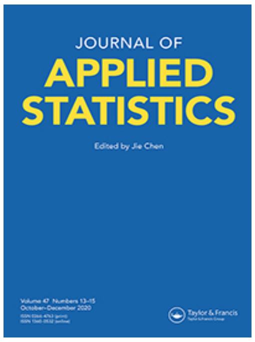

Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Inference for bivariate integer-valued moving average models based on binomial thinning operation

Isabel Silva, Maria Eduarda Silva & Cristina Torres

To cite this article: Isabel Silva, Maria Eduarda Silva & Cristina Torres (2020) Inference for bivariate integer-valued moving average models based on binomial thinning operation, Journal of Applied Statistics, 47:13-15, 2546-2564, DOI: 10.1080/02664763.2020.1747411

To link to this article: https://doi.org/10.1080/02664763.2020.1747411

Published online: 01 Apr 2020.

JOURNAL OF APPLIED STATISTICS
2020, VOL. 47, NOS. 13–15, 2546–2564
https://doi.org/10.1080/02664763.2020.1747411

Inference for bivariate integer-valued moving average models based on binomial thinning operation

Isabel Silva [a], Maria Eduarda Silva [b] and Cristina Torres [c]

[a]Faculdade de Engenharia da Universidade do Porto and CIDMA, Porto, Portugal; [b]Faculdade de Economia da Universidade do Porto and CIDMA, Porto, Portugal; [c]ISCAP-IPP, Porto, Portugal

ABSTRACT

Time series of (small) counts are common in practice and appear in a wide variety of fields. In the last three decades, several models that explicitly account for the discreteness of the data have been proposed in the literature. However, for multivariate time series of counts several difficulties arise and the literature is not so detailed. This work considers Bivariate INteger-valued Moving Average, BINMA, models based on the binomial thinning operation. The main probabilistic and statistical properties of BINMA models are studied. Two parametric cases are analysed, one with the cross-correlation generated through a Bivariate Poisson innovation process and another with a Bivariate Negative Binomial innovation process. Moreover, parameter estimation is carried out by the Generalized Method of Moments. The performance of the model is illustrated with synthetic data as well as with real datasets.

ARTICLE HISTORY
Received 27 November 2018
Accepted 21 March 2020

KEYWORDS
Bivariate discrete distributions; bivariate models; generalized method of moments; moving average; time series of counts

1. Introduction

This paper aims at contributing to integer-valued bivariate time series modelling. The analysis of integer-valued and, in particular, count time series has recently become an active area of research. In fact, time series of counts of certain events or objects in specified time intervals arise in many different contexts such as social science, biology and environmental processes, economics and finance, telecommunications and insurance, see *inter alia* [2,6,8,11,16,21,25,29,40,46]. In many cases the discrete variates are large numbers and it may make sense to approximate them by continuous variates. Often, however, this is not possible or even desirable and it is necessary to develop appropriate modelling strategies for the statistical analysis of time series of counts. A popular approach to model time series of counts is based on the binomial thinning operation proposed by [39] and defined as follows.

Let X be a non-negative integer-valued random variable. Then, for any $\alpha \in [0, 1]$ define the binomial thinning operation as

$$\alpha \circ X := \begin{cases} \sum_{i=1}^{X} Y_i, & X > 0, \\ 0, & X = 0, \end{cases} \tag{1}$$

CONTACT Isabel Silva ims@fe.up.pt

where $\{Y_i\}_i$ is a sequence of independent and identically distributed Bernoulli random variables with $\mathrm{P}(Y_i = 1) = \alpha$, called the counting series of $\alpha \circ X$, which is also independent of X. Note that given X, $\alpha \circ X$ has a binomial distribution with parameters (X, α), $\alpha \circ X | X \sim \mathrm{Bi}(X, \alpha)$. The main properties of the binomial thinning operation are stated in [35].

Models based on the binomial thinning operation were proposed by Alzaid and Al-Osh [3] and McKenzie[26,27], giving rise to the class of (univariate) INteger-valued ARMA (INARMA) models. The literature on these models has been mainly focused in the study of AR (INAR) models and contributions to the study of MA models based upon binomial thinning operation (INMA) are mainly due to [4,5,27]. There are several reasons for this seemingly lack of interest in models with moving average (MA) components, one of which is that MA models are not Markovian.

