{2,n} - \lim_{k\to\infty}\widehat{\theta}^{\dagger\dagger}_{2,k}\big) \end{bmatrix} \\
&= \sqrt{n}(B^{\dagger\dagger}_n)^{-1}\left(\begin{bmatrix} (1+(\widehat{\alpha}^{\dagger\dagger}_n)^2)Y_s - \widehat{\alpha}^{\dagger\dagger}_n(Y_{s-1}+Y_{s+1}) - (1-\widehat{\alpha}^{\dagger\dagger}_n)\widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} \\ (1+(\widehat{\alpha}^{\dagger\dagger}_n)^2)Y_{s+1} - \widehat{\alpha}^{\dagger\dagger}_n(Y_s+Y_{s+2}) - (1-\widehat{\alpha}^{\dagger\dagger}_n)\widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} \end{bmatrix}\right. \\
&\qquad\qquad \left. - B^{\dagger\dagger}_n(B^{\dagger\dagger})^{-1}\begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1}+Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s+Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}\right) \\
&= \sqrt{n}(B^{\dagger\dagger}_n)^{-1}\left(\begin{bmatrix} (1+(\widehat{\alpha}^{\dagger\dagger}_n)^2)Y_s - \widehat{\alpha}^{\dagger\dagger}_n(Y_{s-1}+Y_{s+1}) - (1-\widehat{\alpha}^{\dagger\dagger}_n)\widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} \\ (1+(\widehat{\alpha}^{\dagger\dagger}_n)^2)Y_{s+1} - \widehat{\alpha}^{\dagger\dagger}_n(Y_s+Y_{s+2}) - (1-\widehat{\alpha}^{\dagger\dagger}_n)\widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} \end{bmatrix}\right. \\
&\qquad\qquad \left. - \begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1}+Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s+Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}\right) \\
&\quad + \sqrt{n}\left((B^{\dagger\dagger}_n)^{-1} - (B^{\dagger\dagger})^{-1}\right)\begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1}+Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s+Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix} \\
&= \sqrt{n}(B^{\dagger\dagger}_n)^{-1}\begin{bmatrix} (\widehat{\alpha}^{\dagger\dagger}_n+\alpha)Y_s - (Y_{s-1}+Y_{s+1}) + \widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} & \alpha-1 \\ (\widehat{\alpha}^{\dagger\dagger}_n+\alpha)Y_{s+1} - (Y_s+Y_{s+2}) + \widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} & \alpha-1 \end{bmatrix}\begin{bmatrix} \widehat{\alpha}^{\dagger\dagger}_n - \alpha \\ \widehat{\mu}^{\dagger\dagger}_{\varepsilon,n} - \mu_\varepsilon \end{bmatrix} \\
&\quad + \sqrt{n}(B^{\dagger\dagger}_n)^{-1}\big(B^{\dagger\dagger} - B^{\dagger\dagger}_n\big)(B^{\dagger\dagger})^{-1}\begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1}+Y_{s+1}) - (1-\alpha)\mu_\varepsilon \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s+Y_{s+2}) - (1-\alpha)\mu_\varepsilon \end{bmatrix}.
\end{aligned}$$

Hence

$$\begin{bmatrix} \sqrt{n}\big(\widehat{\theta}_{1,n}^{\dagger\dagger} - \lim_{k\to\infty} \widehat{\theta}_{1,k}^{\dagger\dagger}\big) \\ \sqrt{n}\big(\widehat{\theta}_{2,n}^{\dagger\dagger} - \lim_{k\to\infty} \widehat{\theta}_{2,k}^{\dagger\dagger}\big) \end{bmatrix} = D_{n,\alpha,\varepsilon} \begin{bmatrix} \sqrt{n}(\widehat{\alpha}_n^{\dagger\dagger} - \alpha) \\ \sqrt{n}(\widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} - \mu_\varepsilon) \end{bmatrix} \tag{3.7.12}$$

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

$$D_{n,\alpha,\varepsilon} := (B_n^{\dagger\dagger})^{-1} \begin{bmatrix} (\widehat{\alpha}_n^{\dagger\dagger} + \alpha)Y_s - Y_{s-1} - Y_{s+1} + \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} & \alpha - 1 \\ (\widehat{\alpha}_n^{\dagger\dagger} + \alpha)Y_{s+1} - Y_s - Y_{s+2} + \widehat{\mu}_{\varepsilon,n}^{\dagger\dagger} & \alpha - 1 \end{bmatrix}$$

$$+ (B_n^{\dagger\dagger})^{-1} \begin{bmatrix} -(\widehat{\alpha}_n^{\dagger\dagger} + \alpha) & 1 \\ 1 & -(\widehat{\alpha}_n^{\dagger\dagger} + \alpha) \end{bmatrix} (B^{\dagger\dagger})^{-1} \begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1} + Y_{s+1}) - (1-\alpha)\mu_\varepsilon & 0 \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s + Y_{s+2}) - (1-\alpha)\mu_\varepsilon & 0 \end{bmatrix}.$$

By (3.7.3) and (3.7.4), we have $D_{n,\alpha,\varepsilon}$ converges almost surely as $n \to \infty$ to

$$(B^{\dagger\dagger})^{-1} \begin{bmatrix} 2\alpha Y_s - Y_{s-1} - Y_{s+1} + \mu_\varepsilon & \alpha - 1 \\ 2\alpha Y_{s+1} - Y_s - Y_{s+2} + \mu_\varepsilon & \alpha - 1 \end{bmatrix}$$

$$+ (B^{\dagger\dagger})^{-1} \begin{bmatrix} -2\alpha & 1 \\ 1 & -2\alpha \end{bmatrix} (B^{\dagger\dagger})^{-1} \begin{bmatrix} (1+\alpha^2)Y_s - \alpha(Y_{s-1} + Y_{s+1}) - (1-\alpha)\mu_\varepsilon & 0 \\ (1+\alpha^2)Y_{s+1} - \alpha(Y_s + Y_{s+2}) - (1-\alpha)\mu_\varepsilon & 0 \end{bmatrix} = D_{\alpha,\varepsilon}.$$

Hence, by (3.7.12), (3.7.10) and Slutsky's lemma, we have (3.7.11). □

4 The INAR(1) model with innovational outliers

4.1 The model and some preliminaries

In this section we introduce INAR(1) models contaminated with innovational outliers and we also give some preliminaries.

4.1.1 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 $I \in \mathbb{N}$ *and* $s_i,\ \theta_i \in \mathbb{N}$, $i = 1, \ldots, I$. *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 is given in the following proposition.

4.1.1 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) model 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

$$Z_k := \begin{cases} 0 & \text{if } k = 0,1,\dots,s-1,\\ \theta & \text{if } k = s,\\ \sum_{j=X_{k-1}+1}^{X_{k-1}+Z_{k-1}} \xi_{k,j} & \text{if } k \geqslant s+1. \end{cases} \tag{4.1.1}$$

Moreover, 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. Clearly, $Y_j = X_j + Z_j$ for $j = 0,1,\dots,s-1$, and

$$Y_s = \sum_{j=1}^{Y_{s-1}} \xi_{s,j} + \eta_s = \sum_{j=1}^{X_{s-1}} \xi_{s,j} + \varepsilon_s + \theta = X_s + \theta = X_s + Z_s,$$

$$\begin{aligned} Y_{s+1} &= \sum_{j=1}^{Y_s} \xi_{s+1,j} + \eta_{s+1} = \sum_{j=1}^{X_s+Z_s} \xi_{s+1,j} + \varepsilon_{s+1} = \sum_{j=1}^{X_s} \xi_{s+1,j} + \sum_{j=X_s+1}^{X_s+Z_s} \xi_{s+1,j} + \varepsilon_{s+1}\\ &= X_{s+1} + \sum_{j=X_s+1}^{X_s+Z_s} \xi_{s+1,j} = X_{s+1} + Z_{s+1}. \end{aligned}$$

By induction, we easily conclude that $Y_k(\omega) = X_k(\omega) + Z_k(\omega)$ for all $\omega\in\Omega$ and $k\in\mathbb{Z}_+$.

In proving the independence of the processes X and Z, it is enough to check that the conditions of Lemma 5.2 (see Appendix) 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

$$\begin{aligned} \mathsf{P}(X_n = i_n, Z_n = j_n \mid A) &= \mathsf{P}\left(\sum_{j=1}^{i_{n-1}} \xi_{n,j} + \varepsilon_n = i_n, \sum_{j=i_{n-1}+1}^{i_{n-1}+j_{n-1}} \xi_{n,j} = j_n\right)\\ &= \mathsf{P}\left(\sum_{j=1}^{i_{n-1}} \xi_{n,j} + \varepsilon_n = i_n\right) \mathsf{P}\left(\sum_{j=i_{n-1}+1}^{i_{n-1}+j_{n-1}} \xi_{n,j} = j_n\right), \end{aligned} \tag{4.1.2}$$

where we used the measurability of (X_{n-1}, Z_{n-1}) with respect to the σ–algebra $\sigma(\xi_{i,j} : i = 1,\dots,n-1,\ j\in\mathbb{N})$ and that the random variables ε_n, $(\xi_{n,1},\dots,\xi_{n,i_{n-1}})$ 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 $n > s$,

$$\mathsf{P}(X_n = i_n, Z_n = j_n \mid A) = \mathsf{P}(X_n = i_n, Z_n = j_n \mid X_{n-1} = i_{n-1}, Z_{n-1} = j_{n-1}). \tag{4.1.3}$$

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 (4.1.3) 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 (4.1.2), with the special choice $B := \Omega$ lead to

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

which yields that the conditions of Lemma 5.2 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 sequence $(Z_k(\omega))_{k\geqslant s+1}$ is monotone decreasing for all $\omega \in \Omega$. Using the fact that $Z_k \geqslant 0$, $k \in \mathbb{N}$, we have $(Z_k(\omega))_{k\in\mathbb{Z}_+}$ 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 $X_0, X_1, \ldots, X_k$ and $Z_0, Z_1, \ldots, Z_k$. Using that $\mathsf{E}(Z_k \mid \mathcal{F}_{k-1}^{X,Z}) = \alpha Z_{k-1}$, $k \geqslant s+1$, 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,

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

as desired. We note that the fact that $\mathsf{P}(\lim_{k\to\infty} Z_k = 0) = 1$ is in accordance with Theorem 2 in Chapter I in Athreya and Ney [8].

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 $\left(|Z_k|^p \mathbb{1}_{\{|Z_k|\geqslant M\}}\right)_{k\geqslant s+1}$ is monotone decreasing. Hence

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

which yields the uniformly integrability of $(Z_k^p)_{k\in\mathbb{N}}$. By Lemma 5.3 (see Appendix), we conclude that $Z_k \xrightarrow{L_p} 0$ as $k \to \infty$, i.e., $\lim_{k\to\infty} \mathsf{E}Z_k^p = 0$. □

For our later purposes we need the following lemma about the explicit forms of the first and second moments of the process Z.

4.1.1 Lemma. *We have*

$$\mathsf{E}Z_{s+k} = \theta\alpha^k, \qquad k \in \mathbb{Z}_+, \tag{4.1.4}$$

$$\mathsf{E}Z_{s+k}^2 = \theta^2\alpha^{2k} - \theta\alpha^k(\alpha^k - 1), \qquad k \in \mathbb{Z}_+, \tag{4.1.5}$$

$$\mathsf{E}(Z_{s+k-1}Z_{s+k}) = \alpha \mathsf{E}Z_{s+k-1}^2 = \theta^2\alpha^{2k-1} - \theta\alpha^k(\alpha^{k-1} - 1), \qquad k \in \mathbb{N}. \tag{4.1.6}$$

Proof. Recall that $\mathcal{F}_k^{X,Z}$ denotes the σ–algebra generated by the random variables $X_0, X_1, \ldots, X_k$ and $Z_0, Z_1, \ldots, Z_k$. 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 $\mathsf{E}Z_{s+k} = \alpha^k \mathsf{E}Z_s = \theta\alpha^k$, $k \in \mathbb{Z}_+$. Since $\alpha \in (0,1)$, we have $\lim_{k\to\infty} \mathsf{E}Z_k = 0$. Moreover, using that

(4.1.7)

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

we get

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

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, for all $k \geqslant 0$,

$$\begin{aligned} \begin{bmatrix} \mathsf{E}Z_{s+k} \\ \mathsf{E}Z_{s+k}^2 \end{bmatrix} &= \begin{bmatrix} \alpha^k & 0 \\ (1-\alpha)\alpha^k \sum_{\ell=0}^{k-1} \alpha^\ell & \alpha^{2k} \end{bmatrix} \begin{bmatrix} \mathsf{E}Z_s \\ \mathsf{E}Z_s^2 \end{bmatrix} = \begin{bmatrix} \alpha^k & 0 \\ (1-\alpha)\alpha^k \frac{\alpha^k - 1}{\alpha - 1} & \alpha^{2k} \end{bmatrix} \begin{bmatrix} \theta \\ \theta^2 \end{bmatrix} \\ &= \begin{bmatrix} \theta\alpha^k \\ \theta^2\alpha^{2k} - \theta\alpha^k(\alpha^k - 1) \end{bmatrix}. \end{aligned}$$

Finally, for all $k \in \mathbb{N}$,

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

which yields (4.1.6). □

In case of two (innovational) outliers a similar representation of Y is given in the following proposition.

4.1.2 Proposition. *Let $(Y_k)_{k\in\mathbb{Z}_+}$ be an INAR(1) model with two innovational outliers θ_1 and θ_2 at time points s_1 and s_2, $s_1 < s_2$,*

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

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 $\eta_k := \varepsilon_k + \delta_{k,s_1}\theta_1 + \delta_{k,s_2}\theta_2$, $k \in \mathbb{Z}_+$. Then for all $\omega \in \Omega$

and $k \in \mathbb{Z}_+$, $Y_k(\omega) = X_k(\omega) + Z_k^{(1)}(\omega) + Z_k^{(2)}(\omega)$, *where* $(X_k)_{k\in\mathbb{Z}_+}$ *is an INAR(1) model 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*

$$(4.1.8) \qquad Z_k^{(1)} := \begin{cases} 0 & \text{if } k = 0, 1, \ldots, s_1 - 1, \\ \theta_1 & \text{if } k = s_1, \\ \sum_{j=X_{k-1}+1}^{X_{k-1}+Z_{k-1}^{(1)}} \xi_{k,j} & \text{if } k \geqslant s_1 + 1, \end{cases}$$

and

$$(4.1.9) \qquad Z_k^{(2)} := \begin{cases} 0 & \text{if } k = 0, 1, \ldots, s_2 - 1, \\ \theta_2 & \text{if } k = s_2, \\ \sum_{j=X_{k-1}+Z_{k-1}^{(1)}+1}^{X_{k-1}+Z_{k-1}^{(1)}+Z_{k-1}^{(2)}} \xi_{k,j} & \text{if } k \geqslant s_2 + 1. \end{cases}$$

Moreover, the processes X, $Z^{(1)}$ *and* $Z^{(2)}$ *are (pairwise) independent, and* $\mathsf{P}(\lim_{k\to\infty} Z_k^{(i)} = 0) = 1$, $i = 1, 2$, *and* $Z_k^{(i)} \xrightarrow{L_p} 0$ *as* $k \to \infty$ *for all* $p \in \mathbb{N}$, $i = 1, 2$.

