A bivariate first-order signed integer-valued autoregressive process

Jan Bulla, Christophe Chesneau, Maher Kachour

▶ **To cite this version:**

HAL Id: hal-00655102

https://hal.archives-ouvertes.fr/hal-00655102

Submitted on 26 Dec 2011

A bivariate first-order signed integer-valued autoregressive process

Jan Bulla · Christophe Chesneau · Maher Kachour

Abstract Bivariate integer-valued time series occur in many areas, such as finance, epidemiology, business etc. In this paper, we present bivariate autoregressive integer-valued time series models, based on the signed thinning operator. Compared to classical bivariate INAR models, the new processes have the advantage to allow for negative values for the time series and the autocorrelation functions. Strict stationarity and ergodicity of the processes are established. The moments and the autocovariance functions are determined. Some methods for estimating the model parameters are considered and the asymptotic properties of the obtained estimators are derived. Simulation experiments as well as analysis of real data sets are carried out to assess the models' performance.

Keywords Bivariate integer-valued time series · INAR models · SINAR models · Bivariate Skellam distribution.

2000 Mathematics Subject Classification 62M10, 62M20.

1 Introduction

In many practical situations we have to deal with integer-valued times series. Because of the integer character of the observed series, a standard autoregressive process cannot be considered. Many attempts have therefore been made to define univariate models having the same properties as a real-valued AR

Jan Bulla and Christophe Chesneau
Laboratoire de Mathématiques Nicolas Oresme, CNRS UMR 6139, Université de Caen Basse-Normandie, Campus II, Science 3, 14032 Caen, France. E-mail: chesneau@math.unicaen.fr

Maher Kachour
Laboratoire de Mathématiques Jean Leray, Université de Nantes, 2 Rue de la Houssinière, BP 92208, 44322, Nantes, France. E-mail: maher.kachour@univ-nantes.fr

process and respecting the integer nature of the observations. Some important contributions are due to McKenzie (1985), Al-zaid and Al-Osh (1987) and Du and Li (1991). They used the binomial thinning operator of Steutel and van Harn (1979) to define stochastic processes, denoted by INAR, which resemble AR(1) or AR(p) processes. Gauthier and Latour (1991) also considered a more general version called the GINAR(p) model, based on the generalized Steutel and van Harn operator. For a recent review on discrete-valued time series models, we refer to, e.g., Kedem and Fokianos (2002), McKenzie (2003), Jung and Tremayne (2006) and Weiß (2008).

The literature on multivariate integer-valued time series models is limited. A first model (by Franke et al. 1993), which introduced the multivariate INAR(1) process, is based on the classical binomial thinning operator. Latour (1997) introduced a multivariate GINAR(p) model, based on the generalized thinning operator. Recently, there have been more attempts to derive bivariate integer-valued autoregressive type models. The interested reader is referred to Brannas and Nordstrom (2000), Quoreshi (2006), Heinen and Rengifo (2007) and Pedeli and Karlis (2011).

All models mentioned above (univariate and multivariate) are subject to the same intrinsic limitations. By construction, they cannot fit a time series with negative observations. Moreover, since regression coefficients must be positive, the modelling of series with possible negative correlations is not possible. In order to avoid the shortcomings of previous models, Kachour and Truquet (2011) introduced a general univariate class, denoted by SINAR, based on a modified version of the generalized thinning operator, called the signed thinning operator. Indeed, SINAR allows for negative values both for the series and its autocorrelation function, and can be seen as an extension of the GINAR model of Gauthier and Latour (1991) on $\mathbb{Z}$.

The purpose of this paper is to introduce a simple bivariate integer-valued autoregressive model, based on the signed thinning operator, denoted by B-SINAR(1). More precisely, we present two model specifications: First, with full autoregressive matrix and independent noise components. Second, with diagonal autoregressive matrix and correlated noise. Thus, the first specification is considered as an extension of the multivariate GINAR model on $\mathbb{Z}^2$ (in the bivariate case). Moreover, the second specification can be seen as a general version of the model presented by Pedeli and Karlis (2011).

The remainder of this paper is organized as follows. In Section 2 we present our main results. More precisely, in Section 2.1, two specifications of the first-order bivariate signed integer-valued autoregressive model are introduced. Moreover, we provide theoretical results about the stationarity and the moments of both B-SINAR(1) processes. In Section 3, the estimation of the model parameters is considered. Simulation experiments as well as analysis of real data are discussed in Sections 4 and 5. The proofs of all results are postponed to Section 6.

2 Main results

2.1 The B-SINAR(1) process

The so-called signed thinning operator, originally proposed by Latour and Truquet (2008), is a natural extension of the Steutel and van Harn (1979) operator to $\mathbb{Z}-$valued random variables and is defined as follows.

Definition 1 (*Signed thinning operator*) Let $(Y_i)_{i\in\mathbb{Z}}$ be a sequence of i.i.d. integer-valued random variables with F as common distribution, independent of an integer-valued random variable X. The signed thinning operator, denoted by $F\circ$, is defined by

$$
F \circ X = \begin{cases} \text{sign}(X) \displaystyle\sum_{i=1}^{|X|} Y_i, & \text{if } X \neq 0, \\[2em] 0 & \text{otherwise,} \end{cases} \tag{1}
$$

where, for an integer x, $\text{sign}(x) = 1$ if $x > 0$ and -1 if $x < 0$. The sequence $(Y_i)_{i\in\mathbb{Z}}$ is referred to as a counting sequence.

The model introduced in this paper is based on a random matrix operator defined below.

Definition 2 (*Signed matrix thinning operator*) Let $(Z_1, Z_2)^\tau$ be an integer-valued random vector. The signed matrix thinning operator, denoted by $F\odot$, is defined by

$$
F \odot \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} F_{1,1} \circ Z_1 + F_{1,2} \circ Z_2 \\ F_{2,1} \circ Z_1 + F_{2,2} \circ Z_2 \end{pmatrix},
$$

where

- $F_{i,j}$ is the common distribution of the i.i.d. counting sequence $(Y_k^{(i,j)})_{k\in\mathbb{Z}}$,
- all the counting sequences associated to the operators $F_{i,j}\circ$ are mutually independent

for any $(i,j) \in \{1,2\}^2$.

Definition 3 (*B-SINAR(1) process of first kind*) A bivariate process $(X_{t,1}, X_{t,2})_{t\in\mathbb{Z}}$ is said to be a B-SINAR(1) (for Bivariate Signed INteger-valued AutoRegressive) process of first kind if it admits the following representation

$$
\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = F \odot \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \end{pmatrix}, \tag{2}
$$

where $(\varepsilon_{t,j})_{t \in \mathbb{Z}}$ is an independent stationary process such that

- support$(\varepsilon_{1,j}) \subseteq \mathbb{Z}$,
- it is independent of all counting sequences of the model,
- $(\varepsilon_{t,1})_{t \in \mathbb{Z}}$ and $(\varepsilon_{t,2})_{t \in \mathbb{Z}}$ are independent

for any $j \in \{1, 2\}$.

Adopting the notations of Definitions 2 and 3, we set

$$\gamma_{i,j} = \mathbb{E}(Y_1^{(i,j)}), \qquad \beta_{i,j} = \mathbb{V}(Y_1^{(i,j)}), \qquad \mu_j = \mathbb{E}(\varepsilon_{1,j}), \qquad \sigma_{\varepsilon_j}^2 = \mathbb{V}(\varepsilon_{1,j})$$

and

$$M = \begin{pmatrix} \gamma_{1,1} \ \gamma_{1,2} \\ \\ \gamma_{2,1} \ \gamma_{2,2} \end{pmatrix} \tag{3}$$

for any $(i,j) \in \{1,2\}^2$.

The following theorem provides the conditions which ensure the ergodicity and the stationarity of the B-SINAR(1) process of first kind.

Theorem 1 *Consider a B-SINAR(1) model of first kind, as defined in* (2). *Assume that:*

1. $\varepsilon_1 = (\varepsilon_{1,1}, \varepsilon_{1,2})$ *charges all points of* $E = \mathbb{Z}^2$, *i.e.,* $\mathbb{P}(\varepsilon_1 = x) > 0$, $\forall\, x \in \mathbb{Z}^2$,
2. $\rho(M)$, *the spectral radius of the autoregressive matrix of the process* M (3), *is smaller than* 1.

Then, the process has a unique stationary solution, denoted by $X = (X_1, X_2)$, *and it is also ergodic. Moreover,* $\mathbb{E}(X_j^2) < \infty$ *for any* $j \in \{1, 2\}$.

In the following, we derive some properties of the B-SINAR(1) process of first kind under the assumptions of Theorem 1.

Let $X_1 \stackrel{d}{=} X_{t,1}$ and $X_2 \stackrel{d}{=} X_{t,2}$. By the stochastic representation in (2), we have

$$\mathbb{E}(X_1) = \frac{\mu_1(1 - \gamma_{2,2}) + \gamma_{1,2}\mu_2}{(1 - \gamma_{1,1})(1 - \gamma_{2,2}) - \gamma_{1,2}\gamma_{2,1}},$$

$$\mathbb{V}(X_1) = \frac{\beta_{1,1}\mathbb{E}(|X_1|) + \beta_{1,2}\mathbb{E}(|X_2|) + \gamma_{1,2}^2\mathbb{V}(X_2) + \sigma_{\varepsilon_1}^2}{1 - \gamma_{1,1}^2},$$

$$\mathbb{E}(X_2) = \frac{\mu_2(1 - \gamma_{1,1}) + \gamma_{2,1}\mu_1}{(1 - \gamma_{1,1})(1 - \gamma_{2,2}) - \gamma_{1,2}\gamma_{2,1}}$$

and

$$\mathbb{V}(X_2) = \frac{\beta_{2,1}\mathbb{E}(|X_1|) + \beta_{2,2}\mathbb{E}(|X_2|) + \gamma_{2,1}^2\mathbb{V}(X_1) + \sigma_{\varepsilon_2}^2}{1 - \gamma_{2,2}^2}.$$