The (univariate) INMA(q) process satisfies

$$X_t = \beta_0 \circ_t \varepsilon_t + \beta_1 \circ_t \varepsilon_{t-1} + \cdots + \beta_q \circ_t \varepsilon_{t-q}, \tag{2}$$

with $\beta_0 = 1$ and the index t in $\circ_t$ emphasizing the fact that the thinning operations are performed at each time t (in general, this time index is omitted if there is no risk of misunderstanding). Note that in this model each ε_t has a fixed limited maximum life time of $q + 1$ time units, at times $t, t + 1, \ldots, t + q$. In order to completely define the INMA(q) model, it is necessary to introduce a dependence structure between the thinning operations. This dependence may occur between the thinning operations at time t (that is, $\beta_i \circ_t \varepsilon_{t-i}$ and $\beta_j \circ_t \varepsilon_{t-j}$, for all $i \neq j$) or between the thinning operations involving the same variable at times $t + i$ and $t + j$, (that is, $\beta_i \circ_{t+i} \varepsilon_t$ and $\beta_j \circ_{t+j} \varepsilon_t$,) for all $i \neq j$. For a detailed discussion, see [5,41].

In the model proposed by Al-Osh and Alzaid [4], also denominated changing states model, at each time t (that is, in the time interval $]\, t - 1, t]$), ε_t elements enter the system, comprising the generation t: $E_{t,1}, E_{t,2}, \ldots, E_{t,\varepsilon_t}$. These elements can be *active* or *inactive* during each of their $q + 1$ time units of total life time (that is, the elements can come and go several times during their life times) according to a certain set of probabilities defined by the following Bernoulli random vector $(Y^{(t)}_{(i,1)}, Y^{(t)}_{(i,2)}, \ldots, Y^{(t)}_{(i,q)})$, where $Y^{(t)}_{(i,k)} = 1$ if the ith element of ε_t is *active* in the system at time $t + k$, for $i \in \mathbb{N}; t \in \mathbb{Z}$ and $k = 1, \ldots, q$ (naturally, $Y^{(t)}_{(i,r)} = 0$, otherwise). Moreover, at time t each element of generation $t - i, 0 \leq i \leq q$, has the probability β_i of being *active* (independently of the other elements in the system), that is, $\beta_k = Pr(Y^{(t)}_{(i,k)} = 1)$, for $k = 1, \ldots, q$. Then, $\beta_k \circ_t \varepsilon_{t-k}$ denotes the number of elements of generation $t{-}k$ which are *active* in the system at time t. Thus, $Pr(Y^{(t)}_{(i,k_1)} = 1, Y^{(t)}_{(i,k_2)} = 1, \ldots, Y^{(t)}_{(i,k_j)} = 1) = \prod_{m=1}^{j} \beta_{k_m - k_{m-1}}$, for all $k_0 = 0 \leq k_1 < k_2 < \cdots < k_j \leq q$, generates a dependence structure for $(\beta_0 \circ_t \varepsilon_t, \beta_1 \circ_{t+1} \varepsilon_t, \beta_2 \circ_{t+2} \varepsilon_t, \ldots, \beta_q \circ_{t+q} \varepsilon_t)$, conditional to ε_t. All the other thinning operations are independent.

The INMA models have been extended to threshold INMA models, [47], INMA models with structural changes, [44] and Poisson combined INMA(q) models, [45]. Furthermore, a new INMA(1) model based on the negative binomial thinning operation was proposed in [36]. Additionally, likelihood-based inference was efficiently implemented by [42,43] and diagnostic tests regarding the marginal distribution and the autocorrelation structure of INMA(1) models were proposed in [1].

In several application areas, the need for developing statistical modelling approaches for time series of multivariate count responses is increasing. In the multivariate context, once again, the focus of the literature has been almost solely on AR specifications, see [13,14,17,19,24,29,30,33,34,38] and references therein. A noteworthy exception is the bivariate integer-valued moving average, BINMA, model proposed by Quoreshi [31] and the work of Mamode Khan *et al.* [23,24,38,39] and Silva *et al.* [37]. Quoreshi in [31] proposed a model for the number of transactions in equidistant time intervals with contemporaneous cross-correlation (which may be positive or negative), assuming independence between and within the thinning operations.