Proof. The proof is the very same as the proof of Proposition 4.1.1. We only note that the independence of $Z^{(1)}$ and $Z^{(2)}$ follows by the definitions of the processes $Z^{(1)}$ and $Z^{(2)}$. □

In the sequel we denote by $\mathcal{F}_k^Y$ the σ–algebra generated by the random variables $Y_0, Y_1, \ldots, Y_k$. For all $n \in \mathbb{N}$, $y_0, \ldots, y_n \in \mathbb{R}$ and $\omega \in \Omega$, let us introduce the notations

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

4.2 One outlier, estimation of the mean of the offspring distribution and the outlier's size

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

$$(4.2.1) \qquad \mathsf{E}(Y_k \mid \mathcal{F}_{k-1}^Y) = \alpha Y_{k-1} + \mathsf{E}\eta_k = \alpha Y_{k-1} + \mu_\varepsilon + \delta_{k,s}\theta, \qquad k \in \mathbb{N}.$$

Hence for $n \geqslant s$, $n \in \mathbb{N}$,

$$(4.2.2) \qquad \begin{aligned} &\sum_{k=1}^{n} \big(Y_k - \mathsf{E}(Y_k \mid \mathcal{F}_{k-1}^Y)\big)^2 \\ &\qquad = \sum_{k=1}^{n}{}^{(s)} \big(Y_k - \alpha Y_{k-1} - \mu_\varepsilon\big)^2 + \big(Y_s - \alpha Y_{s-1} - \mu_\varepsilon - \theta\big)^2. \end{aligned}$$

For all $n \geqslant s$, $n \in \mathbb{N}$, we define the function $Q_n : \mathbb{R}^{n+1} \times \mathbb{R}^2 \to \mathbb{R}$, as

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

for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $\alpha', \theta' \in \mathbb{R}$. By definition, for all $n \geqslant s$, a CLS estimator for the parameter $(\alpha, \theta) \in (0,1) \times \mathbb{N}$ is a measurable function $(\widetilde{\alpha}_n, \widetilde{\theta}_n) : S_n \to \mathbb{R}^2$ such that

$$Q_n(\mathbf{y}_n; \widetilde{\alpha}_n(\mathbf{y}_n), \widetilde{\theta}_n(\mathbf{y}_n)) = \inf_{(\alpha', \theta') \in \mathbb{R}^2} Q_n(\mathbf{y}_n; \alpha', \theta') \qquad \forall\ \mathbf{y}_n \in S_n,$$

where S_n is suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 4.2.1). We note that we do not define the CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_n)$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$. We have

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

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

The next lemma is about the existence and uniqueness of the CLS estimator of (α, θ).

4.2.1 Lemma. *There exist subsets $S_n \subset \mathbb{R}^{n+1}$, $n \geqslant \max(3, s+1)$ with the following properties:*

(i) *there exists a unique CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_n) : S_n \to \mathbb{R}^2$,*

(ii) *for all $\mathbf{y}_n \in S_n$, the system of equations*

$$\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \theta') = 0, \qquad \frac{\partial Q_n}{\partial \theta'}(\mathbf{y}_n; \alpha', \theta') = 0, \tag{4.2.3}$$

has the unique solution

$$\widetilde{\alpha}_n(\mathbf{y}_n) = \frac{\sum_{k=1}^{n}{}^{(s)} (y_k - \mu_\varepsilon) y_{k-1}}{\sum_{k=1}^{n}{}^{(s)} y_{k-1}^2}, \tag{4.2.4}$$

$$\widetilde{\theta}_n(\mathbf{y}_n) = y_s - \widetilde{\alpha}_n(\mathbf{y}_n) y_{s-1} - \mu_\varepsilon, \tag{4.2.5}$$

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

Proof. One can easily check that the unique solution of the system of equations (4.2.3) takes the form (4.2.4) and (4.2.5) whenever $\sum_{k=1}^{n}{}^{(s)} y_{k-1}^2 > 0$.

Next we prove that the function $\mathbb{R}^2 \ni (\alpha', \theta') \mapsto Q_n(\mathbf{y}_n; \alpha', \theta')$ is strictly convex for all $\mathbf{y}_n \in S_n$, where

$$S_n := \left\{ \mathbf{y}_n \in \mathbb{R}^{n+1} : \sum_{k=1}^{n}{}^{(s)} y_{k-1}^2 > 0 \right\}.$$

For this it is enough to check that the (2×2)-matrix

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

is (strictly) positive definite for all $\mathbf{y}_n \in S_n$, see, e.g., Berkovitz [10, Theorem 3.3, Chapter III]. For all $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $(\alpha', \theta') \in \mathbb{R}^2$,

$$\begin{aligned}
&\frac{\partial^2 Q_n}{\partial(\alpha')^2}(\mathbf{y}_n; \alpha', \theta') = 2\sum_{k=1}^{n}{}^{(s)}\, y_{k-1}^2 + 2y_{s-1}^2 = 2\sum_{k=1}^{n} y_{k-1}^2, \\
&\frac{\partial^2 Q_n}{\partial\alpha'\partial\theta'}(\mathbf{y}_n; \alpha', \theta') = \frac{\partial^2 Q_n}{\partial\theta'\partial\alpha'}(\mathbf{y}_n; \alpha', \theta') = 2y_{s-1}, \\
&\frac{\partial^2 Q_n}{\partial(\theta')^2}(\mathbf{y}_n; \alpha', \theta') = 2.
\end{aligned}$$

Then $H_n(\mathbf{y}_n; \alpha', \theta')$ has leading principal minors

$$2\sum_{k=1}^{n} y_{k-1}^2 \qquad \text{and} \qquad 4\sum_{k=1}^{n}{}^{(s)}\, y_{k-1}^2,$$

which are positive for all $\mathbf{y}_n \in S_n$. Hence $H_n(\mathbf{y}_n; \alpha', \theta')$ is (strictly) positive definite for all $\mathbf{y}_n \in S_n$.

Since the function $\mathbb{R}^2 \ni (\alpha', \theta') \mapsto Q_n(\mathbf{y}_n; \alpha', \theta')$ is strictly convex for all $\mathbf{y}_n \in S_n$ and the system of equations (4.2.3) has a unique solution for all $\mathbf{y}_n \in S_n$, we get the function in question attains its (global) minimum at this unique solution, which yields (i) and (ii).

Next we check that $\mathbf{Y}_n \in S_n$ holds asymptotically as $n \to \infty$ with probability one. By Proposition 4.1.1, we get

$$\sum_{k=1}^{n}{}^{(s)}\, Y_{k-1}^2 = \sum_{k=1}^{n}{}^{(s)}\, X_{k-1}^2 + 2\sum_{k=1}^{n}{}^{(s)}\, X_{k-1}Z_{k-1} + \sum_{k=1}^{n}{}^{(s)}\, Z_{k-1}^2, \qquad n \geqslant s+1.$$

Using again Proposition 4.1.1 and (2.2.6), we have

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

By Cauchy-Schwartz's inequality,

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

and hence

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}{}^{(s)}\, X_{k-1}Z_{k-1} = 0\right) = 1. \tag{4.2.6}$$

Then

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

which implies that

$$\mathsf{P}\left(\lim_{n\to\infty}\sum_{k=1}^{n}Y_{k-1}^2=\infty\right)=1.$$

Hence $\mathbf{Y}_n\in S_n$ holds asymptotically as $n\to\infty$ with probability one. □

By Lemma 4.2.1, $(\widetilde{\alpha}_n(\mathbf{Y}_n),\widetilde{\theta}_n(\mathbf{Y}_n))$ exists uniquely asymptotically as $n\to\infty$ with probability one. In the sequel we will simply denote it by $(\widetilde{\alpha}_n,\widetilde{\theta}_n)$.

The next result shows that $\widetilde{\alpha}_n$ is a strongly consistent estimator of α, whereas $\widetilde{\theta}_n$ fails to be a strongly consistent estimator of θ.

4.2.1 Theorem. *Consider the CLS estimators $(\widetilde{\alpha}_n,\widetilde{\theta}_n)_{n\in\mathbb{N}}$ of the parameter $(\alpha,\theta)\in(0,1)\times\mathbb{N}$. The sequence $(\widetilde{\alpha}_n)_{n\in\mathbb{N}}$ is strongly consistent for all $(\alpha,\theta)\in(0,1)\times\mathbb{N}$, i.e.,*

$$\mathsf{P}(\lim_{n\to\infty}\widetilde{\alpha}_n=\alpha)=1,\qquad\forall\,(\alpha,\theta)\in(0,1)\times\mathbb{N},\tag{4.2.7}$$

whereas the sequence $(\widetilde{\theta}_n)_{n\in\mathbb{N}}$ is not strongly consistent for any $(\alpha,\theta)\in(0,1)\times\mathbb{N}$, namely,

$$\mathsf{P}\left(\lim_{n\to\infty}\widetilde{\theta}_n=Y_s-\alpha Y_{s-1}-\mu_\varepsilon\right)=1,\qquad\forall\,(\alpha,\theta)\in(0,1)\times\mathbb{N}.\tag{4.2.8}$$

Proof. By (4.2.4) and Proposition 4.1.1, we have asymptotically as $n\to\infty$ with probability one,

$$\widetilde{\alpha}_n=\frac{\sum_{k=1}^{n}{}^{(s)}(X_k-\mu_\varepsilon+Z_k)(X_{k-1}+Z_{k-1})}{\sum_{k=1}^{n}{}^{(s)}(X_{k-1}+Z_{k-1})^2}$$

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

By (2.2.2), to prove (4.2.7), it is enough to check that

(4.2.9)
$$\mathsf{P}\left(\lim_{n\to\infty}\frac{-(X_s-\mu_\varepsilon)X_{s-1}+\sum_{k=s+1}^{n}(X_k-\mu_\varepsilon)Z_{k-1}+\sum_{k=s+1}^{n}X_{k-1}Z_k+\sum_{k=s+1}^{n}Z_{k-1}Z_k}{n}=0\right)=1,$$

and

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{-X_{s-1}^2+2\sum_{k=s+1}^{n}X_{k-1}Z_{k-1}+\sum_{k=s+1}^{n}Z_{k-1}^2}{n}=0\right)=1.\tag{4.2.10}$$

By Proposition 4.1.1 and Cauchy-Schwartz's inequality, we have

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=s+1}^{n}Z_{k-1}^2=0\right)=\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=s+1}^{n}Z_{k-1}Z_k=0\right)=1.$$

Hence, using also (4.2.6), we get (4.2.9) and (4.2.10). By (4.2.5) and (4.2.7), we get (4.2.8). □

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

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

where $\sigma_{\alpha,\varepsilon}^2$ *is defined in* (2.2.9). *Moreover, conditionally on the value* Y_{s-1},

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

Proof. By (4.2.4), we have

$$\widetilde{\alpha}_n-\alpha=\frac{\sum_{k=1}^{n}{}^{(s)}(Y_k-\alpha Y_{k-1}-\mu_\varepsilon)Y_{k-1}}{\sum_{k=1}^{n}{}^{(s)}\,Y_{k-1}^2}$$

holds asymptotically as $n\to\infty$ with probability one. For all $n\geqslant s+1$, by Proposition 4.1.1, we have

$$\begin{aligned}
&\sum_{k=1}^{n}{}^{(s)}(Y_k-\alpha Y_{k-1}-\mu_\varepsilon)Y_{k-1}\\
&\quad=\sum_{k=1}^{n}{}^{(s)}\big[(X_k-\alpha X_{k-1}-\mu_\varepsilon)+(Z_k-\alpha Z_{k-1})\big](X_{k-1}+Z_{k-1})\\
&\quad=\sum_{k=1}^{n}{}^{(s)}(X_k-\alpha X_{k-1}-\mu_\varepsilon)X_{k-1}+\sum_{k=s+1}^{n}(X_k-\alpha X_{k-1}-\mu_\varepsilon)Z_{k-1}+\sum_{k=s+1}^{n}(Z_k-\alpha Z_{k-1})X_{k-1}\\
&\qquad+\sum_{k=s+1}^{n}(Z_k-\alpha Z_{k-1})Z_{k-1},
\end{aligned}$$

and

$$\sum_{k=1}^{n}{}^{(s)}\,Y_{k-1}^2=\sum_{k=1}^{n}{}^{(s)}\,X_{k-1}^2+2\sum_{k=s+1}^{n}X_{k-1}Z_{k-1}+\sum_{k=s+1}^{n}Z_{k-1}^2.$$

By (2.2.1) and (2.2.8), we have

(4.2.13) $$\sqrt{n}\,\frac{\sum_{k=1}^{n}(X_k-\alpha X_{k-1}-\mu_\varepsilon)X_{k-1}}{\sum_{k=1}^{n}X_{k-1}^2}\stackrel{\mathcal{L}}{\longrightarrow}\mathcal{N}(0,\sigma_{\alpha,\varepsilon}^2)\qquad \text{as}\ \ n\to\infty.$$

In what follows we show that

$$\frac{1}{\sqrt{n}} \sum_{k=s+1}^{n} (X_k - \alpha X_{k-1} - \mu_\varepsilon) Z_{k-1} \xrightarrow{L_1} 0 \qquad \text{as} \quad n \to \infty, \tag{4.2.14}$$

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

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

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=s+1}^{n} X_{k-1} Z_{k-1} = 0 \right) = 1, \tag{4.2.17}$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=s+1}^{n} Z_{k-1}^2 = 0 \right) = 1, \tag{4.2.18}$$

where $\xrightarrow{L_1}$ denotes convergence in L_1. We recall that 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 and hence in L_1 to 0 as $n \to \infty$. Hence to prove (4.2.14) it is enough to check that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=s+1}^{n} \mathsf{E}\big[(X_k - \alpha X_{k-1} - \mu_\varepsilon) Z_{k-1}\big] = 0, \tag{4.2.19}$$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left(\sum_{k=s+1}^{n} (X_k - \alpha X_{k-1} - \mu_\varepsilon) Z_{k-1} \right)^2 = 0. \tag{4.2.20}$$

Since $\mathsf{E}X_k = \alpha \mathsf{E}X_{k-1} + \mu_\varepsilon$, $k \in \mathbb{N}$, and the processes X and Z are independent, we have (4.2.19). Similarly, we get

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

By (3.2.14),

$$\mathsf{E}(X_k - \alpha X_{k-1} - \mu_\varepsilon)^2 = \alpha^k (1-\alpha) \mathsf{E}X_0 + \alpha(1 - \alpha^{k-1}) \mu_\varepsilon + \sigma_\varepsilon^2, \qquad k \in \mathbb{N},$$

and then

$$\lim_{k\to\infty} \mathsf{E}(X_k - \alpha X_{k-1} - \mu_\varepsilon)^2 = \lim_{k\to\infty} \big(\alpha^k (1-\alpha) \mathsf{E}X_0 + \alpha(1 - \alpha^{k-1}) \mu_\varepsilon + \sigma_\varepsilon^2\big) = \alpha\mu_\varepsilon + \sigma_\varepsilon^2.$$

Hence there exists some $L > 0$ such that $\mathsf{E}(X_k - \alpha X_{k-1} - \mu_\varepsilon)^2 < L$ for all $k \in \mathbb{N}$. By Proposition 4.1.1, $\lim_{k\to\infty} \mathsf{E}Z_k^2 = 0$, and hence $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}Z_k^2 = 0$, which yields that

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

To prove (4.2.15), it is enough to check that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=s+1}^{n} \mathsf{E}\big[(Z_k - \alpha Z_{k-1})X_{k-1}\big] = 0, \tag{4.2.21}$$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left(\sum_{k=s+1}^{n} (Z_k - \alpha Z_{k-1})X_{k-1}\right)^2 = 0. \tag{4.2.22}$$

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 (4.2.21). Similarly, we get

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

Using that $\lim_{k\to\infty} \mathsf{E}X_k^2 = \mathsf{E}\widetilde{X}^2$ (see, (2.1.2) and (2.2.4)), there exists some $L > 0$ such that $\mathsf{E}X_k^2 < L$ for all $k \in \mathbb{N}$. By Proposition 4.1.1, $\lim_{k\to\infty} \mathsf{E}(Z_k - \alpha Z_{k-1})^2 \leqslant \lim_{k\to\infty} 2\mathsf{E}(Z_k^2 + \alpha^2 Z_{k-1}^2) = 0$, and hence $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathsf{E}(Z_k - \alpha Z_{k-1})^2 = 0$, which yields that

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

To prove (4.2.16), it is enough to check that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=s+1}^{n} \mathsf{E}\big[(Z_k - \alpha Z_{k-1})Z_{k-1}\big] = 0, \tag{4.2.23}$$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left(\sum_{k=s+1}^{n} (Z_k - \alpha Z_{k-1})Z_{k-1}\right)^2 = 0. \tag{4.2.24}$$

Using that $\mathsf{E}\big[(Z_k - \alpha Z_{k-1})Z_{k-1}\big] = \mathsf{E}(Z_{k-1}\mathsf{E}(Z_k - \alpha Z_{k-1} \,|\, Z_{k-1})) = 0$, $k \in \mathbb{N}$, we get (4.2.23). For all $k > \ell$, $k, \ell \in \mathbb{N}$, we get

$$\begin{aligned}
\mathsf{E}[(Z_k - \alpha Z_{k-1})Z_{k-1}(Z_\ell - \alpha Z_{\ell-1})Z_{\ell-1}] &= \mathsf{E}\big[\mathsf{E}[(Z_k - \alpha Z_{k-1})Z_{k-1}(Z_\ell - \alpha Z_{\ell-1})Z_{\ell-1} \,|\, \mathcal{F}_{k-1}^Z]\big] \\
&= \mathsf{E}\big[Z_{k-1}(Z_\ell - \alpha Z_{\ell-1})Z_{\ell-1}\mathsf{E}(Z_k - \alpha Z_{k-1} \,|\, \mathcal{F}_{k-1}^Z)\big] = 0,
\end{aligned}$$

and hence, by (4.1.7), we obtain

$$\begin{aligned}
\frac{1}{n}\mathsf{E}\left(\sum_{k=s+1}^{n} (Z_k - \alpha Z_{k-1})Z_{k-1}\right)^2 &= \frac{1}{n} \sum_{k=s+1}^{n} \mathsf{E}[(Z_k - \alpha Z_{k-1})^2 Z_{k-1}^2] \\
&= \frac{1}{n} \sum_{k=s+1}^{n} \mathsf{E}\big[Z_{k-1}^2 \mathsf{E}((Z_k - \alpha Z_{k-1})^2 \,|\, \mathcal{F}_{k-1}^Z)\big] \\
&= \frac{1}{n} \sum_{k=s+1}^{n} \mathsf{E}\big[Z_{k-1}^2 \alpha(1-\alpha) Z_{k-1}\big] = \frac{\alpha(1-\alpha)}{n} \sum_{k=s+1}^{n} \mathsf{E}Z_{k-1}^3.
\end{aligned}$$

By Proposition 4.1.1, this implies (4.2.24). Condition (4.2.17) was already proved, see (4.2.6). Finally, Proposition 4.1.1 easily yields (4.2.18). Using (4.2.13) – (4.2.18), Slutsky's lemma yields (4.2.11).