Moreover, the autocorrelation functions are given by

$$\rho_{X_1}(k) = \gamma_{1,1}^k + \frac{\gamma_{1,2}}{\mathbb{V}(X_1)} \sum_{j=1}^{k} \gamma_{1,1}^{k-j} \mathbb{C}\text{ov}\left(X_{t,1}, X_{t+j-1,2}\right) \text{ and}$$

$$\rho_{X_2}(k) = \gamma_{2,2}^k + \frac{\gamma_{2,1}}{\mathbb{V}(X_2)} \sum_{j=1}^{k} \gamma_{2,2}^{k-j} \mathbb{C}\text{ov}\left(X_{t,2}, X_{t+j-1,1}\right),$$

for any $k \in \mathbb{N}^*$, where

$$\mathbb{C}\text{ov}\left(X_{t,1}, X_{t+h,2}\right) = \gamma_{2,1} \sum_{i=1}^{h} \gamma_{2,2}^{h-i} \mathbb{C}\text{ov}\left(X_{t,1}, X_{t+i-1,1}\right)$$

$$+ \frac{\gamma_{2,2}^h}{1 - \gamma_{1,1}\gamma_{2,2} - \gamma_{1,2}\gamma_{2,1}} \left(\gamma_{1,1}\gamma_{2,1}\mathbb{V}(X_1) + \gamma_{1,2}\gamma_{2,2}\mathbb{V}(X_2)\right),$$

$$\mathbb{C}\text{ov}\left(X_{t,2}, X_{t+h,1}\right) = \gamma_{1,2} \sum_{i=1}^{h} \gamma_{1,1}^{h-i} \mathbb{C}\text{ov}\left(X_{t,2}, X_{t+i-1,2}\right)$$

$$+ \frac{\gamma_{1,1}^h}{1 - \gamma_{1,1}\gamma_{2,2} - \gamma_{1,2}\gamma_{2,1}} \left(\gamma_{1,1}\gamma_{2,1}\mathbb{V}(X_1) + \gamma_{1,2}\gamma_{2,2}\mathbb{V}(X_2)\right)$$

for any $h \in \mathbb{N}$.

Remark 1 Compared to all bivariate integer-valued autoregressive models based on classical or generalized thinning operators, the B-SINAR(1) process of first kind has several advantages. More precisely, our process is constructed in a way such that it can be used to analyze time series with negative values. Moreover, the second assumption in Theorem 1 is similar to the corresponding condition ensuring stationarity of a real-valued bivariate AR(1) process. Therefore, the B-SINAR(1) process of first kind can reproduce auto- and cross-correlation functions with negative values.

2.2 Particular case of the B-SINAR(1) process of first kind

Next, we present a modified version of the B-SINAR(1) process of first kind.

Definition 4 (*B-SINAR(1) process of second kind*) A bivariate process $(X_{t,1}, X_{t,2})_{t \in \mathbb{Z}}$ is said to be a B-SINAR(1) process of second kind if each component admits the same representation of a SINAR(1) process, i.e.,

$$X_{t,1} = F_1 \circ X_{t-1,1} + \varepsilon_{t,1}, \qquad X_{t,2} = F_2 \circ X_{t-1,2} + \varepsilon_{t,2}, \tag{4}$$

where

- F_j is the common distribution of the i.i.d. counting sequence $(Y_k^{(j)})_{k \in \mathbb{Z}}$ for any $j \in \{1, 2\}$,

− $(\varepsilon_{t,j})_{t\in\mathbb{Z}}$ is an independent stationary process such that
    − support$(\varepsilon_{1,j}) \subseteq \mathbb{Z}$,
    − it is independent of all counting sequences of the model
  for any $j \in \{1,2\}$.

Adopting the notations in Definition 4, we set

$$\gamma_j = \mathbb{E}(Y_1^{(j)}), \qquad \beta_j = \mathbb{V}(Y_1^{(j)})$$

for any $j \in \{1,2\}$

Remark 2 In Definition 4, it is understood that $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$ are not independent, which implies $\mathbb{C}\mathrm{ov}(\varepsilon_{1,1}, \varepsilon_{1,2}) \neq 0$.

Remark 3 The model defined in (4) can be seen as a particular case of the above B-SINAR(1) process of first kind, where the signed matrix thinning operator equals

$$F\odot = \begin{pmatrix} F_1\circ & 0 \\ \\ 0 & F_2\circ \end{pmatrix}$$

and thus the autoregressive matrix has the following form

$$M = \begin{pmatrix} \gamma_1 & 0 \\ \\ 0 & \gamma_2 \end{pmatrix}. \tag{5}$$

Hence, using Theorem 1, the following corollary gives the conditions which ensure the ergodicity and the stationarity of the B-SINAR(1) process of second kind.

Corollary 1 *Consider a B-SINAR(1) model of second kind, as defined in* (4). *Assume that:*

1. $\varepsilon_1 = (\varepsilon_{1,1}, \varepsilon_{1,2})$ charges all points of $E = \mathbb{Z}^2$, i.e., $\mathbb{P}(\varepsilon_t = x) > 0$, $\forall\, x \in \mathbb{Z}^2$,
2. $\max(|\gamma_1|, |\gamma_2|) < 1$.

Then, the process has a unique stationary solution, denoted by $X = (X_1, X_2)$, and it is also ergodic. Moreover, $\mathbb{E}(X_j^2) < \infty$ for any $j \in \{1,2\}$.

Hence, under second order stationarity conditions we obtain

$$\mathbb{E}(X_j) = \frac{\mu_j}{1-\gamma_j}, \tag{6}$$

$$\mathbb{V}(X_j) = \frac{\beta_j \mathbb{E}(|X_j|) + \sigma_{\varepsilon_j}^2}{1-\gamma_j^2} \tag{7}$$

and

$$\rho_{X_j}(k) = \gamma_j^k, \qquad \forall\ k \in \mathbb{N} \tag{8}$$

for any $j \in \{1, 2\}$. Furthermore, we have

$$\mathbb{C}\text{ov}\,(X_{t,1}, X_{t+h,2}) = \frac{\gamma_2^h}{1 - \gamma_1\gamma_2}\,\mathbb{C}\text{ov}\,(\varepsilon_{t,1}, \varepsilon_{t,2}) \tag{9}$$

and

$$\mathbb{C}\text{ov}\,(X_{t,2}, X_{t+h,1}) = \frac{\gamma_1^h}{1 - \gamma_1\gamma_2}\,\mathbb{C}\text{ov}\,(\varepsilon_{t,1}, \varepsilon_{t,2}) \tag{10}$$

for any $h \in \mathbb{N}$.

Remark 4 Since the autoregressive matrix defined in (5) is diagonal, one can see that correlation between the innovations is the only source of dependence between the two series $(X_{t,1})_{t\in\mathbb{Z}}$ and $(X_{t,2})_{t\in\mathbb{Z}}$.

Remark 5 This model can be also viewed as a $\mathbb{Z}^2$-extension of the one studied in Pedeli and Karlis (2011). In addition, due to the second assumption of Corollary 1, the regression coefficients of our process can have negative values, which leads to possibly negative values of the auto- and cross-correlation functions (which cannot be achieved by the model of Pedeli and Karlis (2011) having coefficients belonging to $[0, 1]$).

3 Parameter estimation

3.1 For B-SINAR(1) process of first kind

Next, we state asymptotic results for the estimation of the parameters of B-SINAR(1) process of first kind, defined by (2). In more detail, we consider the Conditional Least Squares (CLS) estimator to determine $\gamma_{i,j}$ and μ_j, i.e., the means of distributions $F_{i,j}$ and ε_j, respectively, for any $(i, j) \in \{1, 2\}^2$.

Let

$$\theta = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \mu_1 \\ \\ \gamma_{2,1} & \gamma_{2,2} & \mu_2 \end{pmatrix}.$$

We suppose that θ belongs to a compact parameter space Θ.

Let $X_0, X_1, \cdots, X_n$ be a bivariate sample from B-SINAR(1) process of first kind, where $X_i = (X_{i,1}, X_{i,2})$ for any $i \in \{0, 1, \ldots, n\}$. Thus, the CLS estimator equals

$$\hat{\theta}^{\text{CLS}} = \begin{pmatrix} \hat{\gamma}_{1,1} & \hat{\gamma}_{1,2} & \hat{\mu}_1 \\ \\ \hat{\gamma}_{2,1} & \hat{\gamma}_{2,2} & \hat{\mu}_2 \end{pmatrix} = \underset{\theta \in \Theta}{\operatorname{argmin}}\, S_n(\theta),$$

with

$$S_n(\theta) = \sum_{t=1}^{n} \| X_t - \theta Y_{t-1} \|^2,$$

where $\| \cdot \|$ denotes the Euclidian norm on $\mathbb{R}^2$ and $Y_{t-1} = (X_{t-1,1}, X_{t-1,2}, 1)^{\tau}$. Note that we have the following explicit expressions

$$\hat{\theta}_1^{\mathrm{CLS}} = \begin{pmatrix} \hat{\gamma}_{1,1} \\ \hat{\gamma}_{1,2} \\ \hat{\mu}_1 \end{pmatrix} = Q^{-1}q_1, \quad \text{and} \quad \hat{\theta}_2^{\mathrm{CLS}} = \begin{pmatrix} \hat{\gamma}_{2,1} \\ \hat{\gamma}_{2,2} \\ \hat{\mu}_2 \end{pmatrix} = Q^{-1}q_2, \qquad (11)$$

where

$$Q = \begin{pmatrix} \sum_{t=1}^{n} X_{t-1,1}^2 & \sum_{t=1}^{n} X_{t-1,1}X_{t-1,2} & \sum_{t=1}^{n} X_{t-1,1} \\ \sum_{t=1}^{n} X_{t-1,1}X_{t-1,2} & \sum_{t=1}^{n} X_{t-1,2}^2 & \sum_{t=1}^{n} X_{t-1,2} \\ \sum_{t=1}^{n} X_{t-1,1} & \sum_{t=1}^{n} X_{t-1,2} & n \end{pmatrix},$$

$$q_1 = \begin{pmatrix} \sum_{t=1}^{n} X_{t,1}X_{t-1,1} \\ \sum_{t=1}^{n} X_{t,1}X_{t-1,2} \\ \sum_{t=1}^{n} X_{t,1} \end{pmatrix}, \quad \text{and} \quad q_2 = \begin{pmatrix} \sum_{t=1}^{n} X_{t,2}X_{t-1,1} \\ \sum_{t=1}^{n} X_{t,2}X_{t-1,2} \\ \sum_{t=1}^{n} X_{t,2} \end{pmatrix}.$$

Let

$$\theta = \begin{pmatrix} \theta_1^* \\ \theta_2^* \end{pmatrix} \quad \text{with} \quad \theta_1^* = \begin{pmatrix} \gamma_{1,1}^* \\ \gamma_{1,2}^* \\ \mu_1^* \end{pmatrix} \quad \text{and} \quad \theta_2^* = \begin{pmatrix} \gamma_{2,1}^* \\ \gamma_{2,2}^* \\ \mu_2^* \end{pmatrix}$$

be the actual value of the parameters. The following theorem establishes the asymptotic properties of the CLS estimators.