In this paper we consider an extension of Quoreshi's BINMA model which builds on the univariate INMA model of Al-Osh and Alzaid [4] introduced above. To motivate this model, consider counting the occurrence of a certain phenomenon (with a finite maximum life-time) in two different locations. There is dependence between the counts at the two locations and, on the other hand, there is serial correlation between counts within the same location. The BINMA model proposed in this work is a bivariate discrete time process that comprises a wide range of auto and cross-correlation structures. Furthermore, can also account for overdispersion when the innovations follow a Bivariate Negative Binomial distribution. Probabilistic and statistical properties are studied and estimation is accomplished by Generalized Method of Moments (GMM), [9,10,12]. We examine the finite sample behaviour of GMM in bivariate INMA time series models using Monte Carlo methods. Finally, we illustrate the application of the BINMA model to a real dataset.

2. Bivariate INteger-valued Moving Average, BINMA(q_1, q_2), model

This section introduces a bivariate INMA model which entertains the dependence structure proposed by Al-Osh and Alzaid [4], thus extending the BINMA model proposed by Quoreshi [31].

Let $\{\mathbf{X}_t\} = \{(X_{1,t}, X_{2,t})\}, t \in \mathbb{Z}$, be a non-negative integer-valued bivariate random variable. Then $\{\mathbf{X}_t\}$ is a Bivariate INteger-valued Moving Average model of order (q_1, q_2), BINMA(q_1, q_2), if satisfies the following recursions

$$\begin{aligned} X_{1,t} &= \varepsilon_{1,t} + \beta_{1,1} \circ \varepsilon_{1,t-1} + \cdots + \beta_{1,q_1} \circ \varepsilon_{1,t-q_1}, \\ X_{2,t} &= \varepsilon_{2,t} + \beta_{2,1} \circ \varepsilon_{2,t-1} + \cdots + \beta_{2,q_2} \circ \varepsilon_{2,t-q_2}, \end{aligned} \tag{3}$$

where '$\circ$' denotes the binomial thinning operation defined in (1), $\beta_{j,r} \in [0,1]$, $\beta_{j,q_j} \neq 0$, for $r = 1, \ldots, q_j; j = 1, 2$, and $\{\boldsymbol{\varepsilon}_t\} = \{(\varepsilon_{1,t}, \varepsilon_{2,t})\}, t \in \mathbb{Z}$, is an i.i.d. sequence of bivariate random variables, usually called **innovation process**, with $\mathrm{E}[\boldsymbol{\varepsilon}_t] = \boldsymbol{\mu} = (\mu_1, \mu_2)$, $\mathrm{Var}[\boldsymbol{\varepsilon}_t] = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ with diagonal elements $\mathrm{Var}[\varepsilon_{j,t}] = \sigma_j^2$ and off diagonal elements $\mathrm{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) = \Lambda \in \mathbb{R}$.

In (3), each equation is a changing states INMA(q_j) model, as defined by (2). Therefore, the serial correlation (within each equation) arises from the special dependence structure between the binomial thinning operations explained in Section 1 while the cross-correlation (between equations) is induced by the bivariate innovation process $\{\boldsymbol{\varepsilon}_t\}$. As a consequence, $\beta_{1,k_1} \circ \varepsilon_{1,t-k_1}$ and $\beta_{2,k_2} \circ \varepsilon_{1,t-k_2}$ are independent for $k_1 \neq k_2$. Note that the special case $q_1 = q_2 = 1$ corresponds to the BINMA(1, 1) model proposed in [31].