By (4.2.5) and (4.2.8),

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

holds asymptotically as $n \to \infty$ with probability one, and hence by (4.2.11), we get (4.2.12). □

4.2.1 Remark. By (4.1.4) and (3.2.13),

$$\mathsf{E}Y_k = \begin{cases} \alpha^k \mathsf{E}Y_0 + \mu_\varepsilon \frac{1-\alpha^k}{1-\alpha} & \text{if } k = 1, \dots, s-1, \\ \alpha^k \mathsf{E}Y_0 + \theta\alpha^{k-s} + \mu_\varepsilon \frac{1-\alpha^k}{1-\alpha} & \text{if } k \geqslant s. \end{cases}$$

Hence $\mathsf{E}(Y_s - \alpha Y_{s-1} - \mu_\varepsilon) = \theta$, $\theta \in \mathbb{N}$. Moreover, by (3.2.14),

$$\begin{aligned}\operatorname{Var}(Y_s - \alpha Y_{s-1} - \mu_\varepsilon) &= \operatorname{Var}(X_s - \alpha X_{s-1} - \mu_\varepsilon + \theta) = \operatorname{Var}(X_s - \alpha X_{s-1} - \mu_\varepsilon) \\ &= \alpha^s(1-\alpha)\mathsf{E}X_0 + \alpha\mu_\varepsilon(1-\alpha^{s-1}) + \sigma_\varepsilon^2 \\ &= \alpha^s(1-\alpha)\mathsf{E}Y_0 + \alpha\mu_\varepsilon(1-\alpha^{s-1}) + \sigma_\varepsilon^2.\end{aligned}$$

If $k \geqslant s+1$, then one can derive a more complicated formula for $\operatorname{Var}(Y_k - \alpha Y_{k-1} - \mu_\varepsilon)$ containing the moments of Z, too.

We also check that $\widetilde{\theta}_n$ is an asymptotically unbiased estimator of θ as $n \to \infty$ for all $(\alpha, \theta) \in (0,1) \times \mathbb{N}$. By (4.2.8), the sequence $\widetilde{\theta}_n - \theta$, $n \in \mathbb{N}$, converges with probability one and hence bounded with probability one, and then the dominated convergence theorem yields that $\lim_{n\to\infty} \mathsf{E}(\widetilde{\theta}_n - \theta) = 0$. □

4.3 One outlier, 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 concentrate on the CLS estimation of $(\alpha, \mu_\varepsilon, \theta)$, where $\theta := \theta_1$. Motivated by (4.2.2), for all $n \geqslant s$, $n \in \mathbb{N}$, we define the function $Q_n : \mathbb{R}^{n+1} \times \mathbb{R}^3 \to \mathbb{R}$, as

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

for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $\alpha', \mu_\varepsilon', \theta' \in \mathbb{R}$. 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 $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n) : S_n \to \mathbb{R}^3$ such that

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

where S_n is suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 4.3.1). We note that we do not define the CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n)$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$. We get

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

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

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

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

4.3.1 Lemma. *There exist subsets* $S_n \subset \mathbb{R}^{n+1}$, $n \geqslant s$ *with the following properties:*

(i) *there exists a unique CLS estimator* $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n) : S_n \to \mathbb{R}^3$,

(ii) *for all* $\mathbf{y}_n \in S_n$, *the system of equations*

$$\begin{aligned} &\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta') = 0, \\ &\frac{\partial Q_n}{\partial \mu_\varepsilon'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta') = 0, \\ &\frac{\partial Q_n}{\partial \theta'}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta') = 0, \end{aligned} \tag{4.3.1}$$

has the unique solution

$$\widehat{\alpha}_n(\mathbf{y}_n) = \frac{(n-1) \sum_{k=1}^{n}{}^{(s)} y_{k-1} y_k - \sum_{k=1}^{n}{}^{(s)} y_k \sum_{k=1}^{n}{}^{(s)} y_{k-1}}{D_n(\mathbf{y}_n)}, \tag{4.3.2}$$

$$\widehat{\mu}_{\varepsilon,n}(\mathbf{y}_n) = \frac{\sum_{k=1}^{n}{}^{(s)} y_{k-1}^2 \sum_{k=1}^{n}{}^{(s)} y_k - \sum_{k=1}^{n}{}^{(s)} y_{k-1} \sum_{k=1}^{n}{}^{(s)} y_{k-1} y_k}{D_n(\mathbf{y}_n)}, \tag{4.3.3}$$

$$\widehat{\theta}_n(\mathbf{y}_n) = y_s - \widehat{\alpha}_n(\mathbf{y}_n) y_{s-1} - \widehat{\mu}_{\varepsilon,n}(\mathbf{y}_n), \tag{4.3.4}$$

where

$$D_n(\mathbf{y}_n) := (n-1) \sum_{k=1}^{n}{}^{(s)} y_{k-1}^2 - \left(\sum_{k=1}^{n}{}^{(s)} y_{k-1} \right)^2, \qquad n \geqslant s,$$

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

Proof. One can easily check that the unique solution of the system of equations (4.3.1) takes the form (4.3.2)-(4.3.3)-(4.3.4) whenever $D_n(\mathbf{y}_n) > 0$.

For all $n \geqslant s+1$, let

$$S_n := \left\{ \mathbf{y}_n \in \mathbb{R}^{n+1} : D_n(\mathbf{y}_n) > 0,\ \Delta_{i,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') > 0,\quad i = 1,2,3,\ \forall\, (\alpha', \mu'_\varepsilon, \theta') \in \mathbb{R}^3 \right\},$$

where $\Delta_{i,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta')$, $i = 1,2,3$, denotes the i-th order leading principal minor of the 3×3 matrix

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

Then the function $\mathbb{R}^3 \ni (\alpha', \mu'_\varepsilon, \theta') \mapsto Q_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta')$ is strictly convex for all $\mathbf{y}_n \in S_n$, see, e.g., Berkovitz [10, Theorem 3.3, Chapter III].

Since the function $\mathbb{R}^3 \ni (\alpha', \mu'_\varepsilon, \theta') \mapsto Q_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta')$ is strictly convex for all $\mathbf{y}_n \in S_n$ and the system of equations (4.3.1) has a unique solution for all $\mathbf{y}_n \in S_n$, we get the function in question attains its (global) minimum at this unique solution, which yields (i) and (ii).

Further, for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $(\alpha', \mu'_\varepsilon, \theta') \in \mathbb{R}^3$, we have

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

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

and

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

$$\frac{\partial^2 Q_n}{\partial \theta' \partial \mu'_\varepsilon}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = \frac{\partial^2 Q_n}{\partial \mu'_\varepsilon \partial \theta'}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = 2,$$

$$\frac{\partial^2 Q_n}{\partial (\theta')^2}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = 2, \qquad \frac{\partial^2 Q_n}{\partial (\mu'_\varepsilon)^2}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = 2n.$$

Then $H_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta')$ has the following leading principal minors

$$\Delta_{1,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = 2 \sum_{k=1}^{n} y_{k-1}^2, \quad \Delta_{2,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') = 4 \left(n \sum_{k=1}^{n} y_{k-1}^2 - \left(\sum_{k=1}^{n} y_{k-1} \right)^2 \right),$$

and

$$\begin{aligned} \Delta_{3,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') &= \det H_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta') \\ &= 8 \left((n-1) \sum_{k=1}^{n} y_{k-1}^2 + 2 Y_{s-1} \sum_{k=1}^{n} y_{k-1} - n (y_{s-1})^2 - \left(\sum_{k=1}^{n} y_{k-1} \right)^2 \right). \end{aligned}$$

Note that $\Delta_{i,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta')$, $i = 1,2,3$, do not depend on $(\alpha', \mu'_\varepsilon, \theta')$, and hence we will simply denote $\Delta_{i,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta')$ by $\Delta_{i,n}(\mathbf{y}_n)$.

Next we check that $\mathbf{Y}_n \in S_n$ holds asymptotically as $n \to \infty$ with probability one. By (2.2.5) and (2.2.6), using the very same arguments as in the proof of Lemma 4.2.1, one can get

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\Delta_{1,n}(\mathbf{Y}_n)}{n} = 2\mathsf{E}\widetilde{X}^2\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\Delta_{2,n}(\mathbf{Y}_n)}{n^2} = 4\operatorname{Var}\widetilde{X}\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\Delta_{3,n}(\mathbf{Y}_n)}{n^2} = 8\operatorname{Var}\widetilde{X}\right) = 1,$$

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

$$\mathsf{P}\left(\lim_{n\to\infty} \Delta_{i,n}(\mathbf{Y}_n) = \infty\right) = 1, \qquad i = 1, 2, 3.$$

By (2.2.5) and (2.2.6), we also get

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{D_n(\mathbf{Y}_n)}{n^2} = \operatorname{Var}\widetilde{X}\right) = 1 \tag{4.3.5}$$

and hence $\mathsf{P}(\lim_{n\to\infty} D_n(\mathbf{Y}_n) = \infty) = 1$. □

By Lemma 4.3.1, $(\widehat{\alpha}_n(\mathbf{Y}_n), \widehat{\mu}_{\varepsilon,n}(\mathbf{Y}_n), \widehat{\theta}_n(\mathbf{Y}_n))$ exists uniquely asymptotically as $n \to \infty$ with probability one. In the sequel we will simply denote it by $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_n)$, we will also denote $D_n(\mathbf{Y}_n)$ by D_n.

By (4.3.2) and (4.3.3), we also get

$$\widehat{\mu}_{\varepsilon,n} = \frac{1}{n-1}\left(\sum_{k=1}^{n}{}^{(s)}\, Y_k - \widehat{\alpha}_n \sum_{k=1}^{n}{}^{(s)}\, Y_{k-1}\right)$$

holds asymptotically as $n \to \infty$ with probability one.

The next result shows that $\widehat{\alpha}_n$ and $\widehat{\mu}_{\varepsilon,n}$ are strongly consistent estimators of α and μ_ε, whereas $\widehat{\theta}_n$ fails to be a strongly consistent estimator of θ.

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

$$\mathsf{P}(\lim_{n\to\infty} \widehat{\alpha}_n = \alpha) = 1, \qquad \forall\, (\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}, \tag{4.3.6}$$

$$\mathsf{P}(\lim_{n\to\infty} \widehat{\mu}_{\varepsilon,n} = \mu_\varepsilon) = 1, \qquad \forall\, (\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}, \tag{4.3.7}$$

whereas the sequence $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ is not strongly consistent for any $(\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}$, namely,

$$\mathsf{P}\left(\lim_{n\to\infty} \widehat{\theta}_n = Y_s - \alpha Y_{s-1} - \mu_\varepsilon\right) = 1, \qquad \forall\, (\alpha, \mu_\varepsilon, \theta) \in (0,1)\times(0,\infty)\times\mathbb{N}. \tag{4.3.8}$$

Proof. By (4.3.2), (4.3.3) and Proposition 4.1.1, we get

$$\begin{bmatrix} \widehat{\alpha}_n \\ \widehat{\mu}_{\varepsilon,n} \end{bmatrix} = \frac{1}{D_n} \begin{bmatrix} K_n \\ L_n \end{bmatrix},$$

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 the very same arguments as in the proof of Theorem 4.2.1, 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 (4.3.5) and (2.2.3), (2.2.4), we obtain

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

and

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

where we used that

$$\begin{aligned}\frac{(1-\alpha)\mathsf{E}\widetilde{X}\mathsf{E}\widetilde{X}^2 - \mu_\varepsilon(\mathsf{E}\widetilde{X})^2}{\operatorname{Var}\widetilde{X}} &= \frac{1}{\operatorname{Var}\widetilde{X}}\left[(1-\alpha)\frac{\mu_\varepsilon}{1-\alpha}\left(\frac{\sigma_\varepsilon^2+\alpha\mu_\varepsilon}{1-\alpha^2} + \frac{\mu_\varepsilon^2}{(1-\alpha)^2}\right) - \mu_\varepsilon\frac{\mu_\varepsilon^2}{(1-\alpha)^2}\right]\\ &= \frac{\mu_\varepsilon}{\operatorname{Var}\widetilde{X}}\frac{\sigma_\varepsilon^2+\alpha\mu_\varepsilon}{1-\alpha^2} = \mu_\varepsilon.\end{aligned}$$

Finally, using (4.3.4), (4.3.6) and (4.3.7) we get (4.3.8). □

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

$$\begin{bmatrix} \sqrt{n}(\widehat{\alpha}_n - \alpha) \\ \sqrt{n}(\widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon) \end{bmatrix} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\ B_{\alpha,\varepsilon}\right) \qquad \text{as } n\to\infty, \tag{4.3.9}$$

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

$$\sqrt{n}(\widehat{\theta}_n - \lim_{k\to\infty}\widehat{\theta}_k) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, [Y_{s-1}\ 0]B_{\alpha,\varepsilon}[Y_{s-1}\ 0]^\top) \qquad \text{as } n\to\infty. \tag{4.3.10}$$

Proof. By (4.3.2) and (4.3.3), we have

$$\begin{bmatrix} \widehat{\alpha}_n - \alpha \\ \widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon \end{bmatrix} = \frac{1}{D_n} \begin{bmatrix} (n-1)\sum_{k=1}^{n}{}^{(s)}(Y_k - \alpha Y_{k-1})Y_{k-1} - \sum_{k=1}^{n}{}^{(s)}(Y_k - \alpha Y_{k-1}) \sum_{k=1}^{n}{}^{(s)} Y_{k-1} \\ \sum_{k=1}^{n}{}^{(s)} Y_{k-1}^2 \sum_{k=1}^{n}{}^{(s)}(Y_k - \mu_\varepsilon) - \sum_{k=1}^{n}{}^{(s)} Y_{k-1} \sum_{k=1}^{n}{}^{(s)}(Y_k - \mu_\varepsilon)Y_{k-1} \end{bmatrix},$$

holds asymptotically as $n \to \infty$ with probability one. By Proposition 4.1.1, we get

$$\begin{bmatrix} \widehat{\alpha}_n - \alpha \\ \widehat{\mu}_{\varepsilon,n} - \mu_\varepsilon \end{bmatrix} = \frac{1}{D_n} \begin{bmatrix} (n-1)\sum_{k=1}^{n}{}^{(s)}(X_k - \alpha X_{k-1})X_{k-1} - \sum_{k=1}^{n}{}^{(s)}(X_k - \alpha X_{k-1}) \sum_{k=1}^{n}{}^{(s)} X_{k-1} + R_n \\ \sum_{k=1}^{n}{}^{(s)} X_{k-1}^2 \sum_{k=1}^{n}{}^{(s)}(X_k - \mu_\varepsilon) - \sum_{k=1}^{n}{}^{(s)} X_{k-1} \sum_{k=1}^{n}{}^{(s)}(X_k - \mu_\varepsilon)X_{k-1} + Q_n \end{bmatrix},$$

holds asymptotically as $n \to \infty$ with probability one, 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