Theorem 2 *Consider a B-SINAR(1) process of first kind as defined in (2). Suppose that the assumptions of Theorem 1 are satisfied and that $\mathbb{E}(|X_j|^4) < \infty$ for any $j \in \{1,2\}$. Then, we obtain*

$$\sqrt{n}\left(\hat{\theta}_1^{\mathrm{CLS}} - \theta_1^*\right) \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(0, V^{-1}\Sigma V^{-1}\right)$$

and

$$\sqrt{n}\left(\hat{\theta}_2^{\mathrm{CLS}} - \theta_2^*\right) \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(0, V^{-1}\Gamma V^{-1}\right),$$

where

$$V = \lim_{n\to\infty} \frac{1}{n}Q = \mathbb{E}\left(Y_{t-1}Y_{t-1}^\tau\right),$$

$$\Sigma = \mathbb{E}\left(Y_{t-1}Y_{t-1}^\tau\left(X_{t,1} - \gamma_{1,1}X_{t-1,1} - \gamma_{1,2}X_{t-1,2} - \mu_1\right)^2\right)$$

and

$$\Gamma = \mathbb{E}\left(Y_{t-1}Y_{t-1}^\tau\left(X_{t,2} - \gamma_{2,1}X_{t-1,1} - \gamma_{2,2}X_{t-1,2} - \mu_2\right)^2\right)$$

for the CLS estimators defined by (11).

Remark 6 According to Theorem 1 of Kachour and Truquet (2011), one can deduce that $\mathbb{E}(|X_j|^4) < \infty$ if $F_{i,j}$ and $\varepsilon_{1,j}$ have a fourth moment for any $(i,j) \in \{1,2\}^2$.

3.2 For B-SINAR(1) process of second kind

In this section we discuss the estimation problem of a B-SINAR(1) process of second kind under some additional parametric assumptions. Indeed, we assume that F_j, the common distribution of the i.i.d. counting sequence $(Y_k^{(j)})_{k\in\mathbb{Z}}$, is given by

$$\mathbb{P}(Y_1^{(j)} = -1) = (1-\alpha_j)^2, \qquad \mathbb{P}(Y_1^{(j)} = 0) = 2\alpha_j(1-\alpha_j), \qquad \mathbb{P}(Y_1^{(j)} = 1) = \alpha_j^2, \tag{12}$$

with $\alpha_j \in (0,1)$ for any $j \in \{1,2\}$. In other words, we set $Y_1^{(j)} \overset{d}{=} Z_j - 1$ for any $j \in \{1,2\}$, where $Z_j \sim \mathrm{Binomial}(2, \alpha_j)$.

Remark 7 From (12), for any $j \in \{1,2\}$, we have

$$|\gamma_j| = |\mathbb{E}(Y_1^{(j)})| = |2\alpha_j - 1| < 1, \qquad \beta_j = \mathbb{V}(Y_1^{(j)}) = 2\alpha_j(1-\alpha_j).$$

Remark 8 Let $j \in \{1,2\}$, $x \in \mathbb{Z}^*$ and $y \in \mathbb{Z}$. Than holds

$$\mathbb{P}\left(F_j \circ x = y\right) = \mathbb{P}\left(T_x^{(j)} = \mathrm{sign}(x)(x+y)\right),$$

where $T_x^{(j)} \sim \mathrm{Binomial}(2|x|, \alpha_j)$, i.e.,

$$\mathbb{P}\left(F_j \circ x = y\right) = \binom{2|x|}{|x| + y\,\mathrm{sign}(x)}\alpha_j^{|x|+y\,\mathrm{sign}(x)}(1-\alpha_j)^{|x|-y\,\mathrm{sign}(x)}\mathbf{1}_{\{y\in\{-|x|,\ldots,|x|\}\}},$$

with $\mathbf{1}$ denoting the indicator function.

Remark 9 In fact, the distribution F_j, defined by (12), can be seen as a natural extension of the binomial distribution, which is used in the definition of the classical thinning operator.

Moreover, we assume that the innovations of the two series in (4) jointly follow the bivariate Skellam distribution presented below.

Definition 5 (*BSkellam*($\lambda_0, \lambda_1, \lambda_2$) *distribution*) Let $\lambda_0 \geq 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$. We say that the bivariate random variable $(\varepsilon_1, \varepsilon_2)$ has the bivariate Skellam distribution, denoted by BSkellam($\lambda_0, \lambda_1, \lambda_2$), if and only if

$$\varepsilon_1 \overset{d}{=} U_1 - U_0, \qquad \varepsilon_2 \overset{d}{=} U_2 - U_0,$$

where U_0, U_1 and U_2 are three independent random variables such that $U_i \sim$ Poisson(λ_i) for any $i \in \{0, 1, 2\}$. By convention, we set $U_0 = 0$ if $\lambda_0 = 0$.

Remark 10 When $\lambda_0 = 0$, the distribution of $(\varepsilon_1, \varepsilon_2)$ is that of a couple of independent Poisson random variables. Now, suppose that $\lambda_0 > 0$. Then,

– the probability mass function of $(\varepsilon_1, \varepsilon_2)$ is given by

$$\mathbb{P}(\varepsilon_1 = x_1, \varepsilon_2 = x_2) = e^{-(\lambda_1 + \lambda_2 + \lambda_0)} \lambda_1^{x_1} \lambda_2^{x_2} \sum_{i=\max(0, -x_1, -x_2)}^{\infty} \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x_1 + i)!(x_2 + i)!i!},$$

$(x_1, x_2) \in \mathbb{Z}^2$.

– the mean of $(\varepsilon_1, \varepsilon_2)$ is $(\lambda_1 - \lambda_0, \lambda_2 - \lambda_0)$ and the covariance matrix equals

$$\Sigma = \begin{pmatrix} \lambda_1 + \lambda_0 & \lambda_0 \\ \\ \lambda_0 & \lambda_2 + \lambda_0 \end{pmatrix}.$$

A detailed study of the bivariate Skellam distribution is postponed to the annex.

Remark 11 Under the above parametric assumption, it follows from Corollary 1 that the B-SINAR(1) process of second kind defined by (4) has a unique stationary solution, and it is also ergodic.

Definition 6 (*SBP-SINAR(1) process*) A B-SINAR(1) process of second kind with the above parametric assumptions will be denoted by SBP-SINAR(1) (for the first-order Skellam Bivariate Parametric SINAR process) in the following.

Next, two estimation methods are proposed to estimate the parameters of the SBP-SINAR(1) process, i.e., $\theta = (\alpha_1, \alpha_2, \lambda_0, \lambda_1, \lambda_2)$.

3.2.1 The method of moments

Based on (6), (8), and (9) (or (10)), moment estimation is straightforward. Let

$$\overline{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{i,j}, \quad S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{i,j} - \overline{X}_j)^2,$$

$$\overline{|X|}_j = \frac{1}{n}\sum_{i=1}^{n}|X_{i,j}|, \text{ and } C_{1,2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{i,1} - \overline{X}_1)(X_{i,2} - \overline{X}_2)$$

for any $j \in \{1,2\}$, and denote the first order sample autocorrelation function by $\hat{\rho}_{X_j}(1)$. Then, one obtains

$$\hat{\alpha}_j = \frac{\hat{\rho}_{X_j}(1) + 1}{2}, \tag{13}$$

$$\hat{\lambda}_j = \hat{\lambda}_0 + 2\overline{X}_j(1 - \hat{\alpha}_j) \tag{14}$$

and

$$\hat{\lambda}_0 = C_{1,2}\left(1 - \hat{\rho}_{X_1}(1)\hat{\rho}_{X_2}(1)\right) \tag{15}$$

for any $j \in \{1,2\}$.

Remark 12 Based on (7), one may also introduce other natural estimators for λ_j, i.e.,

$$\tilde{\lambda}_j = (1 - \hat{\alpha}_j)\left(\overline{X}_j + \hat{\alpha}_j(2S_j^2 - \overline{|X|}_j)\right) \tag{16}$$

for $j \in \{1,2\}$. At first glance, $\tilde{\lambda}_j$ seems slightly more complex, and we study the differences between the two estimators by means of a small simulation study in Section 5.

3.2.2 The conditional maximum likelihood estimator

Here, we consider the Conditional Maximum Likelihood (CML) estimator of $\theta = (\alpha_1, \alpha_2, \lambda_0, \lambda_1, \lambda_2)$. It is assumed that θ belongs to a compact parameter space Θ. Let $x = (x_0, x_1, \cdots, x_n)$ be a bivariate sample from SBP-SINAR(1) process with $x_i = (x_{i,1}, x_{i,2})$ for any $i \in \{0, 1, \ldots, n\}$. Thus, the CML estimator is

$$\hat{\theta}^{\mathrm{CML}} = \operatorname*{argmax}_{\theta \in \Theta} L(\mathbf{x}, \theta),$$

with

$$L(\mathbf{x}, \theta) = \prod_{t=1}^{n} \mathbb{P}\left(X_t = x_t \mid X_{t-1} = x_{t-1}\right).$$

Note that the conditional distribution $X_t \mid X_{t-1}$ in the above equation is equivalent to $\Gamma(\cdot, \cdot)$, the one-step transition probability of the Markov chain $(X_t)_{t \in \mathbb{Z}}$ defined by (4). Hence,

$$
\begin{aligned}
\Gamma((a,b),(c,d)) &= \mathbb{P}\left(X_t = (a,b) \mid X_{t-1} = (c,d)\right) \\
&= \mathbb{P}\left(F_1 \circ c + \varepsilon_{t,1} = a, F_2 \circ d + \varepsilon_{t,2} = b\right).
\end{aligned}
$$

Let $T_x^{(j)}$ be a Binomial$(2|x|, \alpha_j)$ random variable for all $x \in \mathbb{Z}^*$ and any $j \in \{1, 2\}$. Then, using Remark 8, we distinguish four cases:

– For $c = d = 0$ holds

$$
\Gamma((a,b),(0,0)) = \mathbb{P}\left(\varepsilon_{t,1} = a, \varepsilon_{t,2} = b\right);
$$

– For $c \neq 0$ and $d = 0$ holds

$$
\Gamma((a,b),(c,0)) = \sum_{k=a-|c|}^{a+|c|} \mathbb{P}\left(T_c^{(1)} = \mathrm{sign}(c)(c+a-k)\right) \mathbb{P}\left(\varepsilon_{t,1} = k, \varepsilon_{t,2} = b\right);
$$

– For $c = 0$ and $d \neq 0$ holds

$$
\Gamma((a,b),(0,d)) = \sum_{k=b-|d|}^{b+|d|} \mathbb{P}\left(T_d^{(2)} = \mathrm{sign}(d)(d+b-k)\right) \mathbb{P}\left(\varepsilon_{t,1} = a, \varepsilon_{t,2} = k\right);
$$

– For $c \neq 0$ and $d \neq 0$ holds

$$
\begin{aligned}
\Gamma((a,b),(c,d)) = \sum_{k_1=a-|c|}^{a+|c|} \sum_{k_2=b-|d|}^{b+|d|} &\mathbb{P}\left(\varepsilon_{t,1} = k_1, \varepsilon_{t,2} = k_2\right) \times \\
&\mathbb{P}\left(T_c^{(1)} = \mathrm{sign}(c)(c+a-k_1)\right) \mathbb{P}\left(T_d^{(2)} = \mathrm{sign}(d)(d+b-k_2)\right).
\end{aligned}
$$

Note that the joint probability mass function of the two innovation process is given in Remark 10.

Remark 13 Indeed, CML estimation is more complicated than the method of moments, but still tractable. Numerical maximization is straightforward with standard statistical packages, but may be computationally costly for long sequences of observations.

Remark 14 Since $(X_t)_{t \in \mathbb{Z}}$ is a stationary Markov chain, one can obtain asymptotic normality of the parameter estimators from Theorem 2.2 in Billingsley (1961) under standard technical assumptions. For more details we refer to Franke et al. (1993).

4 Simulation study

In this part we shortly investigate some properties of the parameter estimation techniques of the SBP-SINAR(1) process, as presented in Section 3.2. More precisely, we consider a SBP-SINAR(1) model with parameters $(\lambda_0, \lambda_1, \lambda_2) = (7, 3, 8)$ and $(\alpha_1, \alpha_2) = (0.3, 0.8)$. Figure 1 shows random samples with 200 observations of the two components X_1 and X_2, the dashed lines correspond to the expected values of the process. The two components are correlated, in this example the estimated correlation coefficient equals 0.206 (p = 0.00332).

The main focus of the Monte Carlo study lies on comparing the properties of the methods of moments and the the maximum likelihood estimator. Therefore, we generated 200 sequences of observations of length $n = 50$ (respectively 100, 200), using the parameters above. Subsequently, the parameters were estimated by the method of moments and by maximization of the conditional likelihood. With respect to the methods of moments estimator, it should be noted that two methods exist for estimating the parameters λ_1 and λ_2 (see Equations (14) and (16)). Figure 2 shows Box plots of the estimated parameters. The maximum likelihood estimator is colored dark gray, the methods of moments estimators from Equation (14) have white background. Box plots with light gray background result from the estimator of Equation (16), and the horizontal dashed lines correspond to the true parameter values. Several aspects are apparent: First, all estimators seem to be consistent and unbiased, only α_i, $i \in \{1, 2\}$ might be subject to a slight underestimation for short sequences. Second, the maximum likelihood estimator has the lowest variability. Third, comparing the methods of moments estimators, it seems advisable to prefer Equation (16) due to the lower variability of the estimates. Finally, it may be noted that normality of the estimators cannot be rejected for the majority of samples at 5%-level (26 of 36, Shapiro-Wilk).

5 Analysis of the annual Swedish population rates and the Harvest index data

In the following, we fit different models to Swedish population rates and Harvest index data. The two top panels of Figure 3 shows the time series of the annual Swedish population rates (per thousand population) and the Swedish Harvest index, denoted P_t and h_t, respectively, for the period $1750 - 1849$, as reported by Thomas (1940). Although such population variation should be real-valued, the recorded rates are rounded values on the one thousand scale. The rates vary from -27 to 16, with a sample mean of 6.69 and sample variance of 34.6. Originally, the Swedish grain harvest was measured on a half-point scale with a total crop failure scored as zero and a superabundant crop scored as nine. In order to avoid observations such as 3.5 we refine the scale of this series by multiplying all the observations by 2. Thus, the Harvest index takes values between 0 and 18, with a sample mean of 10.5 and sample

variance of 27.8. The sample ACF and PACF of $(P_t)_{t\in\mathbb{Z}}$ and $(h_t)_{t\in\mathbb{Z}}$ are given in the middle panels of Figure 3.

Remark 15 Since the annual Swedish rates series has negative observations, the bivariate time series $X_t = (P_t, h_t)^\tau$ cannot be fitted by any existing integer-valued model based on the classical or generalized thinning operator.

Remark 16 For both series, sample ACF and PACF have values significantly different to zero only at lag 1, if at all. Moreover, the sample correlation equals only 0.0589 (p = 0.56), but the cross-correlation of P_t and h_{t-1} has the value 0.400 (p = 4.11e-05, see also the lower panel of Figure 3). Therefore, one may expect the B-SINAR(1) model to be more adequate for this data series.

Fit of a B-SINAR(1) model

To analyze the present bivariate time series, we consider the following B-SINAR(1) of first kind model

$$
\begin{pmatrix} P_t \\ h_t \end{pmatrix} = \begin{pmatrix} F_{1,1}\circ & F_{1,2}\circ \\ F_{2,1}\circ & F_{2,2}\circ \end{pmatrix} \begin{pmatrix} P_{t-1} \\ h_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}.
$$

Let $\theta = (\gamma_{1,1}, \gamma_{1,2}, \gamma_{2,1}, \gamma_{2,2}, \mu_1, \mu_2)^\tau$ be the model parameters, where $\gamma_{i,j}$ and μ_j represent the respective means of $F_{i,j}$ and ε_t^j for any $(i,j) \in \{1,2\}^2$ and $t \in \mathbb{Z}$. We estimated the parameters using (11). This resulted in

$$
\hat{\gamma}_{1,1} = \underset{(0.0890)}{0.442}, \quad \hat{\gamma}_{1,2} = \underset{(0.0964)}{0.414}, \quad \hat{\mu}_1 = \underset{(1.23)}{-0.606},
$$

$$
\hat{\gamma}_{2,1} = \underset{(00.0829)}{-0.119}, \quad \hat{\gamma}_{2,2} = \underset{(0.0890)}{0.398}, \quad \text{and} \quad \hat{\mu}_2 = \underset{(1.07)}{6.97}.
$$

Standard errors, reported on brackets below each estimate, result from a parametric bootstrap with 200 repetitions.

Remark 17 The spectral radius of the obtained autoregressive matrix, i.e.,

$$
\hat{M} = \begin{pmatrix} 0.442 & 0.414 \\ -0.119 & 0.398 \end{pmatrix},
$$

is smaller than 1, thus the stationarity condition of Theorem 1 is satisfied.

Hence, the one-step ahead least squares predictions can be calculated by

$$
\hat{P}_{t+1} = 0.442P_t + 0.414h_t - 0.606
$$

and

$$
\hat{h}_{t+1} = -0.119P_t + 0.398h_t + 6.97.
$$

Remark 18 In general, $\hat{P}_{t+1}$ and $\hat{h}_{t+1}$ are real-valued, a mapping into the discrete support of the series is obtained by rounding to the nearest integer.

Figure 4 shows the observed values of the series of Swedish population rates and Harvest index, with the corresponding one-step ahead forecasts based on the B-SINAR(1) model of first kind. The horizontal lines correspond to the sample mean values of both series. The mean absolute error (MAE) for both series equals

$$\mathrm{MAE(P_t)} = \frac{1}{99} \sum_{t=1}^{99} |\hat{P}_{t+1} - P_{t+1}| = 2.95$$

and

$$\mathrm{MAE(h_t)} = \frac{1}{99} \sum_{t=1}^{99} |\hat{h}_{t+1} - h_{t+1}| = 3.95.$$

Fit of a SBP-SINAR(1) model

In order to fit the bivariate time series of the annual Swedish population rates and the Swedish Harvest index, we consider the following SBP-SINAR(1) process

$$\begin{pmatrix} P_t \\ h_t \end{pmatrix} = \begin{pmatrix} F_1\circ & 0 \\ 0 & F_2\circ \end{pmatrix} \begin{pmatrix} P_{t-1} \\ h_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix},$$

with F_j as defined in (12), and $\varepsilon_t = \left(\varepsilon_t^1, \varepsilon_t^1\right)$ follows the bivariate Skellam distribution (see Definition 5), for any $j \in \{1,2\}$ and $t \in \mathbb{Z}$. Let $\theta = (\alpha_1, \alpha_2, \lambda_0, \lambda_1, \lambda_2)^\tau$ be the model parameters, where α_1 and α_2) are the parameters of the distribution F_1 (respectively F_2), and λ_0, λ_1 and λ_2 are the parameters of the bivariate Skellam distribution.