The mean and variance of this changing states BINMA model are given by (for $j = 1, 2$)

$$\mathrm{E}[X_{j,t}] = \mu_{X_j} = \mu_j \left(1 + \sum_{i=1}^{q_j} \beta_{j,i}\right),$$
$$\mathrm{Var}[X_{j,t}] = \sigma^2_{X_j} = \sigma^2_j \left(1 + \sum_{i=1}^{q_j} \beta^2_{j,i}\right) + \mu_j \sum_{i=1}^{q_j} \beta_{j,i}(1 - \beta_{j,i}).$$

The autocovariance function, $\gamma_{X_j}(k) = \mathrm{Cov}(X_{j,t-k}, X_{j,t}), j = 1, 2$, can be written as:

$$\gamma_{X_j}(k) = \sigma^2_j \left(\beta_{j,k} + \sum_{i=1}^{q_j-k} \beta_{j,i}\beta_{j,k+i}\right) + \mu_j \sum_{i=1}^{q_j-k} \beta_{j,i}(\beta_{j,k} - \beta_{j,k+i}),$$

for $k = 1, \ldots, q_j$, being zero after lag $q_j + 1$.

Finally, the cross-covariance function of the model is defined as (proof in the Appendix 1),

$$\gamma_{X_1,X_2}(0) = \mathrm{Cov}(X_{1,t}, X_{2,t}) = \Lambda \left(1 + \sum_{i=1}^{\min(q_1,q_2)} \beta_{1,i}\beta_{2,i}\right),$$
$$\gamma_{X_1,X_2}(k) = \mathrm{Cov}(X_{1,t}, X_{2,t-k}) = \begin{cases} \Lambda \left(\beta_{1,k} + \sum_{i=1}^{q_1-k} \beta_{1,k+i}\beta_{2,i}\right), & k = 1, \ldots, q_1, \\ 0, & k \geq q_1 + 1, \end{cases}$$
$$\gamma_{X_2,X_1}(k) = \mathrm{Cov}(X_{1,t-k}, X_{2,t}) = \begin{cases} \Lambda \left(\beta_{2,k} + \sum_{i=1}^{q_2-k} \beta_{2,k+i}\beta_{1,i}\right), & k = 1, \ldots, q_2, \\ 0, & k \geq q_2 + 1. \end{cases}$$

Note that both the contemporaneous and the lagged cross-correlation may be positive or negative.

3. Parametric cases

We now present two parametric cases for the innovation process in the BINMA(q_1, q_2) model.

3.1. BINMA (q_1, q_2) model with bivariate Poisson innovations

Lets consider that the innovation process, $\{\boldsymbol{\varepsilon}_t\}$, follows a Bivariate Poisson (BP) distribution. One of the approaches to generate the Bivariate Poisson distribution is the trivariate reduction method (see [18]). Let X_1, X_2 and X_0 be independently distributed Poisson variables with means $\lambda_1, \lambda_2 \in \mathbb{R}^+$ and $\phi \in [\,0, \min(\lambda_1, \lambda_2)[$, respectively and define $X = X_1 + X_0$ and $Y = X_2 + X_0$. Then, the bivariate random variable (X, Y) has a Bivariate Poisson distribution with parameters λ_1, λ_2 and ϕ, $(X, Y) \sim \mathrm{BP}(\lambda_1, \lambda_2, \phi)$. Marginally, $X \sim P(\lambda_1 + \phi)$ and $Y \sim P(\lambda_2 + \phi)$ and the covariance of X and Y is ϕ. The joint probability function of (X, Y) is given by $Pr[X = x, Y = y] = \mathrm{e}^{-(\lambda_1+\lambda_2+\phi)} \sum_{i=0}^{\min(x,y)} \frac{\lambda_1^{x-i}\lambda_2^{y-i}\phi^i}{(x-i)!(y-i)!i!}$. The joint probability generating function (pgf) of the Bivariate Poisson distribution is given by $G(s_1, s_2) = \exp((\lambda_1 + \phi)(s_1 - 1) + (\lambda_2 + \phi)(s_2 - 1) + \phi(s_1 - 1)(s_2 - 1))$. For a comprehensive treatment of the Bivariate Poisson distribution see [18].