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

By (4.3.5), (2.3.1) and Slutsky's lemma, we have

$$\frac{\sqrt{n}}{D_n} \begin{bmatrix} n\sum_{k=1}^{n}{}^{(s)}(X_k - \alpha X_{k-1})X_{k-1} - \sum_{k=1}^{n}{}^{(s)}(X_k - \alpha X_{k-1}) \sum_{k=1}^{n}{}^{(s)} X_{k-1} \\ \sum_{k=1}^{n}{}^{(s)} X_{k-1}^2 \sum_{k=1}^{n}{}^{(s)}(X_k - \mu_\varepsilon) - \sum_{k=1}^{n}{}^{(s)} X_{k-1} \sum_{k=1}^{n}{}^{(s)}(X_k - \mu_\varepsilon)X_{k-1} \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\ B_{\alpha,\varepsilon} \right),$$

as $n \to \infty$, and hence to prove (4.3.9), by (4.3.5) and Slutsky's lemma, it is enough to check that

(4.3.11) $$\frac{R_n}{n^{3/2}} \xrightarrow{\mathsf{P}} 0 \qquad \text{as } n \to \infty,$$

and

(4.3.12) $$\frac{Q_n}{n^{3/2}} \xrightarrow{\mathsf{P}} 0 \qquad \text{as } n \to \infty,$$

where $\xrightarrow{\mathsf{P}}$ denotes convergence in probability. By (4.2.15) and (4.2.16), to prove (4.3.11) it remains to check that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} (X_k - \alpha X_{k-1}) Z_{k-1} \xrightarrow{\mathsf{P}} 0 \qquad \text{as} \quad n \to \infty, \tag{4.3.13}$$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} (Z_k - \alpha Z_{k-1}) \cdot \frac{1}{n} \sum_{k=1}^{n}{}^{(s)} (X_{k-1} + Z_{k-1}) \xrightarrow{\mathsf{P}} 0 \qquad \text{as} \quad n \to \infty, \tag{4.3.14}$$

$$\frac{1}{n} \sum_{k=1}^{n}{}^{(s)} (X_k - \alpha X_{k-1}) \cdot \frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} Z_{k-1} \xrightarrow{\mathsf{P}} 0 \qquad \text{as} \quad n \to \infty. \tag{4.3.15}$$

Using (4.2.14) and that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} (X_k - \alpha X_{k-1}) Z_{k-1} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} (X_k - \alpha X_{k-1} - \mu_\varepsilon) Z_{k-1} + \mu_\varepsilon \frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} Z_{k-1},$$

to prove (4.3.13), it is enough to check that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} Z_{k-1} \xrightarrow{\mathsf{P}} 0 \qquad \text{as} \quad n \to \infty. \tag{4.3.16}$$

Using that $Z_k \geqslant 0$, $k \in \mathbb{N}$, by Markov's inequality, it is enough to check that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} \mathsf{E} Z_{k-1} = 0. \tag{4.3.17}$$

Since, by (4.1.4), $\mathsf{E} Z_{s+k} = \theta \alpha^k$, $k \geqslant 0$, we have

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} \mathsf{E} Z_{k-1} \leqslant \frac{\theta}{\sqrt{n}} \sum_{k=0}^{n} \alpha^k = \frac{\theta}{\sqrt{n}} \frac{\alpha^{n+1} - 1}{\alpha - 1} \to 0 \qquad \text{as} \quad n \to \infty. \tag{4.3.18}$$

Using that

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

to prove (4.3.14) it is enough to check that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} (Z_k - \alpha Z_{k-1}) \xrightarrow{L_1} 0 \qquad \text{as} \quad n \to \infty. \tag{4.3.19}$$

To verify (4.3.19) it is enough to show that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} \mathsf{E}(Z_k - \alpha Z_{k-1}) = 0, \tag{4.3.20}$$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left(\sum_{k=1}^{n}{}^{(s)} (Z_k - \alpha Z_{k-1}) \right)^2 = 0. \tag{4.3.21}$$

Using that $\mathsf{E}Z_k = \alpha \mathsf{E}Z_{k-1}$, $k \geqslant s+1$, we get (4.3.20) is satisfied. Using that $\mathsf{E}[(Z_k - \alpha Z_{k-1})(Z_\ell - \alpha Z_{\ell-1})] = 0$ for all $k \neq \ell$, $k, \ell \geqslant s+1$, we have

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

as we showed in the proof of Theorem 4.2.2. Using that

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

to prove (4.3.15) it is enough to verify (4.3.16) which was already done.

Now we turn to prove (4.3.12). Using (4.3.16) and that

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

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

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

it is enough to verify that

$$\text{(4.3.22)} \qquad \frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}\,X_{k-1}Z_{k-1} \stackrel{\mathsf{P}}{\longrightarrow} 0 \qquad \text{as} \ \ n \to \infty,$$

$$\text{(4.3.23)} \qquad \frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}\,Z_{k-1}^2 \stackrel{\mathsf{P}}{\longrightarrow} 0 \qquad \text{as} \ \ n \to \infty,$$

$$\text{(4.3.24)} \qquad \frac{1}{\sqrt{n}}\sum_{k=1}^{n}{}^{(s)}\,Z_{k-1}Z_k \stackrel{\mathsf{P}}{\longrightarrow} 0 \qquad \text{as} \ \ n \to \infty.$$

To prove (4.3.22), using that the processes X and Z are non-negative, by Markov's inequality, it is enough to verify 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 $\lim_{k\to\infty}\mathsf{E}X_{k-1} = \mathsf{E}\widetilde{X}$, as in the proof of (4.2.15), we get it is enough to check that

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

which follows by (4.3.18). Similarly, to prove (4.3.23), it is enough to show that

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

By (4.1.5), we have for all $n \geqslant s+1$,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} \mathsf{E} Z_{k-1}^2 \leqslant \frac{1}{\sqrt{n}} \sum_{k=0}^{n} (\theta^2 \alpha^{2k} - \theta \alpha^k (\alpha^k - 1)) \leqslant \frac{\theta^2}{\sqrt{n}} \frac{\alpha^{2(n+1)} - 1}{\alpha^2 - 1} + \frac{\theta}{\sqrt{n}} \frac{\alpha^{n+1} - 1}{\alpha - 1} \to 0,$$

as $n \to \infty$. Similarly, to prove (4.3.24), it is enough to check that

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

By (4.1.6), we have for all $n \geqslant s+1$,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n}{}^{(s)} \mathsf{E}(Z_{k-1} Z_k) \leqslant \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (\theta^2 \alpha^{2k-1} + \theta \alpha^k (1 - \alpha^k)) \leqslant \frac{\theta^2}{\alpha\sqrt{n}} \frac{\alpha^{2n} - 1}{\alpha^2 - 1} + \frac{\theta}{\sqrt{n}} \frac{\alpha^n - 1}{\alpha - 1} \to 0,$$

as $n \to \infty$.

Finally, using (4.3.4) and (4.3.8), 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 (4.3.9), we have (4.3.10). □

4.4 Two outliers, estimation of the mean of the offspring distribution and the outliers' sizes

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 $n \geqslant \max(s_1, s_2)$,

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

For all $n \geqslant \max(s_1, s_2)$, $n \in \mathbb{N}$, we define the function $Q_n : \mathbb{R}^{n+1} \times \mathbb{R}^3 \to \mathbb{R}$, as

$$Q_n(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$$
$$:= \sum_{k=1}^{n}{}^{(s_1,s_2)} \big(y_k - \alpha' y_{k-1} - \mu_\varepsilon\big)^2 + \big(y_{s_1} - \alpha' y_{s_1-1} - \mu_\varepsilon - \theta_1'\big)^2 + \big(y_{s_2} - \alpha' y_{s_2-1} - \mu_\varepsilon - \theta_2'\big)^2,$$

for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $\alpha', \theta'_1, \theta'_2 \in \mathbb{R}$. By definition, for all $n \geqslant \max(s_1, s_2)$, a CLS estimator for the parameter $(\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2$ is a measurable function $(\widetilde{\alpha}_n, \widetilde{\theta}_{1,n}, \widetilde{\theta}_{2,n}) : S_n \to \mathbb{R}^3$ such that

$$Q_n(\mathbf{y}_n; \widetilde{\alpha}_n(\mathbf{y}_n), \widetilde{\theta}_{1,n}(\mathbf{y}_n), \widetilde{\theta}_{2,n}(\mathbf{y}_n)) = \inf_{(\alpha',\theta'_1,\theta'_2)\in\mathbb{R}^3} Q_n(\mathbf{y}_n; \alpha', \theta'_1, \theta'_2) \qquad \forall\ \mathbf{y}_n \in S_n,$$

where S_n is suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 4.4.1). We note that we do not define the CLS estimator $(\widetilde{\alpha}_n, \widetilde{\theta}_{1,n}, \widetilde{\theta}_{2,n})$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$. For all $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $\alpha', \theta'_1, \theta'_2 \in \mathbb{R}$,

$$\begin{aligned}
\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \theta'_1, \theta'_2) &= \sum_{k=1}^{n\ (s_1,s_2)} (-2)\big(y_k - \alpha' y_{k-1} - \mu_\varepsilon\big) y_{k-1} - 2\big(y_{s_1} - \alpha' y_{s_1-1} - \mu_\varepsilon - \theta'_1\big) y_{s_1-1} \\
&\qquad - 2\big(y_{s_2} - \alpha' y_{s_2-1} - \mu_\varepsilon - \theta'_2\big) y_{s_2-1}, \\
\frac{\partial Q_n}{\partial \theta'_1}(\mathbf{y}_n; \alpha', \theta'_1, \theta'_2) &= -2\big(y_{s_1} - \alpha' y_{s_1-1} - \mu_\varepsilon - \theta'_1\big), \\
\frac{\partial Q_n}{\partial \theta'_2}(\mathbf{y}_n; \alpha', \theta'_1, \theta'_2) &= -2\big(y_{s_2} - \alpha' y_{s_2-1} - \mu_\varepsilon - \theta'_2\big).
\end{aligned}$$

The next lemma is about the existence and uniqueness of the CLS estimator of $(\alpha, \theta_1, \theta_2)$.

4.4.1 Lemma. *There exist subsets* $S_n \subset \mathbb{R}^{n+1}$, $n \geqslant \max(3, s_1, s_2)$ *with the following properties:*

(i) *there exists a unique CLS estimator* $(\widetilde{\alpha}_n, \widetilde{\theta}_{1,n}, \widetilde{\theta}_{2,n}) : S_n \to \mathbb{R}^3$,

(ii) *for all* $\mathbf{y}_n \in S_n$, *the system of equations*

$$\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \theta'_1, \theta'_2) = 0, \quad \frac{\partial Q_n}{\partial \theta'_1}(\mathbf{y}_n; \alpha', \theta'_1, \theta'_2) = 0, \quad \frac{\partial Q_n}{\partial \theta'_2}(\mathbf{y}_n; \alpha', \theta'_1, \theta'_2) = 0, \tag{4.4.1}$$

has the unique solution

$$\widetilde{\alpha}_n(\mathbf{y}_n) = \frac{\sum_{k=1}^{n\ (s_1,s_2)} (y_k - \mu_\varepsilon) y_{k-1}}{\sum_{k=1}^{n\ (s_1,s_2)} y_{k-1}^2}, \tag{4.4.2}$$

$$\widetilde{\theta}_{i,n}(\mathbf{y}_n) = y_{s_i} - \widetilde{\alpha}_n(\mathbf{y}_n) y_{s_i-1} - \mu_\varepsilon, \qquad i = 1, 2, \tag{4.4.3}$$

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

Proof. One can easily check that the unique solution of the system of equations (4.4.1) takes the form (4.4.2) and (4.4.3) whenever $\sum_{k=1}^{n}{}^{(s_1,s_2)} y_{k-1}^2 > 0$.

For all $n \geqslant \max(3, s_1, s_2)$, let

$$S_n := \left\{ \mathbf{y}_n \in \mathbb{R}^{n+1} : \Delta_{i,n}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') > 0, \quad i = 1, 2, 3, \ \forall\, (\alpha', \theta_1', \theta_2') \in \mathbb{R}^3 \right\},$$

where $\Delta_{i,n}(\mathbf{y}_n; \alpha', \mu_\varepsilon', \theta')$, $i = 1, 2, 3$, denotes the i-th order leading principal minor of the 3×3 matrix

$$H_n(\mathbf{y}_n; \alpha', \theta_1', \theta_2') := \begin{bmatrix} \frac{\partial^2 Q_n}{\partial (\alpha')^2} & \frac{\partial^2 Q_n}{\partial \theta_1' \partial \alpha'} & \frac{\partial^2 Q_n}{\partial \theta_2' \partial \alpha'} \\ \frac{\partial^2 Q_n}{\partial \alpha' \partial \theta_1'} & \frac{\partial^2 Q_n}{\partial (\theta_1')^2} & \frac{\partial^2 Q_n}{\partial \theta_2' \partial \theta_1'} \\ \frac{\partial^2 Q_n}{\partial \alpha' \partial \theta_2'} & \frac{\partial^2 Q_n}{\partial \theta_1' \partial \theta_2'} & \frac{\partial^2 Q_n}{\partial (\theta_2')^2} \end{bmatrix} (\mathbf{y}_n; \alpha', \theta_1', \theta_2')$$

Then the function $\mathbb{R}^3 \ni (\alpha', \theta_1', \theta_2') \mapsto Q_n(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$ is strictly convex for all $\mathbf{y}_n \in S_n$, see, e.g., Berkovitz [10, Theorem 3.3, Chapter III]. Further, for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $(\alpha', \theta_1', \theta_2') \in \mathbb{R}^3$, we have

$$\frac{\partial^2 Q_n}{\partial (\alpha')^2}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 2 \sum_{k=1}^{n}{}^{(s_1,s_2)} y_{k-1}^2 + 2y_{s_1-1}^2 + 2y_{s_2-1}^2 = 2 \sum_{k=1}^{n} y_{k-1}^2,$$

$$\frac{\partial^2 Q_n}{\partial \alpha' \partial \theta_1'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = \frac{\partial^2 Q_n}{\partial \theta_1' \partial \alpha'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 2y_{s_1-1},$$

$$\frac{\partial^2 Q_n}{\partial \alpha' \partial \theta_2'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = \frac{\partial^2 Q_n}{\partial \theta_2' \partial \alpha'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 2y_{s_2-1},$$

$$\frac{\partial^2 Q_n}{\partial \theta_1' \partial \theta_2'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = \frac{\partial^2 Q_n}{\partial \theta_2' \partial \theta_1'}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 0,$$

$$\frac{\partial^2 Q_n}{\partial (\theta_1')^2}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 2, \qquad \frac{\partial^2 Q_n}{\partial (\theta_2')^2}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 2.$$

This yields that the system of equations (4.4.1) has a unique solution for all $\mathbf{y}_n \in S_n$. Using also that the function $\mathbb{R}^3 \ni (\alpha', \theta_1', \theta_2') \mapsto Q_n(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$ is strictly convex for all $\mathbf{y}_n \in S_n$, we get the function in question attains its (global) minimum at this unique solution, which yields (i) and (ii).

Then $H_n(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$ has the following leading principal minors

$$\Delta_{1,n}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 2 \sum_{k=1}^{n} y_{k-1}^2, \qquad \Delta_{2,n}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 4 \left(\sum_{k=1}^{n} y_{k-1}^2 - (y_{s_1-1})^2 \right),$$

and

$$\Delta_{3,n}(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = \det H_n(\mathbf{y}_n; \alpha', \theta_1', \theta_2') = 8 \left(\sum_{k=1}^{n} y_{k-1}^2 - (y_{s_1-1})^2 - (y_{s_2-1})^2 \right).$$

Note that $\Delta_{i,n}(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$, $i = 1, 2, 3$, do not depend on $(\alpha', \theta_1', \theta_2')$, and hence we will simply denote $\Delta_{i,n}(\mathbf{y}_n; \alpha', \theta_1', \theta_2')$ by $\Delta_{i,n}(\mathbf{y}_n)$.