We estimated the model parameters by maximization of the likelihood due to the superior performance of the estimator in the previous section. The resulting parameter estimates are

$$\hat{\alpha}_1 = \underset{(0.0428)}{0.790}, \quad \hat{\alpha}_2 = \underset{(0.0431)}{0.597}, \quad \hat{\lambda}_0 = \underset{(1.48)}{10.0},$$

$$\hat{\lambda}_1 = \underset{(1.52)}{12.7}, \quad \text{and} \quad \hat{\lambda}_2 = \underset{(1.55)}{18.4}.$$

As before, standard errors, reported on brackets below each estimate, result from a parametric bootstrap with 200 repetitions. Thus, the one-step ahead least squares predictions lead to

$$\tilde{P}_{t+1} = (2\hat{\alpha}_1 - 1)P_t + \left(\hat{\lambda}_1 - \hat{\lambda}_0\right)$$
$$= 0.581 P_t + 2.73,$$

and

$$\tilde{h}_{t+1} = (2\hat{\alpha}_2 - 1)h_t + \left(\hat{\lambda}_2 - \hat{\lambda}_0\right)$$
$$= 0.195P_t + 8.39,$$

Figure 5 shows the observed values of the series of Swedish population rates and Harvest index and the corresponding one-step ahead forecasts resulting from the estimated SBP-SINAR(1) model. The mean absolute error (MAE) for both series equals

$$\text{MAE}(P_t) = \frac{1}{99}\sum_{t=1}^{99}|\tilde{P}_{t+1} - P_{t+1}| = 3.38,$$

and

$$\text{MAE}(h_t) = \frac{1}{99}\sum_{t=1}^{99}|\tilde{h}_{t+1} - h_{t+1}| = 3.98.$$

Summarizing, both proposed models have their small advantages and disadvantages: On the one hand, the B-SINAR(1) of first kind model obtains a better in-sample fit in terms of MAE. On the other hand, the SBP-SINAR(1) constitutes the more parsimonious model, however, the difference in MAE is rather small. This result can be explained by the low correlation but significant cross-correlation of the series at lag 1. A relevant advantage of the SBP-SINAR(1) model may be expected for data with significantly correlation at lag 0, in particular in a simulation context.

6 Proofs

6.1 Proof of Theorem 1

Let $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^2$, and denote the corresponding operator norm associated to a matrix of size 2×2 in the same way. For all $x \in E = \mathbb{Z}^2$, we define the empirical measure μ_n by

$$\mu_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n}\pi^i(x,\cdot),$$

where π^i denotes the i-step transition probability of the Markov chain $(X_t)_{t\in\mathbb{Z}}$ defined by (2) for any $i \in \{1, 2, \ldots, n\}$. The idea is to prove that this sequence of probability distribution $(\mu_n)_{n\in\mathbb{Z}}$ has a sub-sequence converging to some probability distribution μ. By construction, this limit will automatically be a stationary probability distributions of the B-SINAR(1) process.

Let $V : \mathbb{R}^2 \to \mathbb{R}^+$ be the Lyapunov function defined by $V(x) = \|x\|^2$. Since $\rho(M) < 1$, we have that $\|M^k\| < 1$ if k is large enough. For such a k, using (Kachour and Truquet 2011, Lemma 3), we obtain

$$\limsup_{\|x\|\to+\infty} \frac{\mathbb{E}\left(V(X_k) \mid X_0 = x\right)}{V(x)} < 1,$$

and the Lyapunov criterion is satisfied. Therefore, the sequence $(\mu_n)_{n \in \mathbb{N}}$ is tight and any cluster point μ is an invariant distribution with $\mu(V) < \infty$ (see (Duflo 1997, Proposition 2.1.6)). It follows that μ is a measure $\pi-$invariant. From assumption 1, one can deduce that B-SINAR(1) is an irreducible and aperiodic Markov chain on $\mathbb{Z}^2$. Therefore, the Markov chain $(X_t)_{t \in \mathbb{Z}}$, as defined in (2), is positive recurrent on E and the invariant measure μ is unique. The existence and uniqueness of a stationary process $X = (X_1, X_2)$ satisfying (2) follows and $\mu(V) = \mathbb{E}(\|X\|^2) < \infty$. Finally, the ergodicity of the process X easily follows since $(X_t)_{t \in \mathbb{Z}}$ is positive recurrent on E. ∎

6.2 Proof of Theorem 2

The asymptotic properties of the CLS estimators for the parameters can be established by using the results in Klimko and Nelson (1978). In the following, we present a direct proof inspired by the proof of Theorem 3.2 in Zhang et al. (2010). For any $(i,j) \in \{1,2\}^2$, set

$$M_n^{(i,j)} = -\frac{1}{2}\frac{\partial S_n(\theta)}{\partial \gamma_{i,j}} \qquad \text{and} \qquad \mathrm{M}_n^{(j)} = -\frac{1}{2}\frac{\partial \mathrm{S}_n(\theta)}{\partial \mu_j}.$$

Observe that

$$\sqrt{n}\left(\hat{\theta}_1^{\mathrm{CLS}} - \theta_1^*\right) = \left(\frac{1}{n}Q\right)^{-1}\frac{1}{\sqrt{n}}\begin{pmatrix} M_n^{(1,1)} \\ \\ M_n^{(1,2)} \\ \\ M_n^{(1)} \end{pmatrix}, \qquad (17)$$

and

$$\sqrt{n}\left(\hat{\theta}_2^{\mathrm{CLS}} - \theta_2^*\right) = \left(\frac{1}{n}Q\right)^{-1}\frac{1}{\sqrt{n}}\begin{pmatrix} M_n^{(2,1)} \\ \\ M_n^{(2,2)} \\ \\ M_n^{(2)} \end{pmatrix}. \qquad (18)$$

Let $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$ and $R_t^1 = X_{t,1} - \gamma_{1,1}X_{t-1,1} - \gamma_{1,2}X_{t-1,2} - \mu_1$. Then we have

$$\begin{aligned} \mathbb{E}\left(M_n^{(1,1)} \mid \mathcal{F}_{n-1}\right) &= \mathbb{E}\left(M_{n-1}^{(1,1)} + R_n^1 X_{n-1,1} \mid \mathcal{F}_{n-1}\right) \\ &= M_{n-1}^{(1,1)} + \mathbb{E}\left(R_n^1 X_{n-1,1} \mid \mathcal{F}_{n-1}\right) \\ &= M_{n-1}^{(1,1)}. \end{aligned}$$

Hence, $\{M_n^{(1,1)}, \mathcal{F}_n, n \geq 1\}$ is a martingale. Under the moments assumptions in Theorem 2, we obtain

$$\frac{1}{n} \sum_{t=1}^{n} \left(R_t^1 X_{t-1,1} \right)^2 \xrightarrow{a.s.} \mathbb{E} \left(R_t^1 \right)^2 X_{t-1,1}^2 = \Sigma_{1,1}$$

from the ergodic theorem. Moreover, using (Hall and Heyde 1980, Corollary 3.2), we get

$$\frac{1}{\sqrt{n}} M_n^{(1,1)} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma_{1,1}\right).$$

Similarly, one can prove that

$$\frac{1}{\sqrt{n}} M_n^{(1,2)} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma_{2,2}\right)$$

and

$$\frac{1}{\sqrt{n}} M_n^{(1)} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma_{3,3}\right),$$

Therefore, if follows that

$$\frac{1}{\sqrt{n}} \mathbf{c}^\tau \begin{pmatrix} M_n^{(1,1)} \\ M_n^{(1,2)} \\ M_n^{(1)} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbb{E}\left(\left(c_1 X_{t-1,1} + c_2 X_{t-1,2} + c_3\right)^2 \left(R_t^1\right)^2 \right)\right)$$

for any $\mathbf{c} = (c_1, c_2, c_3)^\tau \in (\mathbb{R}^*)^3$. From the Cramer-Wold device, one can deduce that

$$\frac{1}{\sqrt{n}} \begin{pmatrix} M_n^{(1,1)} \\ M_n^{(1,2)} \\ M_n^{(1)} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma\right). \tag{19}$$

Thus, in the same way, we find

$$\frac{1}{\sqrt{n}} \begin{pmatrix} M_n^{(2,1)} \\ M_n^{(2,2)} \\ M_n^{(2)} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Gamma\right). \tag{20}$$

Combining (17) and (19) on the one hand, and (18) and (20) on the other hand, we conclude the proof of Theorem 2. ∎

Annex: some complements on the BSkellam($\lambda_0, \lambda_1, \lambda_2$) distribution

Basic properties

First of all, let us recall the definition of the Skellam distribution.

Definition 7 (*Skellam(λ_1, λ_2) distribution*) Let $\lambda_1 > 0$ and $\lambda_2 > 0$. We say that the random variable X has the Skellam distribution, denoted by Skellam(λ_1, λ_2), if and only if

$$X \stackrel{d}{=} U_1 - U_2,$$

where U_1 and U_2 are two independent random variables such that $U_i \sim$ Poisson(λ_i) for any $i \in \{1, 2\}$.

The probability mass function of X is given by

$$\mathbb{P}(X = x) = e^{-(\lambda_1 + \lambda_2)} \lambda_1^x \sum_{i=\max(0,-x)}^{\infty} \frac{(\lambda_1 \lambda_2)^i}{(x+i)! i!}, \qquad x \in \mathbb{Z}.$$

Further details can be found, e.g., in Skellam (1946), Karlis and Ntzoufras (2009) and Al-Zaid and Omair (2010).

Remark 19 Note that $X_1 \sim$ Skellam(λ_1, λ_0), $X_2 \sim$ Skellam(λ_2, λ_0) and $X_1 - X_2 \sim$ Skellam(λ_1, λ_2) if $(X_1, X_2) \sim$ BSkellam($\lambda_0, \lambda_1, \lambda_2$). Moreover, X_1 and X_2 are independent if and only if $\lambda_0 = 0$.

Lemma 1 below investigates some basic properties of the BSkellam($\lambda_0, \lambda_1, \lambda_2$) distribution.

Lemma 1 *Let $\lambda_0 \geq 0$, $\lambda_1 > 0$, $\lambda_2 > 0$ and $(X_1, X_2) \sim$ BSkellam($\lambda_0, \lambda_1, \lambda_2$). Then the characteristic function of (X_1, X_2) is*

$$\phi_{(X_1, X_2)}(t_1, t_2) = e^{\lambda_1(e^{it_1} - 1)} e^{\lambda_2(e^{it_2} - 1)} e^{\lambda_0(e^{-it_1 - it_2} - 1)}, \qquad (t_1, t_2) \in \mathbb{Z}^2.$$

Remark 20 Using the characteristic function of $(X_1, X_2) \sim$ BSkellam($\lambda_0, \lambda_1, \lambda_2$), the probability mass function of (X_1, X_2) can be expressed by

$$P(X_1 = x_1, X_2 = x_2) =$$
$$\frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi_{(X_1, X_2)}(t_1, t_2) e^{-ix_1 t_1} e^{-ix_2 t_2} dt_1 dt_2, \qquad (x_1, x_2) \in \mathbb{Z}^2.$$

In particular, this yields the following equality

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{\lambda_1(e^{it_1} - 1)} e^{\lambda_2(e^{it_2} - 1)} e^{\lambda_0(e^{-it_1 - it_2} - 1)} e^{-ixt_1} e^{-iyt_2} dt_1 dt_2 =$$

$$(2\pi)^2 e^{-(\lambda_1 + \lambda_2 + \lambda_0)} \lambda_1^x \lambda_2^y \sum_{i=\max(0,-x,-y)}^{\infty} \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x+i)!(y+i)! i!}, \qquad (x, y) \in \mathbb{Z}^2.$$

Lemma 2 below presents a Gaussian approximation result on the $\mathrm{BSkellam}(\lambda_0, \lambda_1, \lambda_2)$ distribution for large $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$.