Now, assuming that $\boldsymbol{\varepsilon}_t \sim \mathrm{BP}(\lambda_1, \lambda_2, \phi)$, for $\lambda_1, \lambda_2 > 0$ and $\phi \in [0, \min(\lambda_1, \lambda_2)[$, in the BINMA model defined in (3), the mean and variance simplify to $\mathrm{E}[X_{j,t}] = \mathrm{Var}[X_{j,t}] = (\lambda_j + \phi)(1 + \sum_{i=1}^{q_j} \beta_{j,i})$. The autocovariance and cross-covariance functions are given by

$$\gamma_{X_j}(k) = (\lambda_j + \phi)\left(\beta_{j,k} + \sum\nolimits_{i=1}^{q_j-k} \beta_{j,i}\beta_{j,k}\right),$$

$$\gamma_{X_1,X_2}(0) = \phi\left(1 + \sum\nolimits_{i=1}^{\min(q_1,q_2)} \beta_{1,i}\beta_{2,i}\right),$$

$$\gamma_{X_1,X_2}(k) = \phi\left(\beta_{1,k} + \sum\nolimits_{i=1}^{q_1-k} \beta_{1,k+i}\beta_{2,i}\right),$$

$$\gamma_{X_2,X_1}(k) = \phi\left(\beta_{2,k} + \sum\nolimits_{i=1}^{q_2-k} \beta_{2,k+i}\beta_{1,i}\right),$$

for $k = 1, \ldots, q_j$, respectively; being zero after lag $q_j + 1$. Note that now the cross-correlation is always positive.

Remark 3.1: The simplest model is the first-order BINMA model with Bivariate Poisson innovations:

$$\begin{aligned} X_{1,t} &= \varepsilon_{1,t} + \beta_1 \circ \varepsilon_{1,t-1}, \\ X_{2,t} &= \varepsilon_{2,t} + \beta_2 \circ \varepsilon_{2,t-1}, \end{aligned} \tag{4}$$

where '$\circ$' denotes the binomial thinning operation defined in (1), $\beta_j \in]0, 1]$, for $j = 1, 2$, and the innovation process follows a Bivariate Poisson distribution with parameters $(\lambda_1, \lambda_2, \phi)$. Following [4] we obtain the conditional mean and variance for the marginal processes,

$$\mathrm{E}[X_{j,t}|\mathcal{F}_{j,t-1}] = (\lambda_j + \phi) + \frac{\beta_j X_{j,t-1}}{1 + \beta_j}, \quad \mathrm{Var}[X_{j,t}|\mathcal{F}_{j,t-1}] = (\lambda_j + \phi) + \frac{\beta_j X_{j,t-1}}{(1 + \beta_j)^2}, \tag{5}$$

for $j = 1, 2$ and where $\mathcal{F}_{j,t-1}$ be the σ-algebra generated by $\{X_{j,1}, \ldots, X_{j,t-1}\}$.

3.2. BINMA(q_1, q_2) model with bivariate negative binomial innovations

It is well known that Poisson distribution is not suitable for modelling and analysis of integer-valued time series when equidispersion condition is not satisfied. In an attempt to overcome this problem, a BINMA(q_1, q_2) model with Bivariate Negative Binomial innovations is now defined. Consider then that the innovation process, $\{\boldsymbol{\varepsilon}_t\}$, follows a Bivariate Negative Binomial (BNB) distribution. As in the univariate case, there are several ways to construct the BNB distribution. In what follows it was considered a BNB distribution that is obtained by a gamma mixture of two independent Poisson random variables, proposed by Cheon *et al.* [7] and Marshall and Olkin [20], called BNB-type I distribution. Let $\theta \sim Gamma(\tau^{-1}, \tau^{-1})$ and $X_i|\theta \sim \mathrm{P}(\theta\lambda_i), i = 1, 2$. Then, the bivariate random variable (X_1, X_2) has BNB-type I distribution with parameters λ_1, λ_2 and τ, $(X_1, X_2) \sim \mathrm{BNB}(\lambda_1, \lambda_2, \tau)$. Note that the marginal distribution of X_j is univariate Negative Binomial with parameters τ^{-1} and $p_j = \tau^{-1}/(\lambda_j + \tau^{-1})$, $j = 1, 2$, and the covariance of X_1 and X_2

is $\lambda_1\lambda_2\tau$. The joint probability function of (X_1, X_2) is given by $Pr[X_1 = x_1, X_2 = x_2] = \frac{\Gamma(\tau^{-1}+x_1+x_2)}{\Gamma(\tau^{-1})\Gamma(x_1+1)\Gamma(x_2+1)}\lambda_1^{x_1}\lambda_2^{x_2}\tau^{-\tau^{-1}}(\lambda_1+\lambda_2+\tau^{-1})^{-(x_1+x_2+\tau^{-1})}$, where $\lambda_1, \lambda_2, \tau > 0$ (τ is a dispersion parameter) and $\Gamma(\cdot)$ is the gamma function. For a comprehensive treatment of the Bivariate Negative Binomial-type I distribution see [7,20].