Next we check that $\mathbf{Y}_n \in S_n$ holds asymptotically as $n \to \infty$ with probability one. By (2.2.5) and (2.2.6), using the very same arguments as in the proof of Lemma 4.2.1, one can get

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\Delta_{1,n}(\mathbf{Y}_n)}{n} = 2\mathsf{E}\widetilde{X}^2\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\Delta_{2,n}(\mathbf{Y}_n)}{n} = 4\mathsf{E}\widetilde{X}^2\right) = 1,$$

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{\Delta_{3,n}(\mathbf{Y}_n)}{n} = 8\mathsf{E}\widetilde{X}^2\right) = 1,$$

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

$$\mathsf{P}\left(\lim_{n\to\infty} \Delta_{i,n}(\mathbf{Y}_n) = \infty\right) = 1, \qquad i = 1, 2, 3,$$

which yields that $\mathbf{Y}_n \in S_n$ holds asymptotically as $n \to \infty$ with probability one. □

By Lemma 4.4.1, $(\widetilde{\alpha}_n(\mathbf{Y}_n), \widetilde{\theta}_{1,n}(\mathbf{Y}_n), \widetilde{\theta}_{2,n}(\mathbf{Y}_n))$ exists uniquely asymptotically as $n \to \infty$ with probability one. In the sequel we will simply denote it by $(\widetilde{\alpha}_n, \widetilde{\theta}_{1,n}, \widetilde{\theta}_{2,n})$.

The next result shows that $\widetilde{\alpha}_n$ is a strongly consistent estimator of α, whereas $\widetilde{\theta}_{i,n}$, $i = 1, 2$, fail to be strongly consistent estimators of $\theta_{i,n}$, $i = 1, 2$, respectively.

4.4.1 Theorem. *Consider the CLS estimators* $(\widetilde{\alpha}_n, \widetilde{\theta}_{1,n}, \widetilde{\theta}_{2,n})_{n\in\mathbb{N}}$ *of the parameter* $(\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2$. *The sequence* $(\widetilde{\alpha}_n)_{n\in\mathbb{N}}$ *is strongly consistent for all* $(\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2$, *i.e.,*

$$\mathsf{P}(\lim_{n\to\infty} \widetilde{\alpha}_n = \alpha) = 1, \qquad \forall\, (\alpha, \theta_1, \theta_2) \in (0,1) \times \mathbb{N}^2, \tag{4.4.4}$$

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

$$\mathsf{P}\left(\lim_{n\to\infty} \widetilde{\theta}_{i,n} = Y_{s_i} - \alpha Y_{s_i - 1} - \mu_\varepsilon\right) = 1, \qquad \forall\, (\alpha, \theta_1, \theta_1) \in (0,1) \times \mathbb{N}^2, \qquad i = 1, 2. \tag{4.4.5}$$

Proof. By (4.4.2) and Proposition 4.1.2, we get

$$\widetilde{\alpha}_n = \frac{\sum_{k=1}^{n}{}^{(s_1,s_2)} (X_k - \mu_\varepsilon) X_{k-1} + K_n}{\sum_{k=1}^{n}{}^{(s_1,s_2)} X_{k-1}^2 + L_n},$$

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

$$K_n := \sum_{k=1}^{n}{}^{(s_1,s_2)} (Z_k^{(1)} + Z_k^{(2)})(X_{k-1} + Z_{k-1}^{(1)} + Z_{k-1}^{(2)}) + \sum_{k=1}^{n}{}^{(s_1,s_2)} (X_k - \mu_\varepsilon)(Z_{k-1}^{(1)} + Z_{k-1}^{(2)}),$$

$$L_n := \sum_{k=1}^{n}{}^{(s_1,s_2)} \left[(Z_{k-1}^{(1)})^2 + (Z_{k-1}^{(2)})^2 + 2X_{k-1}Z_{k-1}^{(1)} + 2X_{k-1}Z_{k-1}^{(2)} + 2Z_{k-1}^{(1)}Z_{k-1}^{(2)}\right].$$

Using the very same arguments as in the proof of Theorem 4.2.1, one can obtain

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{K_n}{n}=0\right)=1 \qquad \text{and} \qquad \mathsf{P}\left(\lim_{n\to\infty}\frac{L_n}{n}=0\right)=1.$$

Indeed, the only fact that was not verified in the proof of Theorem 4.2.1 is that

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}{}^{(s_1,s_2)}\, Z_k^{(1)}Z_k^{(2)}=0\right)=1. \tag{4.4.6}$$

By Cauchy-Schwartz's inequality and Proposition 4.1.2,

$$\frac{1}{n}\sum_{k=1}^{n}{}^{(s_1,s_2)}\, Z_k^{(1)}Z_k^{(2)} \leqslant \sqrt{\frac{1}{n}\sum_{k=1}^{n}(Z_k^{(1)})^2\frac{1}{n}\sum_{k=1}^{n}(Z_k^{(2)})^2} \to 0 \qquad \text{as } n\to\infty,$$

with probability one.

Finally, by (4.4.3) and (4.4.4), we have (4.4.5). □

4.4.1 Remark. Since

$$\begin{bmatrix} Y_{s_1}-\alpha Y_{s_1-1}-\mu_\varepsilon \\ Y_{s_2}-\alpha Y_{s_2-1}-\mu_\varepsilon \end{bmatrix} = \begin{bmatrix} X_{s_1}-\alpha X_{s_1-1}-\mu_\varepsilon+\theta_1 \\ X_{s_2}-\alpha X_{s_2-1}-\mu_\varepsilon+\theta_2 \end{bmatrix},$$

and

$$\begin{aligned}
&\operatorname{Cov}(X_{s_1}-\alpha X_{s_1-1}-\mu_\varepsilon+\theta_1, X_{s_2}-\alpha X_{s_2-1}-\mu_\varepsilon+\theta_2) \\
&\quad= \mathsf{E}\big[(X_{s_1}-\alpha X_{s_1-1}-\mu_\varepsilon)(X_{s_2}-\alpha X_{s_2-1}-\mu_\varepsilon)\big] \\
&\quad= \mathsf{E}\left[\left(\sum_{j=1}^{X_{s_1-1}}(\xi_{s_1,j}-\alpha)+(\varepsilon_{s_1}-\mu_\varepsilon)\right)\left(\sum_{j=1}^{X_{s_2-1}}(\xi_{s_2,j}-\alpha)+(\varepsilon_{s_2}-\mu_\varepsilon)\right)\right]=0,
\end{aligned}$$

by Remark 4.2.1, we get

$$\begin{aligned}
&\operatorname{Var}\begin{bmatrix} Y_{s_1}-\alpha Y_{s_1-1}-\mu_\varepsilon \\ Y_{s_2}-\alpha Y_{s_2-1}-\mu_\varepsilon \end{bmatrix} \\
&\quad= \begin{bmatrix} \alpha^{s_1}(1-\alpha)\mathsf{E}Y_0+\alpha\mu_\varepsilon(1-\alpha^{s_1-1})+\sigma_\varepsilon^2 & 0 \\ 0 & \alpha^{s_2}(1-\alpha)\mathsf{E}Y_0+\alpha\mu_\varepsilon(1-\alpha^{s_2-1})+\sigma_\varepsilon^2 \end{bmatrix}.
\end{aligned}$$

□

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

$$\sqrt{n}(\widetilde{\alpha}_n-\alpha) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma_{\alpha,\varepsilon}^2) \qquad \text{as } n\to\infty, \tag{4.4.7}$$

where $\sigma_{\alpha,\varepsilon}^2$ *is defined in* (2.2.9). *Moreover, conditionally on the values* Y_{s_1-1} *and* Y_{s_2-1},

(4.4.8)

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

Proof. By (4.4.2), we get

$$\widetilde{\alpha}_n - \alpha = \frac{\sum_{k=1}^{n}{}^{(s_1,s_2)}(Y_k - \alpha Y_{k-1} - \mu_\varepsilon)Y_{k-1}}{\sum_{k=1}^{n}{}^{(s_1,s_2)} Y_{k-1}^2},$$

holds asymptotically as $n \to \infty$ with probability one. To prove (4.4.7), using Proposition 4.1.2 and (4.2.14)–(4.2.18), it is enough to check that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (Z_k^{(1)} - \alpha Z_{k-1}^{(1)}) Z_{k-1}^{(2)} \xrightarrow{\mathsf{P}} 0 \qquad \text{as } n \to \infty, \tag{4.4.9}$$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (Z_k^{(2)} - \alpha Z_{k-1}^{(2)}) Z_{k-1}^{(1)} \xrightarrow{\mathsf{P}} 0 \qquad \text{as } n \to \infty. \tag{4.4.10}$$

To prove (4.4.9) it is enough to verify that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathsf{E}[(Z_k^{(1)} - \alpha Z_{k-1}^{(1)}) Z_{k-1}^{(2)}] = 0, \tag{4.4.11}$$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E} \left(\sum_{k=1}^{n} (Z_k^{(1)} - \alpha Z_{k-1}^{(1)}) Z_{k-1}^{(2)} \right)^2 = 0. \tag{4.4.12}$$

Since the processes $Z^{(1)}$ and $Z^{(2)}$ are independent, we have

$$\mathsf{E}[(Z_k^{(1)} - \alpha Z_{k-1}^{(1)}) Z_{k-1}^{(2)}] = \mathsf{E}(Z_k^{(1)} - \alpha Z_{k-1}^{(1)}) \mathsf{E} Z_{k-1}^{(2)} = 0, \qquad k \in \mathbb{N},$$

which yields (4.4.11). Using that for all $k, \ell \in \mathbb{N}$, $k > \ell$,

$$\begin{aligned}
\mathsf{E}\big[(Z_k^{(1)} - \alpha Z_{k-1}^{(1)}) Z_{k-1}^{(2)} (Z_\ell^{(1)} - \alpha Z_{\ell-1}^{(1)}) Z_{\ell-1}^{(2)}\big] &= \mathsf{E}\big[(Z_k^{(1)} - \alpha Z_{k-1}^{(1)})(Z_\ell^{(1)} - \alpha Z_{\ell-1}^{(1)})\big] \mathsf{E}(Z_{k-1}^{(2)} Z_{\ell-1}^{(2)}) \\
&= \mathsf{E}\big[(Z_\ell^{(1)} - \alpha Z_{\ell-1}^{(1)}) \mathsf{E}(Z_k^{(1)} - \alpha Z_{k-1}^{(1)} \,|\, \mathcal{F}_{k-1}^{Z^{(1)}})\big] \mathsf{E}(Z_{k-1}^{(2)} Z_{\ell-1}^{(2)}) \\
&= 0 \cdot \mathsf{E}(Z_{k-1}^{(2)} Z_{\ell-1}^{(2)}) = 0,
\end{aligned}$$

we get

$$\begin{aligned}
\frac{1}{n} \mathsf{E} \left(\sum_{k=1}^{n} (Z_k^{(1)} - \alpha Z_{k-1}^{(1)}) Z_{k-1}^{(2)} \right)^2 &= \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}[(Z_k^{(1)} - \alpha Z_{k-1}^{(1)})^2 (Z_{k-1}^{(2)})^2] \\
&= \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}(Z_k^{(1)} - \alpha Z_{k-1}^{(1)})^2 \mathsf{E}(Z_{k-1}^{(2)})^2 \\
&\leqslant \frac{2}{n} \sum_{k=1}^{n} \mathsf{E}((Z_k^{(1)})^2 + \alpha^2 (Z_{k-1}^{(1)})^2) \mathsf{E}(Z_{k-1}^{(2)})^2.
\end{aligned}$$

Hence, by (4.1.5),

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

$$\leqslant \frac{2}{n}\sum_{k=0}^{n}\Bigg[\Big(\theta_1^2\alpha^{2k}+\theta_1\alpha^k(1-\alpha^k)+\alpha^2\theta_1^2\alpha^{2(k-1)}+\alpha^2\theta_1\alpha^{k-1}(1-\alpha^{k-1})\Big)$$

$$\times\Big(\theta_2^2\alpha^{2(k-1)}+\theta_2\alpha^{k-1}(1-\alpha^{k-1})\Big)\Bigg]$$

$$\leqslant \frac{2(\theta_1^2+\theta_1)}{n}\Big(\theta_2^2\frac{\alpha^{2n}-1}{\alpha^2-1}+\theta_2\frac{\alpha^n-1}{\alpha-1}+\theta_2^2\frac{\alpha^{2n}-1}{\alpha^2-1}+\alpha\theta_2\frac{\alpha^n-1}{\alpha-1}\Big)\to 0 \qquad \text{as } n\to\infty.$$

Similarly one can check (4.4.10).

Moreover, conditionally on the values Y_{s_1-1} and Y_{s_2-1}, by (4.4.3), (4.4.5) and (4.4.7),

$$\begin{bmatrix}\sqrt{n}\big(\widetilde{\theta}_{1,n}-\lim_{k\to\infty}\widetilde{\theta}_{1,k}\big)\\ \sqrt{n}\big(\widetilde{\theta}_{2,n}-\lim_{k\to\infty}\widetilde{\theta}_{2,k}\big)\end{bmatrix}=\sqrt{n}\begin{bmatrix}\widetilde{\theta}_{1,n}-(Y_{s_1}-\alpha Y_{s_1-1}-\mu_\varepsilon)\\ \widetilde{\theta}_{2,n}-(Y_{s_2}-\alpha Y_{s_2-1}-\mu_\varepsilon)\end{bmatrix}=\sqrt{n}\begin{bmatrix}-(\widetilde{\alpha}_n-\alpha)Y_{s_1-1}\\ -(\widetilde{\alpha}_n-\alpha)Y_{s_2-1}\end{bmatrix}$$

$$\stackrel{\mathcal{L}}{\longrightarrow}\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\sigma_{\alpha,\varepsilon}^2\begin{bmatrix}Y_{s_1-1}^2 & Y_{s_1-1}Y_{s_2-1}\\ Y_{s_1-1}Y_{s_2-1} & Y_{s_2-1}^2\end{bmatrix}\right) \qquad \text{as } n\to\infty.$$

□

4.5 Two outliers, estimation of the mean of the offspring and innovation distributions and the outliers' sizes

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 $(\alpha,\mu_\varepsilon,\theta_1,\theta_2)$. 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 $n\geqslant\max(s_1,s_2)$, $n\in\mathbb{N}$,

$$\sum_{k=1}^{n}\big(Y_k-\mathsf{E}(Y_k\,|\,\mathcal{F}_{k-1}^Y)\big)^2$$

$$=\sum_{k=1}^{n}{}^{(s_1,s_2)}\big(Y_k-\alpha Y_{k-1}-\mu_\varepsilon\big)^2+\big(Y_{s_1}-\alpha Y_{s_1-1}-\mu_\varepsilon-\theta_1\big)^2+\big(Y_{s_2}-\alpha Y_{s_2-1}-\mu_\varepsilon-\theta_2\big)^2.$$

For all $n\geqslant\max(s_1,s_2)$, $n\in\mathbb{N}$, we define the function $Q_n:\mathbb{R}^{n+1}\times\mathbb{R}^4\to\mathbb{R}$, as

$$Q_n(\mathbf{y}_n;\alpha',\mu_\varepsilon',\theta_1',\theta_2')$$

$$:=\sum_{k=1}^{n}{}^{(s_1,s_2)}\big(y_k-\alpha' y_{k-1}-\mu_\varepsilon'\big)^2+\big(y_{s_1}-\alpha' y_{s_1-1}-\mu_\varepsilon'-\theta_1'\big)^2+\big(y_{s_2}-\alpha' y_{s_2-1}-\mu_\varepsilon'-\theta_2'\big)^2,$$

for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2 \in \mathbb{R}$. By definition, for all $n \geqslant \max(s_1, s_2)$, 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 $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_{1,n}, \widehat{\theta}_{2,n}) : S_n \to \mathbb{R}^4$ such that

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

where S_n is suitable subset of $\mathbb{R}^{n+1}$ (defined in the proof of Lemma 4.5.1). We note that we do not define the CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_{1,n}, \widehat{\theta}_{2,n})$ for all samples $\mathbf{y}_n \in \mathbb{R}^{n+1}$. For all $\mathbf{y}_n \in \mathbb{R}^{n+1}$, $\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2 \in \mathbb{R}$,

$$\begin{aligned}
&\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \\
&\qquad = -2\sum_{k=1}^{n}{}^{(s_1,s_2)} \big(y_k - \alpha' y_{k-1} - \mu'_\varepsilon\big) y_{k-1} - 2\big(y_{s_1} - \alpha' y_{s_1-1} - \mu'_\varepsilon - \theta'_1\big) y_{s_1-1} \\
&\qquad\quad - 2\big(y_{s_2} - \alpha' y_{s_2-1} - \mu'_\varepsilon - \theta'_2\big) y_{s_2-1}, \\
&\frac{\partial Q_n}{\partial \mu'_\varepsilon}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = -2\sum_{k=1}^{n}{}^{(s_1,s_2)} \big(y_k - \alpha' y_{k-1} - \mu'_\varepsilon\big) - 2\big(y_{s_1} - \alpha' y_{s_1-1} - \mu'_\varepsilon - \theta'_1\big) \\
&\qquad\qquad\qquad - 2\big(y_{s_2} - \alpha' y_{s_2-1} - \mu'_\varepsilon - \theta'_2\big), \\
&\frac{\partial Q_n}{\partial \theta'_1}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = -2\big(y_{s_1} - \alpha' y_{s_1-1} - \mu'_\varepsilon - \theta'_1\big), \\
&\frac{\partial Q_n}{\partial \theta'_2}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = -2\big(y_{s_2} - \alpha' y_{s_2-1} - \mu'_\varepsilon - \theta'_2\big).
\end{aligned}$$

The next lemma is about the existence and uniqueness of the CLS estimator of $(\alpha, \mu_\varepsilon, \theta_1, \theta_2)$.