Lemma 2 *For $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$ large enough, we have the approximation*

$$\mathrm{BSkellam}(\lambda_0, \lambda_1, \lambda_2) \approx \mathcal{N}_2 \left(\begin{pmatrix} \lambda_1 - \lambda_0 \\ \\ \lambda_2 - \lambda_0 \end{pmatrix}, \begin{pmatrix} \lambda_1 + \lambda_0 & \lambda_0 \\ \\ \lambda_0 & \lambda_2 + \lambda_0 \end{pmatrix} \right).$$

Estimation

In this section, we carry out the estimation of the unknown parameters via the method of moments and the method of the maximum likelihood. Let $\lambda_0 \geq 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$ be unknown parameters and $(X_{1,1}, X_{2,1}), \ldots, (X_{1,n}, X_{2,n})$ be n i.i.d. bivariate random variables with the common distribution $\mathrm{BSkellam}(\lambda_0, \lambda_1, \lambda_2)$.

The method of moments.

Set, for any $j \in \{1, 2\}$,

$$\overline{X}_j = \frac{1}{n} \sum_{i=1}^n X_{j,i}, \qquad S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{j,i} - \overline{X}_j)^2$$

and

$$C_{1,2} = \frac{1}{n-1} \sum_{i=1}^n (X_{1,i} - \overline{X}_1)(X_{2,i} - \overline{X}_2).$$

Due to Lemma 1, the method of moment gives

– the estimator

$$\hat{\lambda}_j = \frac{1}{2}(\overline{X}_j + S_j^2).$$

  for λ_j with $j \in \{1, 2\}$
– the estimator $\hat{\lambda}_0 = C_{1,2}$ for λ_0.

The maximum likelihood method.

For any $z = (x, y) \in \mathbb{Z}^2$ and $\lambda = (\lambda_0, \lambda_1, \lambda_2) \in [0, \infty[^3$, set

$$G(\lambda, z) = \lambda_1^x \lambda_2^y \sum_{i=0}^\infty \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x+i)!(y+i)!i!}.$$

Lemma 3 *For any* $(x, y) \in \mathbb{Z}^2$, *we have*

$$\frac{\partial G(\lambda, z)}{\partial \lambda_1} = \frac{x}{\lambda_1} G(\lambda, z) + \frac{\lambda_0}{\lambda_1} G(\lambda, (x + 1, y + 1)),$$

$$\frac{\partial G(\lambda, z)}{\partial \lambda_2} = \frac{y}{\lambda_2} G(\lambda, z) + \frac{\lambda_0}{\lambda_2} G(\lambda, (x + 1, y + 1))$$

and

$$\frac{\partial G(\lambda, z)}{\partial \lambda_0} = G(\lambda, (x + 1, y + 1)).$$

For any $i \in \{1, \ldots, n\}$, set $z_i = (x_{1,i}, x_{2,i}) \in \mathbb{Z}^2$, $z = (z_1, \ldots, z_n)$ and $\lambda = (\lambda_0, \lambda_1, \lambda_2) \in [0, \infty[^3$. The likelihood function is

$$L(\lambda, z) = \prod_{i=1}^{n} \mathbb{P}(X_1 = x_{1,i}, X_2 = x_{2,i}) = e^{-n(\lambda_0 + \lambda_1 + \lambda_2)} \prod_{i=1}^{n} G(\lambda, z_i).$$

Therefore the j-th partial derivative of the log–likelihood function is

$$\frac{\partial \ln L(\lambda, z)}{\partial \lambda_j} = -n + \sum_{i=1}^{n} \frac{\frac{\partial G(\lambda, z_i)}{\partial \lambda_j}}{G(\lambda, z_i)}.$$

for any $j \in \{0, 1, 2\}$. It follows from Lemma 3 that

$$\frac{\partial \ln L(\lambda, z)}{\partial \lambda_1} = -n + \frac{1}{\lambda_1} \sum_{i=1}^{n} x_{1,i} + \frac{\lambda_0}{\lambda_1} \sum_{i=1}^{n} \frac{G(\lambda, z_i + 1)}{G(\lambda, z_i)},$$

$$\frac{\partial \ln L(\lambda, z)}{\partial \lambda_2} = -n + \frac{1}{\lambda_2} \sum_{i=1}^{n} x_{2,i} + \frac{\lambda_0}{\lambda_2} \sum_{i=1}^{n} \frac{G(\lambda, z_i + 1)}{G(\lambda, z_i)}$$

and

$$\frac{\partial \ln L(\lambda, z)}{\partial \lambda_0} = -n + \sum_{i=1}^{n} \frac{G(\lambda, z_i + 1)}{G(\lambda, z_i)}.$$

Set $Z_i = (X_{1,i}, X_{2,i})$ for any $i \in \{1, \ldots, n\}$. The maximum likelihood estimate $\hat{\lambda} = (\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2)$ satisfies

$$\begin{cases} 0 = -n + \dfrac{1}{\hat{\lambda}_1} \displaystyle\sum_{i=1}^{n} X_{1,i} + \dfrac{\hat{\lambda}_0}{\hat{\lambda}_1} \displaystyle\sum_{i=1}^{n} \dfrac{G(\hat{\lambda}, Z_i + 1)}{G(\hat{\lambda}, Z_i)}, \\[4mm] 0 = -n + \dfrac{1}{\hat{\lambda}_2} \displaystyle\sum_{i=1}^{n} X_{2,i} + \dfrac{\hat{\lambda}_0}{\hat{\lambda}_2} \displaystyle\sum_{i=1}^{n} \dfrac{G(\hat{\lambda}, Z_i + 1)}{G(\hat{\lambda}, Z_i)}, \\[4mm] 0 = -n + \displaystyle\sum_{i=1}^{n} \dfrac{G(\hat{\lambda}, Z_i + 1)}{G(\hat{\lambda}, Z_i)}. \end{cases} \qquad (21)$$

Therefore $\sum_{i=1}^{n} G(\hat{\lambda}, Z_i + 1)/G(\hat{\lambda}, Z_i) = n$ and, a fortiori,

$$\hat{\lambda}_1 = \overline{X}_1 + \hat{\lambda}_0, \qquad \hat{\lambda}_2 = \overline{X}_2 + \hat{\lambda}_0. \tag{22}$$

Substituting (22) into the last equation in (21), one obtains $\hat{\lambda}_0$. Putting $\hat{\lambda}_0$ in(22), we obtain $\hat{\lambda}_1$ and $\hat{\lambda}_2$.

Numerical study

In this section we briefly demonstrate some properties of the bivariate Skellam distribution and the parameter estimation procedures. More precisely, we consider two settings (termed Setting 1 and 2 in the following), dealing with a Bivariate Skellam distribution with parameters $(\lambda_0, \lambda_1, \lambda_2) = (4, 2, 3)$ and $(\lambda_0, \lambda_1, \lambda_2) = (4, 35, 40)$, respectively. Figure 6 shows random samples with 500 observations each from Setting 1 in the left and Setting 2 in the right panel. The two settings are subject to different correlation, which equals 0.617 (respectively 0.0966), and their estimated values for the samples displayed are 0.628 (p < 2.2e-16) and 0.0975 (p = 0.0292).

In a first Monte Carlo experiment, we investigate the departure from normality in the two settings. For this purpose, we generated 1000 samples for both settings with 1000 observations each and perform the Shapiro-Wilk test for multivariate normality. In Setting 1, having comparably small values of λ_i, $i = 1, 2, 3$, normality is rejected at 5%-level for 87.2% of the samples. However, the corresponding proportion in the second setting reduces to 30.1%, supporting the results of Lemma 2.

In a second Monte Carlo experiment, we investigate the properties of the method of moments and maximum likelihood estimators. To this end, the two parameter sets described above were used for generating 100 sequences of observations of length $n = 50, 100$, and 200, respectively. Subsequently, the parameters were estimated by the method of moments and by maximization of the likelihood. Figure 7 shows Box plots of the estimated parameters, where the method of moments estimates have white background and the boxes corresponding to the maximum likelihood estimator are coloured gray. The dashed black lines represent the true parameter values. Several aspects are apparent: First, both estimators seem to be consistent. Second, both estimators seem unbiased. Third, the maximum likelihood estimator has a lower variability than the method of moments estimator. Finally, it may be noted that normality of the estimators cannot be rejected for the large majority of samples at 5%-level (30 of 36, Shapiro-Wilk).