Now, assuming that $\boldsymbol{\varepsilon}_t \sim \text{BNB}(\lambda_1, \lambda_2, \tau)$, $\lambda_1, \lambda_2, \tau > 0$ in the changing states BINMA model (3), the mean and variance are given by $\text{E}[X_{j,t}] = \lambda_j(1 + \sum_{i=1}^{q_j}\beta_{j,i})$, and $\text{Var}[X_{j,t}] = \lambda_j(1 + \sum_{i=1}^{q_j}\beta_{j,i}) + \tau\lambda_j^2(1 + \sum_{i=1}^{q_j}\beta_{j,i}^2)$, respectively. The autocovariance and cross-covariance functions can be written as

$$\gamma_{X_j}(k) = \lambda_j\left(\beta_{j,k} + \sum_{i=1}^{q_j-k}\beta_{j,i}\beta_{j,k}\right) + \tau\lambda_j^2\left(\beta_{j,k} + \sum_{i=1}^{q_j-k}\beta_{j,i}\beta_{j,k+i}\right),$$

$$\gamma_{X_1,X_2}(0) = \lambda_1\lambda_2\tau\left(1 + \sum_{i=1}^{\min(q,q_2)}\beta_{1,i}\beta_{2,i}\right),$$

$$\gamma_{X_1,X_2}(k) = \lambda_1\lambda_2\tau\left(\beta_{1,k} + \sum_{i=1}^{q_1-k}\beta_{1,k+i}\beta_{2,i}\right),$$

$$\gamma_{X_2,X_1}(k) = \lambda_1\lambda_2\tau\left(\beta_{2,k} + \sum_{i=1}^{q_2-k}\beta_{2,k+i}\beta_{1,i}\right),$$

for $k = 1, \ldots, q_j$, respectively; being zero after lag $q_j + 1$.

Note that, for this model, the index of dispersion is given by $\frac{\sigma_{X_j}^2}{\mu_{X_j}} = 1 + \tau\lambda_j\frac{1+\sum_{i=1}^{q_j}\beta_{j,i}^2}{1+\sum_{i=1}^{q_j}\beta_{j,i}} > 1$, which means that this model accounts for overdispersion.

4. Parameter estimation and Monte Carlo simulation results

In this section, we discuss a Generalized Method of Moments (GMM) approach to the estimation of the BINMA (q_1, q_2) model from an estimating functions perspective. In fact, estimating functions provide a very general framework for statistical inference and are particularly useful in non-Gaussian settings such as count data. We further illustrate the small sample properties in the particular case of Bivariate Poisson and Bivariate Negative Binomial innovations.

4.1. Generalized method of moments

To estimate the parameters of the proposed model, we opt for the Generalized Method of Moment (GMM) methodology based on first- and second-order moments of the process. The GMM estimator (see Appendix 2) was firstly introduced by Hansen [9] into the econometric literature and, since then, has been widely applied in several fields.

Suppose that we have an observed sample $\{\mathbf{X}_t\} = \{(X_{1,t}, X_{2,t})\}$, $t = 1, \ldots, n$, of a BINMA(q_1, q_2) process defined as in (3) with $(5 + q_1 + q_2)$ unknown parameters $\boldsymbol{\theta} = (\beta_{1,1}, \ldots, \beta_{1,q_1}, \mu_1, \sigma_1^2, \beta_{2,1}, \ldots, \beta_{2,q_2}, \mu_2, \sigma_2^2, \Lambda)$. The GMM estimator is defined as