4.5.1 Lemma. *There exist subsets $S_n \subset \mathbb{R}^{n+1}$, $n \geqslant \max(s_1, s_2)$ with the following properties:*

(i) *there exists a unique CLS estimator $(\widehat{\alpha}_n, \widehat{\mu}_{\varepsilon,n}, \widehat{\theta}_{1,n}, \widehat{\theta}_{2,n}) : S_n \to \mathbb{R}^4$,*

(ii) *for all $\mathbf{y}_n \in S_n$, the system of equations*

$$\begin{aligned}
&\frac{\partial Q_n}{\partial \alpha'}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 0, \qquad && \frac{\partial Q_n}{\partial \mu'_\varepsilon}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 0, \\
&\frac{\partial Q_n}{\partial \theta'_1}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 0, && \frac{\partial Q_n}{\partial \theta'_2}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 0,
\end{aligned} \tag{4.5.1}$$

has the unique solution

$$(4.5.2) \qquad \widehat{\alpha}_n(\mathbf{y}_n) = \frac{(n-2)\sum_{k=1}^{n}{}^{(s_1,s_2)} y_{k-1}y_k - \sum_{k=1}^{n}{}^{(s_1,s_2)} y_k \sum_{k=1}^{n}{}^{(s_1,s_2)} y_{k-1}}{D_n(\mathbf{y}_n)},$$

$$(4.5.3) \qquad \widehat{\mu}_{\varepsilon,n}(\mathbf{y}_n) = \frac{\sum_{k=1}^{n}{}^{(s_1,s_2)} y_{k-1}^2 \sum_{k=1}^{n}{}^{(s_1,s_2)} y_k - \sum_{k=1}^{n}{}^{(s_1,s_2)} y_{k-1} \sum_{k=1}^{n}{}^{(s_1,s_2)} y_{k-1}y_k}{D_n(\mathbf{y}_n)},$$

$$(4.5.4) \qquad \widehat{\theta}_{i,n}(\mathbf{y}_n) = y_{s_i} - \widehat{\alpha}_n(\mathbf{y}_n)y_{s_i-1} - \widehat{\mu}_{\varepsilon,n}(\mathbf{y}_n), \qquad i = 1, 2,$$

where

$$D_n(\mathbf{y}_n) := (n-2)\sum_{k=1}^{n}{}^{(s_1,s_2)} y_{k-1}^2 - \left(\sum_{k=1}^{n}{}^{(s_1,s_2)} y_{k-1}\right)^2, \qquad n \geqslant \max(s_1, s_2),$$

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

Proof. One can easily check that the unique solution of the system of equations (4.5.1) takes the form (4.5.2)-(4.5.3)-(4.5.4) whenever $D_n(\mathbf{y}_n) > 0$.

For all $n \geqslant \max(s_1, s_2)$, let

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

where $\Delta_{i,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$, $i = 1, 2, 3, 4$, denotes the i-th order leading principal minor of the 4×4 matrix

$$H_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) := \begin{bmatrix} \frac{\partial^2 Q_n}{\partial(\alpha')^2} & \frac{\partial^2 Q_n}{\partial\mu'_\varepsilon\partial\alpha'} & \frac{\partial^2 Q_n}{\partial\theta'_1\partial\alpha'} & \frac{\partial^2 Q_n}{\partial\theta'_2\partial\alpha'} \\ \frac{\partial^2 Q_n}{\partial\alpha'\partial\mu'_\varepsilon} & \frac{\partial^2 Q_n}{\partial(\mu'_\varepsilon)^2} & \frac{\partial^2 Q_n}{\partial\theta'_1\partial\mu'_\varepsilon} & \frac{\partial^2 Q_n}{\partial\theta'_2\partial\mu'_\varepsilon} \\ \frac{\partial^2 Q_n}{\partial\alpha'\partial\theta'_1} & \frac{\partial^2 Q_n}{\partial\mu'_\varepsilon\partial\theta'_1} & \frac{\partial^2 Q_n}{\partial(\theta'_1)^2} & \frac{\partial^2 Q_n}{\partial\theta'_2\partial\theta'_1} \\ \frac{\partial^2 Q_n}{\partial\alpha'\partial\theta'_2} & \frac{\partial^2 Q_n}{\partial\mu'_\varepsilon\partial\theta'_2} & \frac{\partial^2 Q_n}{\partial\theta'_1\partial\theta'_2} & \frac{\partial^2 Q_n}{\partial(\theta'_2)^2} \end{bmatrix} (\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2).$$

Then the function $\mathbb{R}^4 \ni (\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \mapsto Q_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$ is strictly convex for all $\mathbf{y}_n \in S_n$, see, e.g., Berkovitz [10, Theorem 3.3, Chapter III].

Since the function $\mathbb{R}^4 \ni (\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \mapsto Q_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$ is strictly convex for all $\mathbf{y}_n \in S_n$ and the system of equations (4.5.1) has a unique solution for all $\mathbf{y}_n \in S_n$, we get the function in question attains its (global) minimum at this unique solution, which yields (i) and (ii).

Further, for all $\mathbf{y}_n \in \mathbb{R}^{n+1}$ and $(\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) \in \mathbb{R}^4$,

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

and

$$\begin{aligned}
&\frac{\partial^2 Q_n}{\partial\theta'_1\partial\theta'_2}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = \frac{\partial^2 Q_n}{\partial\theta'_2\partial\theta'_1}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 0, \\
&\frac{\partial^2 Q_n}{\partial(\theta'_1)^2}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 2, \qquad \frac{\partial^2 Q_n}{\partial(\theta'_2)^2}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 2.
\end{aligned}$$

Then $H_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$ has the following leading principal minors

$$\begin{aligned}
&\Delta_{1,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 2\sum_{k=1}^{n} y_{k-1}^2, \\
&\Delta_{2,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 4\left(n\sum_{k=1}^{n} y_{k-1}^2 - \left(\sum_{k=1}^{n} y_{k-1}\right)^2\right), \\
&\Delta_{3,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) = 8\left((n-1)\sum_{k=1}^{n} y_{k-1}^2 - \left(\sum_{k=1}^{n} y_{k-1}\right)^2 + 2y_{s_1-1}\sum_{k=1}^{n} y_{k-1} - n(y_{s_1-1})^2\right), \\
&\Delta_{4,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2) := \det H_n(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2).
\end{aligned}$$

Note that $\Delta_{i,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$, $i = 1, 2, 3, 4$, do not depend on $(\alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$, and hence we will simply denote $\Delta_{i,n}(\mathbf{y}_n; \alpha', \mu'_\varepsilon, \theta'_1, \theta'_2)$ by $\Delta_{i,n}(\mathbf{y}_n)$.

Next we check that $\mathbf{Y}_n \in S_n$ holds asymptotically as $n \to \infty$ with probability one. By

(2.2.5) and (2.2.6), using the very same arguments as in the proof of Lemma 4.2.1, one can get

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{\Delta_{1,n}(\mathbf{Y}_n)}{n}=2\mathsf{E}\widetilde{X}^2\right)=1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{\Delta_{2,n}(\mathbf{Y}_n)}{n^2}=4\operatorname{Var}\widetilde{X}\right)=1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{\Delta_{3,n}(\mathbf{Y}_n)}{n^2}=8\operatorname{Var}\widetilde{X}\right)=1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{\Delta_{4,n}(\mathbf{Y}_n)}{n^2}=16\operatorname{Var}\widetilde{X}\right)=1,$$

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

$$\mathsf{P}\left(\lim_{n\to\infty}\Delta_{i,n}(\mathbf{Y}_n)=\infty\right)=1,\qquad i=1,2,3,4.$$

By (2.2.5) and (2.2.6), we also get

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{D_n(\mathbf{Y}_n)}{n^2}=\operatorname{Var}\widetilde{X}\right)=1,\tag{4.5.5}$$

and hence $\mathsf{P}(\lim_{n\to\infty}D_n(\mathbf{Y}_n)=\infty)=1$. □

By Lemma 4.5.1,

$$(\widehat{\alpha}_n(\mathbf{Y}_n),\widehat{\mu}_{\varepsilon,n}(\mathbf{Y}_n),\widehat{\theta}_{1,n}(\mathbf{Y}_n),\widehat{\theta}_{2,n}(\mathbf{Y}_n))$$

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

The next result shows that $\widehat{\alpha}_n$ and $\widehat{\mu}_{\varepsilon,n}$ are strongly consistent estimators of α and μ_ε, respectively, whereas $\widetilde{\theta}_{i,n}$, $i=1,2$, fail to be strongly consistent estimators of $\theta_{i,n}$, $i=1,2$, respectively.

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

$$\mathsf{P}(\lim_{n\to\infty}\widehat{\alpha}_n=\alpha)=1,\qquad\forall\,(\alpha,\mu_\varepsilon,\theta_1,\theta_2)\in(0,1)\times(0,\infty)\times\mathbb{N}^2,\tag{4.5.6}$$

$$\mathsf{P}(\lim_{n\to\infty}\widehat{\mu}_{\varepsilon,n}=\mu_\varepsilon)=1,\qquad\forall\,(\alpha,\mu_\varepsilon,\theta_1,\theta_2)\in(0,1)\times(0,\infty)\times\mathbb{N}^2,\tag{4.5.7}$$

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

$$\mathsf{P}\left(\lim_{n\to\infty}\widehat{\theta}_{i,n}=Y_{s_i}-\alpha Y_{s_i-1}-\mu_\varepsilon\right)=1\tag{4.5.8}$$

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

Proof. To prove (4.5.6) and (4.5.7), using Proposition 4.1.2 and the proof of Theorem 4.3.1, it is enough to check that

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}{}^{(s_1,s_2)}(Z_{k-1}^{(1)}+Z_{k-1}^{(2)})(Z_k^{(1)}+Z_k^{(2)})=0\right)=1,$$

$$\mathsf{P}\left(\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n}{}^{(s_1,s_2)}(Z_{k-1}^{(1)}+Z_{k-1}^{(2)})^2=0\right)=1.$$

The above relations follows by (4.4.6).

By (4.5.4) and (4.5.6), (4.5.7), we have (4.5.8). □

The asymptotic distribution of the CLS estimation is given in the next theorem.

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

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

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

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

where

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

Proof. Using Proposition 4.1.2, the proof of Theorem 4.3.2, and (4.4.9), (4.4.10), one can obtain (4.5.9). By (4.5.4) and (4.5.8),

$$\begin{aligned}\begin{bmatrix}\sqrt{n}\big(\widehat{\theta}_{1,n}-\lim_{k\to\infty}\widehat{\theta}_{1,k}\big)\\ \sqrt{n}\big(\widehat{\theta}_{2,n}-\lim_{k\to\infty}\widehat{\theta}_{2,k}\big)\end{bmatrix}&=\sqrt{n}\begin{bmatrix}Y_{s_1}-\widehat{\alpha}_nY_{s_1-1}-\widehat{\mu}_{\varepsilon,n}-(Y_{s_1}-\alpha Y_{s_1-1}-\mu_\varepsilon)\\ Y_{s_2}-\widehat{\alpha}_nY_{s_2-1}-\widehat{\mu}_{\varepsilon,n}-(Y_{s_2}-\alpha Y_{s_2-1}-\mu_\varepsilon)\end{bmatrix}\\ &=\begin{bmatrix}-Y_{s_1-1}&-1\\ -Y_{s_2-1}&-1\end{bmatrix}\begin{bmatrix}\sqrt{n}(\widehat{\alpha}_n-\alpha)\\ \sqrt{n}(\widehat{\mu}_{\varepsilon,n}-\mu_\varepsilon)\end{bmatrix}\end{aligned}$$

holds asymptotically as $n\to\infty$ with probability one. Using (4.5.9) we obtain (4.5.10). □

5 Appendix

5.1 Lemma. *If* $\alpha\in(0,1)$ *and* $\mathsf{E}\varepsilon_1<\infty$, *then the INAR(1) model in* (2.1.1) *has a unique stationary distribution.*

Proof. We follow the train of thoughts given in Section 6.3 in Hall and Heyde [35], but we also complete the proof given there. For all $n\in\mathbb{Z}_+$, let P_n denote the probability generating function of X_n, i.e., $P_n(s) := \mathsf{E} s^{X_n}$, $|s| \leqslant 1$, $s \in \mathbb{C}$. Let A and B be the probability generating function of the offspring $(\xi_{1,1})$ and the innovation (ε_1) distribution, respectively. With the notation

$$A^{(k)}(s) := (\underbrace{A \circ \cdots \circ A}_{k\text{-times}})(s), \qquad |s| \leqslant 1,\ s \in \mathbb{C},\ k \in \mathbb{N},$$

we get for all $|s| \leqslant 1, s \in \mathbb{C}$, and $n \in \mathbb{N}$,

$$\begin{aligned} P_n(s) &= \mathsf{E}(\mathsf{E}(s^{X_n} \mid \mathcal{F}_{n-1}^X)) = \mathsf{E}\Big[\mathsf{E}\big(s^{\sum_{j=1}^{X_{n-1}} \xi_{n,j}} \mid \mathcal{F}_{n-1}^X\big)\mathsf{E}(s^{\varepsilon_n} \mid \mathcal{F}_{n-1}^X)\Big] \\ &= \mathsf{E}(A(s)^{X_{n-1}} B(s)) = P_{n-1}(A(s))B(s). \end{aligned}$$

By iteration, we have

$$\begin{aligned} P_n(s) &= P_{n-1}(A(s))B(s) = P_{n-2}((A \circ A)(s))B(A(s))B(s) = \cdots \\ &= P_0(A^{(n)}(s))B(s) \prod_{k=1}^{n-1} B(A^{(k)}(s)), \qquad |s| \leqslant 1,\ s \in \mathbb{C},\ n \in \mathbb{N}. \end{aligned} \tag{5.11}$$

We check that $\lim_{n\to\infty} P_0(A^{(n)}(s)) = P_0(1) = 1$, $s \in \mathbb{C}$, and verify that the sequence $\prod_{k=1}^{n} B(A^{(k)}(s)), n \in \mathbb{N}$, is convergent for all $s \in [0, 1]$. By iteration, for all $n \in \mathbb{N}$,

$$\begin{aligned} A^{(n)}(s) &= A^{(n-1)}(1 - \alpha + \alpha s) = A^{(n-2)}(1 - \alpha + \alpha(1 - \alpha + \alpha s)) \\ &= A^{(n-2)}(1 - \alpha + \alpha(1 - \alpha) + \alpha^2 s) = \cdots = (1 - \alpha) \sum_{k=0}^{n-1} \alpha^k + \alpha^n s \\ &= (s - 1)\alpha^n + 1, \end{aligned} \tag{5.12}$$

and hence $\lim_{n\to\infty} A^{(n)}(s) = 1$, $s \in \mathbb{C}$. Then $\lim_{n\to\infty} P_0(A^{(n)}(s)) = P_0(1) = 1$, $s \in \mathbb{C}$. Since $0 \leqslant B(v) \leqslant 1$, $v \in [0, 1], v \in \mathbb{R}$, we get for all $s \in [0, 1]$, the sequence $\prod_{k=1}^{n} B(A^{(k)}(s))$, $n \in \mathbb{N}$, is nonnegative and monotone decreasing and hence convergent.