Proofs

Proof of Lemma 1. Let $(t_1, t_2) \in \mathbb{Z}^2$. The characteristic function of (X_1, X_2) is then given by

$$
\begin{aligned}
\phi_{(X_1,X_2)}(t_1, t_2) &= \mathbb{E}(e^{it_1 X_1 + it_2 X_2}) = \mathbb{E}(e^{it_1(U_1 - U_0) + it_2(U_2 - U_0)}) \\
&= \mathbb{E}(e^{it_1 U_1 + it_2 U_2 - i(t_1 + t_2)U_0}) = \phi_{U_1}(t_1)\phi_{U_2}(t_2)\phi_{U_0}(-t_1 - t_2) \\
&= e^{\lambda_1(e^{it_1} - 1)} e^{\lambda_2(e^{it_2} - 1)} e^{\lambda_0(e^{-it_1 - it_2} - 1)}.
\end{aligned}
$$

This ends the proof of Lemma 1. ■

   Proof of Lemma 2. Set $W_1 = X_1/(\lambda_1 + \lambda_0)$ and $W_2 = X_2/(\lambda_2 + \lambda_0)$ and let $(t_1, t_2) \in \mathbb{Z}^2$. Using Lemma 1 and $e^x \underset{0}{\sim} 1 + x + x^2/2$, we have

$$
\begin{aligned}
\phi_{(W_1,W_2)}(t_1, t_2) &= \phi_{(X_1,X_2)}(t_1/(\lambda_1 + \lambda_0), t_2/(\lambda_2 + \lambda_0)) \\
&= e^{\lambda_1(e^{it_1/(\lambda_1 + \lambda_0)} - 1)} e^{\lambda_2(e^{it_2/(\lambda_2 + \lambda_0)} - 1)} e^{\lambda_0(e^{-it_1/(\lambda_1 + \lambda_0) - it_2/(\lambda_2 + \lambda_0)} - 1)} \\
&\approx e^{\lambda_1(it_1/(\lambda_1 + \lambda_0) - (1/2)t_1^2/(\lambda_1 + \lambda_0)^2)} e^{\lambda_2(it_2/(\lambda_2 + \lambda_0) - (1/2)t_2^2/(\lambda_2 + \lambda_0)^2)} \times \\
&\quad e^{\lambda_0(-it_1/(\lambda_1 + \lambda_0) - it_2/(\lambda_2 + \lambda_0) - (1/2)(t_1/(\lambda_1 + \lambda_0) + t_2/(\lambda_2 + \lambda_0))^2)} \\
&= e^{it_1(\lambda_1 - \lambda_0)/(\lambda_1 + \lambda_0)} e^{it_2(\lambda_2 - \lambda_0)/(\lambda_2 + \lambda_0)} \times \\
&\quad e^{-(1/2)(t_1^2/(\lambda_1 + \lambda_0) + t_2^2/(\lambda_2 + \lambda_0) + 2t_1 t_2 \lambda_0/(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0))}.
\end{aligned}
$$

for any $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$ large enough. The last term is the characteristic function of a bivariate random variable

$$
Z \sim \mathcal{N}_2\left(\begin{pmatrix} (\lambda_1 - \lambda_0)/(\lambda_1 + \lambda_0) \\ (\lambda_2 - \lambda_0)/(\lambda_2 + \lambda_0) \end{pmatrix}, \begin{pmatrix} 1/(\lambda_1 + \lambda_0) & \lambda_0/(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0) \\ \lambda_0/(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0) & 1/(\lambda_2 + \lambda_0) \end{pmatrix} \right).
$$

Therefore (W_1, W_2) can be approximated by this distribution and, a fortiori,

$$
(X_1, X_2) \approx \mathcal{N}_2\left(\begin{pmatrix} \lambda_1 - \lambda_0 \\ \lambda_2 - \lambda_0 \end{pmatrix}, \begin{pmatrix} \lambda_1 + \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_2 + \lambda_0 \end{pmatrix} \right).
$$

Lemma 2 is proved. ■

Proof of Lemma 3. For any $(x, y) \in \mathbb{Z}^2$, we have

$$
\begin{aligned}
\frac{\partial G(\lambda, z)}{\partial \lambda_1} &= \lambda_1^x \lambda_2^y \sum_{i=0}^{\infty} \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x+i)!(y+i)!i!} \\
&= x \lambda_1^{x-1} \lambda_2^y \sum_{i=0}^{\infty} \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x+i)!(y+i)!i!} + \lambda_1^x \lambda_2^y \sum_{i=1}^{\infty} i \frac{(\lambda_0 \lambda_2)^i \lambda_1^{i-1}}{(x+i)!(y+i)!i!} \\
&= \frac{x}{\lambda_1} G(\lambda, z) + \lambda_1^x \lambda_2^y \sum_{i=0}^{\infty} \frac{(\lambda_0 \lambda_2)^{i+1} \lambda_1^i}{(x+i+1)!(y+i+1)!i!} \\
&= \frac{x}{\lambda_1} G(\lambda, z) + \frac{\lambda_0}{\lambda_1} G(\lambda, (x+1, y+1)).
\end{aligned}
$$