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}}\left\{\mathbf{h}_n(\boldsymbol{\theta}, \mathbf{X}_n)'\mathbf{W}_n\,\mathbf{h}_n(\boldsymbol{\theta}, \mathbf{X}_n)\right\},$$

where $\mathbf{h}_n(\boldsymbol{\theta}, \mathbf{X}_n) = [\mathbf{h}_1\ \mathbf{h}_2\ \mathbf{h}_3\ \mathbf{h}_4\ \mathbf{h}_5\ \mathbf{h}_6]'$ is a $(5 + 2(q_1 + q_2))-$dimensional vector based on the summary statistics concerning the (unconditional) first- and second-order moments: mean, variance, autocovariance and the cross-covariance up to lag q_j, $(j = 1, 2)$. Letting

$$\mu'_{X_{j_1},X_{j_2}}(\tau) = \mathrm{E}[X_{j_1,t-\tau}\, X_{j_2,t}] = \gamma_{X_{j_1},X_{j_2}}(\tau) + \mu_{X_{j_1}}\mu_{X_{j_2}},$$

for $\tau = 0, 1, \ldots,$ $\gamma_{X_{j_i},X_{j_i}} = \gamma_{X_{j_i}}$ and $j_1, j_2 \in \{1, 2\}$, the elements of $\mathbf{h}$ are as follows: $\mathbf{h}_1$ and $\mathbf{h}_2$ are 2-dimensional vectors with components $h_{1k} = \frac{1}{n}\sum_{t=1}^{n}(X_{k,t} - \mu_{X_k})$ and $h_{2k} = \frac{1}{n-1}\sum_{t=1}^{n}(X_{k,t}^2 - \mu'_{X_k,X_k}(0))$, for $k = 1, 2$, respectively; $\mathbf{h}_3$ and $\mathbf{h}_4$ are q_1- and q_2-dimensional vectors with components $h_{3k} = \frac{1}{n}\sum_{t=1+k}^{n}(X_{1,t-k}X_{1,t} - \mu'_{X_1,X_1}(k))$, for $k = 1, \ldots, q_1$, and $h_{4k} = \frac{1}{n}\sum_{t=1+k}^{n}(X_{2,t-k}X_{2,t} - \mu'_{X_2,X_2}(k))$, for $k = 1, \ldots, q_2$, respectively; and finally $\mathbf{h}_5$ and $\mathbf{h}_6$ are $(q_2 + 1)-$ and q_1-dimensional vectors with components, respectively, $h_{5k} = \frac{1}{n}\sum_{t=1+k}^{n}(X_{1,t-k}X_{2,t} - \mu'_{X_1,X_2}(k))$, for $k = 0, \ldots, q_2$, and $h_{6k} = \frac{1}{n}\sum_{t=1+k}^{n}(X_{1,t}X_{2,t-k} - \mu'_{X_1,X_2}(k))$, for $k = 1, \ldots, q_1$. $\mathbf{W}_n$ is a symmetric and positive definite weight matrix which is obtained as the inverse of the covariance matrix of $\mathbf{h}_n(\boldsymbol{\theta}, \mathbf{X}_{j,t})$, that is $\mathbf{W}_n = (\mathrm{Cov}(\mathbf{h}_n(\boldsymbol{\theta}, \mathbf{X}_{j,t})))^{-1}$.

The GMM estimator thus defined is asymptotically consistent with the smallest attainable asymptotic variance, see theorems A.1 and A.2 in Appendix 2 and [12] for additional details.

Remark 4.1: Note that, in general is not possible to find an analytical solution for the minimization of the quadratic form $\mathbf{h}_n(\boldsymbol{\theta}, \mathbf{X}_n)'\mathbf{W}_n\mathbf{h}_n(\boldsymbol{\theta}, \mathbf{X}_n)$ and we have to resort to numerical procedures. In order to obtain an efficient GMM estimator we can reformulate the GMM criteria as

$$Q_n(\boldsymbol{\theta}, \mathbf{X}_n) = \mathbf{h}_n(\boldsymbol{\theta}, \mathbf{X}_n)'\ \mathbf{W}_n(\boldsymbol{\theta})\ \mathbf{h}_n(\boldsymbol{\theta}, \mathbf{X}_n), \tag{6}$$

where the weight matrix, which depends on the parameters, is obtained via plug-in or empirical estimation as the covariance matrix of $\mathbf{h}_n$ and minimize this quadratic form with respect to $\boldsymbol{\theta}$. This procedure is called the *continuously updated GMM* estimator. Details and alternative approaches for the estimation of the optimal weight matrix can be found in [10,28].