We will use the following known theorem (see, e.g., Chung [24, Section I.6, Theorem 4 and Section I.7, Theorem 2]. Let $(\xi_n)_{n\in\mathbb{Z}_+}$ be a homogeneous Markov chain with state space I. Let us suppose that there exists some subset D of I such that D is an essential, aperiodic class and $I \setminus D$ is a subset of inessential states. Then either

(a) for all $i \in I, j \in D$ we have $\lim_{n\to\infty} p_{i,j}^{(n)} = 0$, and therefore, there does not exist any stationary distribution,

or

(b) for all $i, j \in D$ we have $\lim_{n\to\infty} p_{i,j}^{(n)} := \pi_j > 0$, and in this case the unique stationary distribution is given by $(\widetilde{\pi}_j)_{j\in I}$ where $\widetilde{\pi}_j := \pi_j$ if $j \in D$ and $\widetilde{\pi}_j := 0$ if $j \in I \setminus D$.

Here $p_{i,j}^{(n)}$ denotes the n-step transition probability from the state i to the state j.

Let us introduce the notation

$$i_{\min} := \min\Big\{ i \in \mathbb{Z}_+ : \mathsf{P}(\varepsilon_1 = i) > 0 \Big\}.$$

Using that the offspring distribution is Bernoulli, i.e., it can take values 0 and 1, and both of them with positive probability, since $\alpha \in (0,1)$, one can think it over that the set of states $D := \Big\{ i \in \mathbb{Z}_+ : i \geqslant i_{\min} \Big\}$ is an essential class. Note also that $I \setminus D$ is a finite set of inessential states. The class D is aperiodic, since

$$p_{i_{\min}, i_{\min}} = \mathsf{P}(X_{n+1} = i_{\min} \,|\, X_n = i_{\min}) \geqslant \mathsf{P}(\varepsilon_{n+1} = i_{\min})(1-\alpha)^{i_{\min}} > 0.$$

Note that if the additional assumption $\mathsf{P}(\varepsilon_1 = 0) > 0$ is satisfied, then the Markov chain is irreducible and aperiodic.

Let us assume that there is no stationary distribution. With the notation

$$\widetilde{P}_n(s) := \frac{P_n(s)}{s^{i_{\min}}} = \sum_{k=0}^{\infty} s^k \mathsf{P}(X_n = k + i_{\min}), \qquad s \in [0,1],$$

we get for all $n \in \mathbb{N}$,

$$\widetilde{\mathsf{P}}_n(0) = \mathsf{P}(X_n = i_{\min}) = \sum_{j=0}^{\infty} \mathsf{P}(X_n = i_{\min} \,|\, X_0 = j)\mathsf{P}(X_0 = j) = \sum_{j=0}^{\infty} p_{j, i_{\min}}^{(n)} \mathsf{P}(X_0 = j).$$

Hence, by part (a) of the above recalled theorem, we get $\lim_{n\to\infty} p_{j,i_{\min}}^{(n)} = 0$ for all $j \in \mathbb{Z}_+$. Then the dominated convergence theorem yields that

$$\lim_{n\to\infty} \widetilde{P}_n(0) = 0.$$

However, we show that $\lim_{n\to\infty} \widetilde{P}_n(0) > 0$, which is a contradiction. Using that $\mathsf{P}(\varepsilon_1 = i_{\min}) > 0$ and that

$$\widetilde{\mathsf{P}}_n(0) = P_0(1-\alpha^n)\mathsf{P}(\varepsilon_1 = i_{\min}) \prod_{k=1}^{n-1} B(1-\alpha^k),$$

we have it is enough to prove that the limit of the sequence $\prod_{k=1}^{n} B(A^{(k)}(0)) = \prod_{k=1}^{n} B(1-\alpha^k)$, $n \in \mathbb{N}$, is positive. It is known that for this it is enough to verify that

$$\sum_{k=1}^{\infty} (1 - B(A^{(k)}(0))) = \sum_{k=1}^{\infty} (1 - B(1-\alpha^k)) \qquad \text{is convergent,}$$

see, e.g., Brémaud [17, Appendix, Theorem 1.9]. Just as in Section 6.3 in Hall and Heyde [35], we show that for all $s \in [0,1)$, $\sum_{k=1}^{\infty} (1 - B(A^{(k)}(s)))$ is convergent. For all $k \in \mathbb{N}$, $s \in [0,1)$,

$$1 - B(A^{(k)}(s)) = \frac{1 - B(A^{(k)}(s))}{1 - A^{(k)}(s)} (1 - A^{(k)}(s)),$$

and, by mean value theorem,

$$\frac{1 - B(A^{(k)}(s))}{1 - A^{(k)}(s)} = \frac{B(A^{(k)}(1)) - B(A^{(k)}(s))}{A^{(k)}(1) - A^{(k)}(s)} = \frac{B'(\theta(s))(A^{(k)}(1) - A^{(k)}(s))}{A^{(k)}(1) - A^{(k)}(s)} = B'(\theta(s)),$$

with some $\theta(s) \in (s, 1)$. Since

$$(5.13) \qquad B'(s) = \mathsf{E}(\varepsilon_1 s^{\varepsilon_1 - 1}) = \sum_{k=1}^{\infty} k s^{k-1} \mathsf{P}(\varepsilon_1 = k) \leqslant \sum_{k=1}^{\infty} k \mathsf{P}(\varepsilon_1 = k) = \mathsf{E}\varepsilon_1, \qquad s \in [0, 1],$$

we have

$$\frac{1 - B(A^{(k)}(s))}{1 - A^{(k)}(s)} \leqslant \mathsf{E}\varepsilon_1 = \mu_\varepsilon, \qquad s \in [0, 1).$$

Furthermore, by (5.12), we get

$$1 - A^{(k)}(s) \leqslant 1 - A^{(k)}(0) = \alpha^k, \qquad k \in \mathbb{N}, \ \ s \in [0, 1),$$

and hence $1 - B(A^{(k)}(s)) \leqslant \mu_\varepsilon \alpha^k$ for all $k \in \mathbb{N}$, $s \in [0, 1)$. Then

$$\sum_{k=1}^{\infty} (1 - B(A^{(k)}(s))) \leqslant \mu_\varepsilon \sum_{k=1}^{\infty} \alpha^k = \frac{\mu_\varepsilon \alpha}{1 - \alpha} < \infty, \qquad s \in [0, 1).$$

Let us denote by $\widetilde{X}$ a random variable on $(\Omega, \mathcal{A}, P)$ with a stationary distribution for the Markov chain $(X_n)_{n \in \mathbb{Z}_+}$. The, by the dominated convergence theorem and part (b) of the above recalled theorem, we have for all $j \in \mathbb{Z}_+$,

$$\begin{aligned}
\lim_{n \to \infty} \mathsf{P}(X_n = j) &= \lim_{n \to \infty} \sum_{i=0}^{\infty} \mathsf{P}(X_n = j \,|\, X_0 = i) \mathsf{P}(X_0 = i) \\
&= \sum_{i=0}^{\infty} \big(\lim_{n \to \infty} \mathsf{P}(X_n = j \,|\, X_0 = i) \big) \mathsf{P}(X_0 = i) \\
&= \mathsf{P}(\widetilde{X} = j) \sum_{i=0}^{\infty} \mathsf{P}(X_0 = i) = \mathsf{P}(\widetilde{X} = j),
\end{aligned}$$

which yields that X_n converges in distribution to $\widetilde{X}$ as $n \to \infty$. By the continuity theorem for probability generating functions (see, e.g., Feller [31, Section 11]), we also have $\widetilde{X}$ has the probability generating function

$$(5.14) \qquad P(s) := B(s) \prod_{k=1}^{\infty} B(A^{(k)}(s)), \qquad s \in (0, 1).$$

The uniqueness of the stationary distribution follows by part (b) of the above recalled theorem. □

Proofs of formulae (2.2.3)**,** (2.2.4) **and** (2.2.10)**.** Let us introduce the probability generating functions

$$A(s) := \mathsf{E}s^\xi = 1 - \alpha + \alpha s, \qquad s > 0,$$

where ξ is a random variable with Bernoulli distribution having parameter $\alpha \in (0, 1)$ and

$$B(s) := \mathsf{E}s^\varepsilon = \sum_{k=0}^{\infty} \mathsf{P}(\varepsilon = k) s^k, \qquad s > 0,$$

where ε is a non-negative integer-valued random variable with the same distribution as ε_1. In what follows we suppose that $\mathsf{E}\varepsilon^3 < \infty$. Since $\alpha \in (0,1)$ and $\mathsf{E}\varepsilon < \infty$, by Lemma 5.1, there exists a uniquely determined stationary distribution of the INAR(1) model in (2.1.1). Let us denote by $\widetilde{X}$ a random variable with this unique stationary distribution. Due to Hall and Heyde [35, formula (6.38)] or by the proof of Lemma 5.1, the probability generating function of $\widetilde{X}$ takes the form

$$P(s) := \mathsf{E}s^{\widetilde{X}} = B(s)B(A(s))B(A(A(s)))\cdots = B(s)\prod_{k=1}^{\infty} B(A^{(k)}(s)), \qquad s \in (0,1), \tag{5.15}$$

where for all $k \in \mathbb{N}$,

$$A^{(k)}(s) = (\underbrace{A \circ \cdots \circ A}_{k\text{-times}})(s), \qquad s \in (0,1).$$

Hence for all $s \in (0,1)$,

$$\begin{aligned} \log P(s) &= \log \mathsf{E}s^{\widetilde{X}} = \log B(s) + \log B(A(s)) + \log B(A(A(s))) + \cdots \\ &= \log B(s) + \sum_{k=1}^{\infty} \log B(A^{(k)}(s)). \end{aligned} \tag{5.16}$$

Using that $\mathsf{E}\varepsilon^3 < \infty$, by Abel's theorem (see, e.g., Brémaud [17, Appendix, Theorems 1.2 and 1.3]), we get

$$\lim_{s\uparrow 1}\left(\frac{\mathrm{d}}{\mathrm{d}s}\log B(s)\right) = \lim_{s\uparrow 1}\frac{\mathsf{E}(\varepsilon s^{\varepsilon-1})}{\mathsf{E}s^{\varepsilon}} = \mathsf{E}\varepsilon,$$

$$\lim_{s\uparrow 1}\left(\frac{\mathrm{d}^2}{\mathrm{d}s^2}\log B(s)\right) = \lim_{s\uparrow 1}\frac{\mathsf{E}(\varepsilon(\varepsilon-1)s^{\varepsilon-2})\mathsf{E}s^{\varepsilon} - (\mathsf{E}\varepsilon s^{\varepsilon-1})^2}{(\mathsf{E}s^{\varepsilon})^2} = \mathsf{E}(\varepsilon(\varepsilon-1)) - (\mathsf{E}\varepsilon)^2,$$

and

$$\lim_{s\uparrow 1}\left(\frac{\mathrm{d}^3}{\mathrm{d}s^3}\log B(s)\right) = \lim_{s\uparrow 1}\frac{N(s)}{(\mathsf{E}s^{\varepsilon})^4},$$

where

$$\begin{aligned} N(s) :=& \mathsf{E}(\varepsilon(\varepsilon-1)(\varepsilon-2)s^{\varepsilon-3})(\mathsf{E}s^{\varepsilon})^3 + \mathsf{E}(\varepsilon(\varepsilon-1)s^{\varepsilon-2})\mathsf{E}(\varepsilon s^{\varepsilon-1})(\mathsf{E}s^{\varepsilon})^2 \\ &- \Big[\mathsf{E}(\varepsilon(\varepsilon-1)s^{\varepsilon-2})\mathsf{E}s^{\varepsilon} - (\mathsf{E}\varepsilon s^{\varepsilon-1})^2\Big]2\mathsf{E}s^{\varepsilon}\mathsf{E}(\varepsilon s^{\varepsilon-1}) \\ &- 2\mathsf{E}(\varepsilon s^{\varepsilon-1})\mathsf{E}(\varepsilon(\varepsilon-1)s^{\varepsilon-2})(\mathsf{E}s^{\varepsilon})^2, \qquad s \in (0,1). \end{aligned}$$

Hence

$$\begin{aligned} \lim_{s\uparrow 1}\left(\frac{\mathrm{d}^3}{\mathrm{d}s^3}\log B(s)\right) &= \mathsf{E}(\varepsilon(\varepsilon-1)(\varepsilon-2)) - \mathsf{E}(\varepsilon(\varepsilon-1))\mathsf{E}\varepsilon - 2[\mathsf{E}(\varepsilon(\varepsilon-1)) - (\mathsf{E}\varepsilon)^2]\mathsf{E}\varepsilon \\ &= \mathsf{E}\varepsilon^3 - 3\mathsf{E}\varepsilon^2 + 2\mathsf{E}\varepsilon - 3(\mathsf{E}\varepsilon^2 - \mathsf{E}\varepsilon)\mathsf{E}\varepsilon + 2(\mathsf{E}\varepsilon)^3. \end{aligned}$$

Then

$$\mathsf{E}\varepsilon = \lim_{s\uparrow 1}\left(\frac{\mathrm{d}}{\mathrm{d}s}\log B(s)\right), \tag{5.17}$$

$$\mathsf{E}\varepsilon^2 = \lim_{s\uparrow 1}\left(\frac{\mathrm{d}^2}{\mathrm{d}s^2}\log B(s)\right) + \mathsf{E}\varepsilon + (\mathsf{E}\varepsilon)^2, \tag{5.18}$$

$$\mathsf{E}\varepsilon^3 = \lim_{s\uparrow 1}\left(\frac{\mathrm{d}^3}{\mathrm{d}s^3}\log B(s)\right) + 3\mathsf{E}\varepsilon^2 - 2\mathsf{E}\varepsilon + 3\mathsf{E}\varepsilon(\mathsf{E}\varepsilon^2 - \mathsf{E}\varepsilon) - 2(\mathsf{E}\varepsilon)^3. \tag{5.19}$$

By (5.16),

$$\log P(s) = \log \mathsf{E}s^{\widetilde{X}} = b(s) + \sum_{k=1}^{\infty} b(A^{(k)}(s)), \qquad s\in(0,1],$$

where $b(s) := \log B(s)$, $s\in(0,1]$. We show that

$$\lim_{s\uparrow 1}\left(\frac{\mathrm{d}}{\mathrm{d}s}\log P(s)\right) = b'(1) + \sum_{k=1}^{\infty}\left[b'(A^{(k)}(1))\prod_{\ell=0}^{k-1}A'(A^{(\ell)}(1))\right]. \tag{5.20}$$

First we note that

$$\frac{\mathrm{d}}{\mathrm{d}s}\log P(s) = b'(s) + \sum_{k=1}^{\infty}\left[b'(A^{(k)}(s))\prod_{\ell=0}^{k-1}A'(A^{(\ell)}(s))\right], \qquad s\in(0,1),$$

and, by (5.12), we get for all $k\in\mathbb{N}$, the functions $b'(A^{(k)}(s))$, $s\in[0,1]$ are well-defined. We check that the functions $b'(A^{(k)}(s))$, $s\in[0,1]$, $k\in\mathbb{N}$, are bounded with a common bound. By (5.12), we have

$$A^{(k)}(s) = (s-1)\alpha^k + 1 \in [1-\alpha^k, 1], \qquad s\in[0,1], \quad k\in\mathbb{N},$$

and hence $A^{(k)}(s)\in[1-\alpha,1]$, $s\in[0,1]$, $k\in\mathbb{N}$. Then, using (5.13), we get

$$b'(A^{(k)}(s)) = \frac{B'(A^{(k)}(s))}{B(A^{(k)}(s))} \leqslant \frac{\mathsf{E}\varepsilon}{B(1-\alpha)} < \infty, \qquad s\in[0,1], \quad k\in\mathbb{N}.$$

Using that $A'(s) = \alpha = \mathsf{E}\xi$, $\forall\ s>0$, and that

$$\sum_{k=1}^{\infty}\alpha^k = \frac{\alpha}{1-\alpha} < \infty,$$

the dominated convergence theorem and (5.17) yield (5.20). Hence, since $A(1)=1$,

$$\begin{aligned}\lim_{s\uparrow 1}\left(\frac{\mathrm{d}}{\mathrm{d}s}\log P(s)\right) &= \mathsf{E}\varepsilon + (\mathsf{E}\varepsilon)\mathsf{E}\xi + (\mathsf{E}\varepsilon)(\mathsf{E}\xi)^2 + \cdots = \sum_{k=0}^{\infty}(\mathsf{E}\varepsilon)(\mathsf{E}\xi)^k \\ &= \frac{\mathsf{E}\varepsilon}{1-\mathsf{E}\xi} = \frac{\mu_\varepsilon}{1-\alpha} < \infty.\end{aligned} \tag{5.21}$$

Just as we derived (5.17), but without supposing $\mathsf{E}\widetilde{X}<\infty$, Abel's theorem yields that

$$\mathsf{E}\widetilde{X} = \lim_{s\uparrow 1}\left(\frac{\mathrm{d}}{\mathrm{d}s}\log P(s)\right).$$

By (5.21), we get $\mathsf{E}\widetilde{X} = \frac{\mu_\varepsilon}{1-\alpha}$, which also shows that $\mathsf{E}\widetilde{X}$ is finite.