The other equalities can be proved similarly. Lemma 3 is proved. ∎

Figures
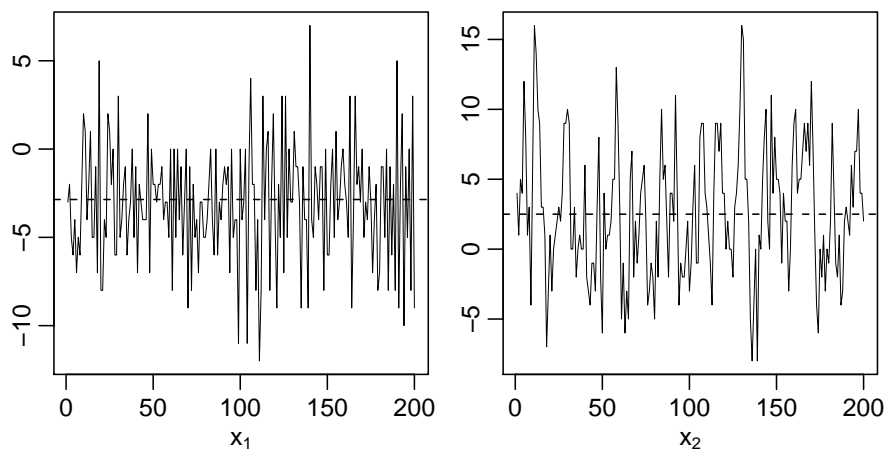
Fig. 1 Sample with 200 observations generated from a SBP-SINAR(1) model with true parameters $(\lambda_0, \lambda_1, \lambda_2) = (7, 3, 8)$ and $(\alpha_1, \alpha_2) = (0.3, 0.8)$. Dashed horizontal lines represent the true means.
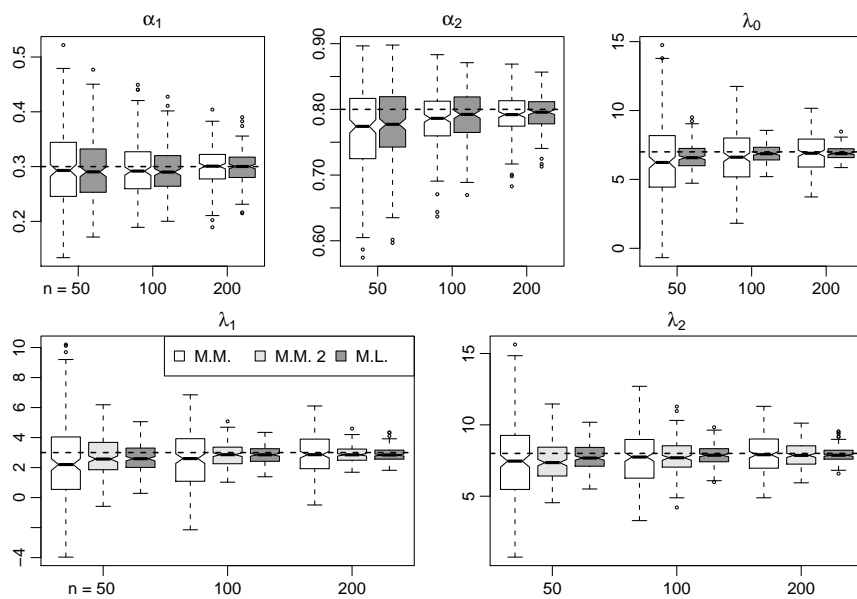
Fig. 2 Parameter estimates from 50 sequences of length $n = 50$, 100, and 200 simulated from a SBB-SINAR(1) model with true parameters $(\lambda_0, \lambda_1, \lambda_2) = (7, 3, 8)$ and $(\alpha_1, \alpha_2) = (0.3, 0.8)$. Box plots corresponding to the method of moments (M.M.) and maximum likelihood estimator (M.L.) have white and gray background, respectively. The dashed black lines represent the true parameter values.
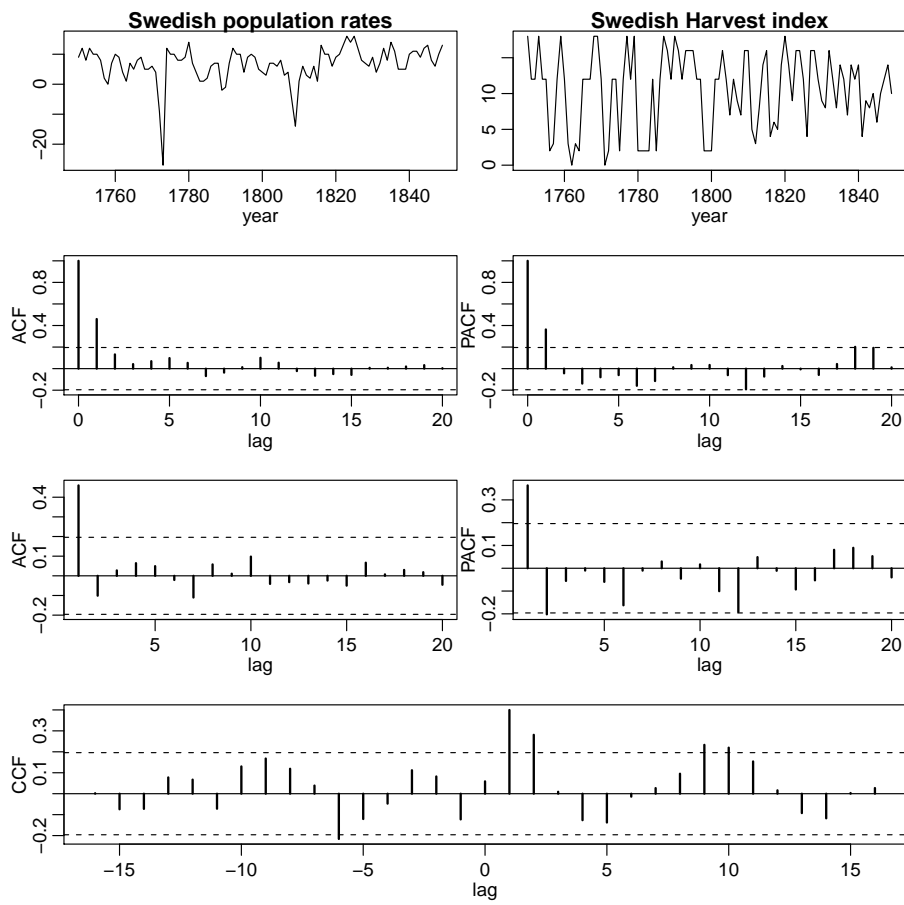
Fig. 3 Annual Swedish population rates (left top panel) and annual Swedish Harvest index (right top panel), 1750 − 1849, with corresponding ACFs and PACFs (middle panels) and CCF (lower panel)
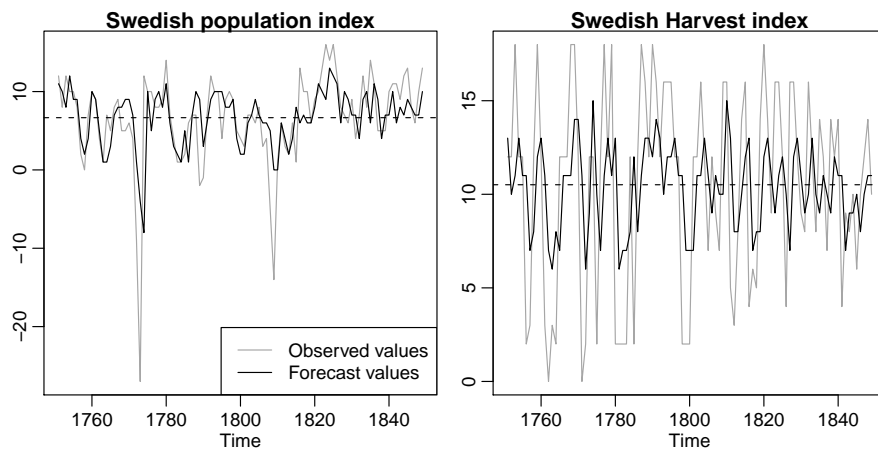
Fig. 4 Comparison between the observed values of the Swedish population (left panel) and Harvest index series (right panel), respectively, and the corresponding one-step ahead forecasting based on the fitted B-SINAR(1) model of first kind. The dashed horizontal lines represent the observed mean values of the series.
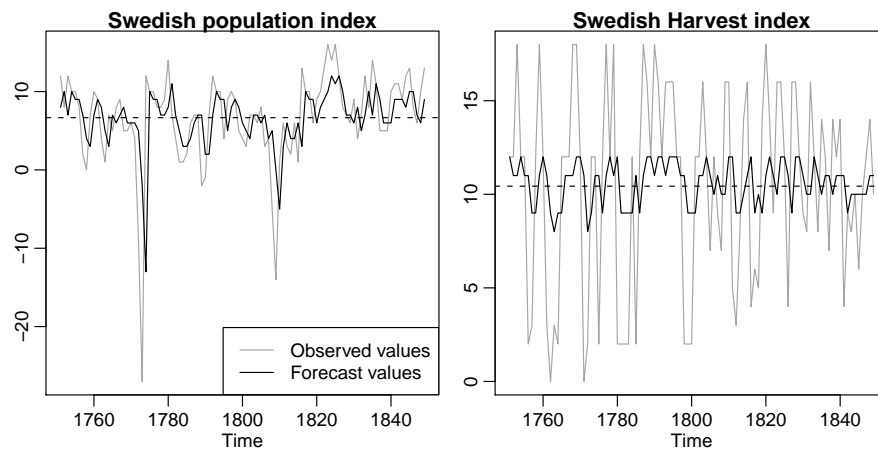
Fig. 5 Comparison between the observed values of the Swedish population (left panel) and Harvest index series (right panel), respectively, and the corresponding one-step ahead forecasting based on the fitted SBP-SINAR(1) model. The dashed horizontal lines represent the observed mean values of the series.
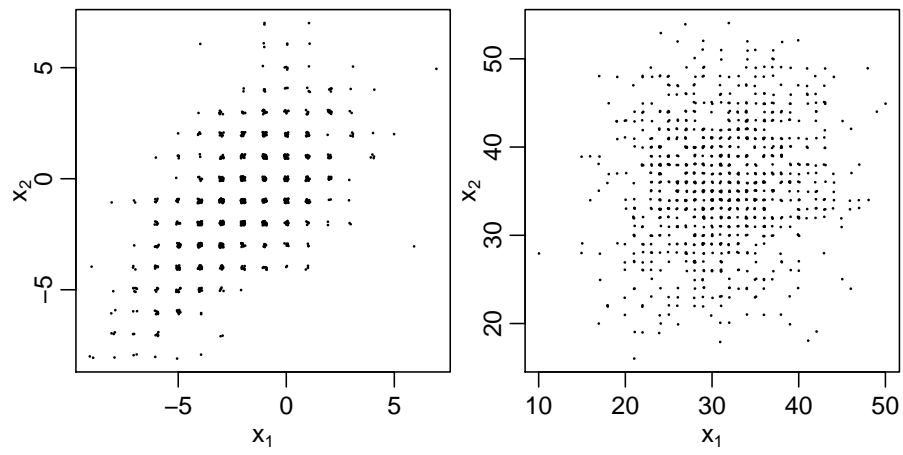
Fig. 6 Two samples with 500 observations each, generated from a bivariate Skellam distribution with true parameters $(\lambda_0, \lambda_1, \lambda_2) = (4, 2, 3)$ and $(\lambda_0, \lambda_1, \lambda_2) = (4, 35, 40)$ in the left (respectively right) panel.
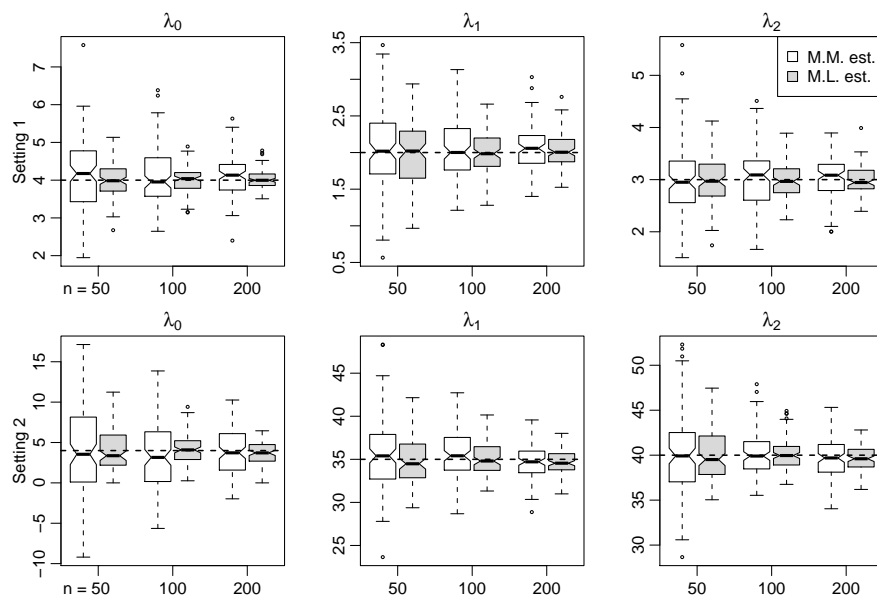
Fig. 7 Parameter estimates from 100 from simulated sequences of length $n = 50$, 100, and 200. The upper three panels were simulated from a bivariate Skellam with parameters $(\lambda_0, \lambda_1, \lambda_2) = (4, 2, 3)$, the lower three panels result from the parameters $(\lambda_0, \lambda_1, \lambda_2) = (4, 35, 40)$ in the left and right panel, respectively. Box plots corresponding to the method of moments (M.M.) and maximum likelihood (M.L.) estimator have white (respectively grey) background. The dashed black lines represent the true parameter values.

References

Al-zaid, A.A. and Al-Osh, M. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.*, 8, (3), 261-275.

Al-Zaid, A.A. and Omair, M.A. (2010). On the poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society*, 8, (33), 17-45.

Billingsley, A. (1961). *Statistical Inference for Markov Processes*, University of Chicago Press.

Brännäs, K. and Nordström, J. (2000). A Bivariate Integer Valued Allocation Model for Guest Nights in Hotels and Cottages. *Umea Economic Studies*, 547.

Du, J.-G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *J. Times Ser. Anal.*, 12, 129-142.

Duflo, M. (1997). *Random Iterative Models*, Volume 34 of Applications of Mathematics (New York). Springer-Verlag, Berlin.

Franke, J., and Subba Rao, T. (1993). Multivariate first-order integer-valued autoregressions, *Technical report, Forschung Universitat Kaiserslautern.*

Gauthier, G. and Latour, A. (1994). Convergence forte des estimateurs des paramètres d'un processus GENAR(p). *Ann. Sc. Math. Quebec*, 18, 37-59.

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

Heinen, A. and Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, 14, 564-583.

Jung, R.C. and Tremayne, A.R. (2006). Binomial thinning models for integer time series. *Stat. Model.*, 6, (2), 81-96.

Kachour, M. and Truquet, L. (2011). A p-order signed integer-valued autoregressive (SINAR(1)) model. *Journal of Time Series Analysis*, 32, (4), 223-236.

Karlis, D. and Ntzoufras, I. (2006). Bayesian analysis of the differences of count data. *Statistics in medicine*, 25, (11), 1885-1905.

Karlis, D. and Ntzoufras, I. (2009). Bayesian modeling of football outcomes: Using the Skellam distribution for the goal difference. *Journal Management Mathematics*, 20, (2), 133-145.

Kedem, B. and Fokianos, K. (2002). *Regression models for time series analysis*. Wiley-Interscience New-York.

Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic process. *Ann. Statist.*, 6, 629-642.

Latour, A. (1997) The multivariate GINAR(p) process. *Adv. Appl. Prob.*, 29, 228-248.

Latour, A. and Truquet, L. (2008). An integer-valued bilinear type model. Preprint.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 21, (4), 645-650.

McKenzie, E. (2003). Discrete variate time series. In Stochastic processes: modeling and simulation, volume 21 of Handbook of Statist., pages 573-606. North- Holland, Amsterdam.

Pedeli, X. and Karlis, D. (2011). A bivariate Poisson INAR(1) model with application, *Statistical Modeling*, 11, 325-349.

Quoreshi, A.M.M.S. (2006). Bivariate Time Series Modeling of Financial Count Data. *Commun. Stat. Theory Meth.*, 35, 1343-1358.

Skellam, J.G. (1946). The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society*, Series A, 109-296.

Steutel, F.W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Probab.*, 7, (5), 893-899.

Thomas, D.S. (1940). Social and Economic aspects of Swedish population mouvements. Macmillan, New York.

Weiß, C.H. (2008). Thinning operations for modeling time series of counts a survey. *J. Appl. Probab.*, 92, (3), 319-341.

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

140, (3), 667-683.