Using that

$$b''(s) = \frac{B''(s)B(s) - (B'(s))^2}{(B(s))^2}, \qquad s \in (0,1),$$

we get

$$b''(A^{(k)}(s)) \leqslant \frac{\mathsf{E}(\varepsilon(\varepsilon-1))\mathsf{E}\varepsilon}{(B(1-\alpha))^2} < \infty, \qquad s \in [0,1], \quad k \in \mathbb{N}.$$

Using also that $b''(1) = \mathsf{E}(\varepsilon(\varepsilon-1)) - (\mathsf{E}\varepsilon)^2$ and $A''(s) = 0$, $s > 0$, by the dominated convergence theorem, one can check that

$$\lim_{s\uparrow 1}\left(\frac{\mathrm{d}^2}{\mathrm{d}s^2}\log P(s)\right) = b''(1) + \sum_{k=1}^{\infty} b''(A^{(k)}(1))\left(\prod_{\ell=0}^{k-1} A'(A^{\ell}(1))\right)^2 = b''(1)\sum_{k=0}^{\infty}(\mathsf{E}\xi)^{2k}$$

$$= \frac{\mathsf{E}(\varepsilon(\varepsilon-1)) - (\mathsf{E}\varepsilon)^2}{1-(\mathsf{E}\xi)^2} = \frac{\operatorname{Var}\varepsilon - \mathsf{E}\varepsilon}{1-\alpha^2} = \frac{\sigma_\varepsilon^2 - \mu_\varepsilon}{1-\alpha^2},$$

which implies that $\mathsf{E}\widetilde{X}^2$ is finite and

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

By a similar argument, using that $\mathsf{E}\varepsilon^3 < \infty$ and

$$b'''(1) = \mathsf{E}(\varepsilon(\varepsilon-1)(\varepsilon-2)) - 3(\mathsf{E}\varepsilon)(\mathsf{E}\varepsilon(\varepsilon-1)) + 2(\mathsf{E}\varepsilon)^3,$$

we get

$$\lim_{s\uparrow 1}\left(\frac{\mathrm{d}^3}{\mathrm{d}s^3}\log P(s)\right) = b'''(1) + \sum_{k=1}^{\infty} b'''(A^{(k)}(1))\left(\prod_{\ell=0}^{k-1} A'(A^{\ell}(1))\right)^3$$

$$= \frac{\mathsf{E}(\varepsilon(\varepsilon-1)(\varepsilon-2)) - 3(\mathsf{E}\varepsilon)(\mathsf{E}\varepsilon(\varepsilon-1)) + 2(\mathsf{E}\varepsilon)^3}{1-(\mathsf{E}\xi)^3}$$

$$= \frac{\mathsf{E}\varepsilon^3 - 3\mathsf{E}\varepsilon^2 + 2\mathsf{E}\varepsilon - 3\mathsf{E}\varepsilon(\mathsf{E}\varepsilon^2 - \mathsf{E}\varepsilon) + 2(\mathsf{E}\varepsilon)^3}{1-\alpha^3}$$

$$= \frac{\mathsf{E}\varepsilon^3 - 3(\sigma_\varepsilon^2 + \mu_\varepsilon^2) + 2\mu_\varepsilon - 3\mu_\varepsilon(\sigma_\varepsilon^2 + \mu_\varepsilon^2 - \mu_\varepsilon) + 2\mu_\varepsilon^3}{1-\alpha^3} = \frac{\mathsf{E}\varepsilon^3 - 3\sigma_\varepsilon^2(1+\mu_\varepsilon) - \mu_\varepsilon^3 + 2\mu_\varepsilon}{1-\alpha^3},$$

which implies that $\mathsf{E}\widetilde{X}^3$ is finite and

$$\mathsf{E}\widetilde{X}^3 = \frac{\mathsf{E}\varepsilon^3 - 3\sigma_\varepsilon^2(1+\mu_\varepsilon) - \mu_\varepsilon^3 + 2\mu_\varepsilon}{1-\alpha^3} + 3\frac{\sigma_\varepsilon^2 + \alpha\mu_\varepsilon}{1-\alpha^2} - 2\frac{\mu_\varepsilon}{1-\alpha}$$

$$+ 3\frac{\mu_\varepsilon(\sigma_\varepsilon^2 + \alpha\mu_\varepsilon)}{(1-\alpha)(1-\alpha^2)} + \frac{\mu_\varepsilon^3}{(1-\alpha)^3}.$$

This yields (2.2.10). One can also write (2.2.10) in the following form

$$\mathsf{E}\widetilde{X}^3 = \frac{1}{1-\alpha^3}\Big[3\alpha^2(1-\alpha)\mathsf{E}\widetilde{X}^2 + 3\alpha^2\mu_\varepsilon\mathsf{E}\widetilde{X}^2 + 3\alpha\mathsf{E}\widetilde{X}(\sigma_\varepsilon^2 + \mu_\varepsilon^2) + \mathsf{E}\varepsilon^3 + 3\alpha(1-\alpha)\mu_\varepsilon\mathsf{E}\widetilde{X}$$

$$+\alpha(1-\alpha)(1-2\alpha)\mathsf{E}\widetilde{X}\Big].$$

□

5.2 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* $\mathsf{P}(X_{n-1} = k,\, Z_{n-1} = \ell) > 0$*,*

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

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

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

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

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

The following result can be found in several textbooks, see, e.g., Theorem 3.6 in Bhattacharya and Waymire [11, Chapter 0]. For completeness we give a proof.

5.3 Lemma. *Let* $(\xi_n)_{n\in\mathbb{N}}$ *be a sequence of random variables such that* $\mathsf{P}(\lim_{n\to\infty} \xi_n = 0) = 1$ *and* $\{\xi_n^p : n \in \mathbb{N}\}$ *is uniformly integrable for some* $p \in \mathbb{N}$*, i.e.,* $\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \mathsf{E}\left(|\xi_n|^p \mathbb{1}_{\{|\xi_n|>M\}}\right) = 0$*. Then* $\xi_n \stackrel{L_p}{\longrightarrow} 0$ *as* $n \to \infty$*, i.e.,* $\lim_{n\to\infty} \mathsf{E}|\xi_n|^p = 0$*.*

Proof. For all $n \in \mathbb{N}$ and $M > 0$, we get

$$\mathsf{E}|\xi_n|^p = \mathsf{E}\left(|\xi_n|^p \mathbb{1}_{\{|\xi_n|^p>M\}}\right) + \mathsf{E}\left(|\xi_n|^p \mathbb{1}_{\{|\xi_n|^p\leqslant M\}}\right) \leqslant \sup_{n\in\mathbb{N}} \mathsf{E}\left(|\xi_n|^p \mathbb{1}_{\{|\xi_n|^p>M\}}\right) + \mathsf{E}\left(|\xi_n|^p \mathbb{1}_{\{|\xi_n|^p\leqslant M\}}\right).$$

By $\mathsf{P}(\lim_{n\to\infty} \xi_n = 0) = 1$,

$$\lim_{n\to\infty} |\xi_n(\omega)|^p \mathbb{1}_{\{|\xi_n(\omega)|^p\leqslant M\}} = 0, \qquad \forall\ \omega \in \Omega,$$

and $\mathsf{E}\left(|\xi_n|^p \mathbb{1}_{\{|\xi_n|^p\leqslant M\}}\right) \leqslant M^p < \infty$ for all $n \in \mathbb{N}$. Hence, by dominated convergence theorem, we have

$$\lim_{n\to\infty} \mathsf{E}\left(|\xi_n|^p \mathbb{1}_{\{|\xi_n|^p\leqslant M\}}\right) = 0$$

for all $M > 0$. Then

$$\limsup_{n\to\infty} \mathsf{E}|\xi_n|^p \leqslant \sup_{n\in\mathbb{N}} \mathsf{E}\left(|\xi_n|^p \mathbb{1}_{\{|\xi_n|^p>M\}}\right), \qquad \forall\ M > 0.$$

By the uniformly integrability of $\{\xi_n^p : n \in \mathbb{N}\}$, we have $\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \mathsf{E}\left(|\xi_n|^p \mathbb{1}_{\{|\xi_n|^p>M\}}\right) = 0$, which yields the assertion. □

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] Ahn S., Gyemin L. and Jongwoo J. (2000) Analysis of the M/D/1-type queue based on an integer-valued autoregressive process. *Operations Research Letters* **27**, 235–241.

[3] 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.

[4] Al-Osh M. A. and Alzaid A. A. (1988) Integer-valued moving average (INMA) process. *Statistische Hefte* **29**(4), 281–300.

[5] Aly E. E. and Bouzar N. (1994) Explicit stationary distributions for some Galton-Watson processes with immigration. *Communications in Statistics. Stochastic Models* **10**(2), 499–517.

[6] Aly E. E. and Bouzar N. (2005) Stationary solutions for integer-valued autoregressive processes. *International Journal of Mathematics and Mathematical Sciences* **1**(1), 1–18.

[7] Andersson J. and Karlis D. (2010) Treating missing values in INAR(1) models: An application to syndromic surveillance data. *Journal of Time Series Analysis* **31**(1), 12–19.

[8] K. B. Athreya and P. E. Ney, *Branching processes*. Springer-Verlag, Berlin, 1972.

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

[10] Berkovitz L. D. (2002) *Convexity and optimization in* $\mathbb{R}^n$. sWiley.

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

[12] Blundell R., Griffith R. and Windmeijer F. (2002) Individual effects and dynamics in count data models. *Journal of Econometrics* **108**, 113–131.

[13] Brännäs K. (1995) Explanatory variables in the AR(1) count data model. *Umeå Economic Studies* **381**.

[14] Brännäs K. and Hall A. (2001) Estimation in integer-valued moving average models. *Applied Stochastic Models in Business and Industry* **17**, 277–291.

[15] Brännäs K. and Hellström J. (2001) Generalized integer-valued autoregression. *Econometric Reviews* **20**, 425–443.

[16] Brännäs K. and Nordström J. (2006) Tourist accomodation effects of festivals. *Tourism Economics* **12**(2), 291–302.

[17] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues.* Springer, 1999.

[18] Bu R. and McCabe B. (2008) Model selection, estimation and forecasting in INAR(p) models: A likelihood-based Markov chain approach. *International Journal of Forecasting* **24**(1), 151-162.

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

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

[21] Cameron A. C. and Trivedi P. (1998) *Regression analysis of count data.* Cambridge University Press, Oxford.

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

[23] 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.

[24] Chung K. L. (1960) *Markov chains with stationary transition probabilities*. Springer.

[25] Doukhan P., Latour A. and Oraichi D. (2006) A simple integer-valued bilinear time series model. *Advances in Applied Probability* **38**(2), 559–578.

[26] Drost F. C., van den Akker R. and Werker B. J. M. (2009) The asymptotic structure of nearly unstable non-negative integer-valued AR(1) models. *Bernoulli* **15**(2), 297–324.

[27] Drost F. C., van den Akker R. and Werker B. J. M. (2009) Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(2), 467–485.

[28] 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.

[29] Drost F. C., van den Akker R. and Werker B. J. M. (2008) Note on integer-valued bilinear time series models. *Statistics & Probability Letters* **78**, 992–996.

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

[31] W. Feller, *An introduction to probability theory and its applications, volume 1, 3rd edition.* Wiley, 1968.

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

[33] 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.

[34] Györfi L., Ispány M., Pap G. and Varga K. (2007) Poisson limit of an inhomogeneous nearly critical INAR(1) model. *Acta Universitatis Szegediensis. Acta Scientiarum Mathematicarum* **73**(3-4), 789–815.

[35] Hall P. and Heyde C. C. (1980) *Martingale limit theory and its application.* Academic Press.

[36] Ispány M. (2008) Limit theorems for normalized nearly critical branching processes with immigration. *Publicationes Mathematicae Debrecen* **72**(1-2), 17–34.

[37] 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.

[38] Ispány M., Pap G. and van Zuijlen M. (2003) Asymptotic behaviour of estimators of the parameters of nearly unstable INAR(1) models. Foundations of statistical inference (Shoresh, 2000), 195–206, *Contributions to Statistics*, Physica, Heidelberg.

[39] Ispány M., Pap G. and van Zuijlen M. (2005) Fluctuation limit of branching processes with immigration and estimation of the means. *Advances in Applied Probability* **37**(2), 523–538.

[40] Jung R. C. and Tremayne A. R. (2006) Binomial thinning models for integer time series. *Statistical Modelling* **6**, 81-96.

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

[42] Kachour M. and Yao J. F. (2009) First-order rounded integer-valued autoregressive (RINAR(1)) process. *Journal of Time Series Analysis* **30**(4), 417–448.

[43] Klimko L. A. and Nelson P. I. (1978) On conditional least squares estimation for stochastic processes. *The Annals of Statistics* **6**, 629–642.

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

[45] McCabe B. P. M. and Martin G. M. (2005) Bayesian prediction of low count time series. *International Journal of Forecasting* **21**, 315–330.

[46] 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**, 968–978.

[47] MacDonald I. and Zucchini W. (1997) *Hidden Markov and other models for discrete-valued time series.* Chapman & Hall, London.

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

[49] McKenzie E. (2003) Discrete variate time series. In: Shanbhag D. N., Rao C. R. (eds) *Handbook of Statistics*. Elsevier Science, 573–606.

[50] Monteiro M., Pereira I. and Scotto M. G. (2008) Optimal alarm systems for count processes. *Communications in Statistics. Theory and Methods* **37**, 3054–3076.

[51] Montgomery D. C. (2005) *Introduction to Statistical Quality Control, 5th edition*. John Wiley & Sons, Inc.

[52] 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.

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

[54] Quoreshi A. M. M. S. (2006) Bivariate time series modelling of financial count data. *Communications in Statistics. Theory and Methods* **35**, 1343–1358.

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

[56] 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.

[57] Silva M. E. and Oliveira V. L. (2005) Difference equations for the higher-order moments and cumulants of the INAR(p) model. *Journal of Time Series Analysis* **26**(1), 17–36.

[58] Silva I. and Silva M. E. (2006) Asymptotic distribution of the Yule-Walker estimator for INAR(p) processes. *Statistics & Probability Letters* **76**(15), 1655–1663.

[59] Silva I., Silva M. E., Pereira I. and Silva N. (2005) Replicated INAR(1) processes. *Methodology and Computing in Applied Probability* **7**, 517–542.

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

[61] Steutel F. W. and van Harn K. (2004) *Infinite divisibility of probability distributions on the real line*. Dekker, New York.

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

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

[64] Weiß C. H. (2008) Thinning operations for modelling time series of counts–a survey. *Advances in Statistical Analysis* **92**(3), 319-341.

[65] Weiß C. H. (2008) Serial dependence and regression of poisson INARMA models. *Journal of Statistical Planning and Inference* **138**, 2975-2990.

[66] Weiß C. H. (2008) The combined INAR(p) models for time series of counts. *Statistics & Probability Letters* **78**(13), 1817–1822.

[67] Zheng H. T., Basawa I. V. and Datta S. (2006) Inference for pth-order random coefficient integer-valued autoregressive processes. *Journal of Time Series Analysis* **27**, 411–440.

[68] Zheng H. T., Basawa I. V. and Datta S. (2007) First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference* **173**, 212–229.

[69] Zhou J. and Basawa I. V. (2005) Least-squared estimation for bifurcation autoregressive processes. *Statistics & Probability Letters* **74**, 77–88.

[70] Zhu R. and Joe H. (2003) A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. *Stochastic Models* **19**(2), 235–254.