Quoreshi [31] also used a GMM approach to the estimation of the BINMA (q_1, q_2) based on prediction errors. Thus, the summary statistics as well as the quadratic forms to be minimized are different. Note that the weight matrix in [31] is fixed while in our approach it is updated between iterations.

4.2. Monte Carlo simulation results

To illustrate the estimation procedure and to analyse the small sample properties of the GMM estimators for the parameters we focus on the BINMA(1,1) model with BP and BNB innovation processes, respectively. Thus, 5000 independent replicates of time series of length $n = 200$, 500 and 1000 are generated from each model. The mean estimate and the standard errors of the estimates are obtained from the 5000 replications. The practical implementation of GMM explained in last section is adapted to each model, with $q_1 = 1$ and $q_2 = 1$.

The minimization in (6) is performed by the R function *optim*, which accomplished a general-purpose optimization based on Nelder-Mead, quasi-Newton and conjugate-gradient algorithms and includes an option for box-constrained optimization [32]. The initial estimates for the minimization of the quadratic form are obtained from the method of moments.

BINMA(1,1) *model with bivariate Poisson innovations*

The minimization in (6) is subject to $\beta_j \in]0,1[$, $\lambda_j > 0$ and $\phi \in]0, \min(\lambda_1, \lambda_2)[$ for $j = 1, 2$. The results are summarized in Table 1 by the (mean) estimates and corresponding standard errors. The results indicate that the bias and standard errors decrease as n increases as expected from the GMM estimators properties. Furthermore, even in small samples, the distributions of the estimators is fairly symmetric. Note that the estimates

Table 1. Sample mean and standard errors (in brackets) of the estimates for BINMA(1,1) models with Bivariate Poisson innovations.

$\boldsymbol{\theta} = (\beta_1, \lambda_1, \beta_2, \lambda_2, \phi)$	n	β_1	λ_1	β_2	λ_2	ϕ
	200	0.142	2.963	0.469	1.125	0.440
		(0.129)	(0.396)	(0.217)	(0.279)	(0.176)
$\boldsymbol{\theta} = (0.1, 3.0, 0.5, 1.0, 0.5)$	500	0.111	3.009	0.493	1.049	0.473
		(0.082)	(0.277)	(0.150)	(0.198)	(0.125)
	1000	0.104	3.006	0.501	1.017	0.491
		(0.059)	(0.202)	(0.113)	(0.152)	(0.093)
	200	0.135	2.966	0.671	2.577	0.934
		(0.109)	(0.429)	(0.256)	(0.727)	(0.317)
$\boldsymbol{\theta} = (0.1, 3.0, 0.9, 2.0, 1.0)$	500	0.109	2.999	0.781	2.264	0.976
		(0.070)	(0.298)	(0.188)	(0.471)	(0.215)
	1000	0.102	3.013	0.824	2.165	0.985
		(0.049)	(0.224)	(0.149)	(0.341)	(0.157)
	200	0.600	2.149	0.650	2.241	0.920
		(0.243)	(0.520)	(0.249)	(0.548)	(0.269)
$\boldsymbol{\theta} = (0.6, 2.0, 0.7, 2.0, 1.0)$	500	0.609	2.055	0.691	2.091	0.966
		(0.186)	(0.379)	(0.198)	(0.394)	(0.190)
	1000	0.606	2.033	0.707	2.035	0.981
		(0.152)	(0.310)	(0.164)	(0.307)	(0.142)
	200	0.594	2.073	0.662	2.123	0.492
		(0.240)	(0.445)	(0.243)	(0.458)	(0.235)
$\boldsymbol{\theta} = (0.6, 2.0, 0.7, 2.0, 0.5)$	500	0.611	2.019	0.696	2.045	0.492
		(0.177)	(0.305)	(0.185)	(0.312)	(0.156)
	1000	0.609	2.009	0.704	2.018	0.493
		(0.133)	(0.224)	(0.147)	(0.237)	(0.112)
	200	0.297	1.073	0.320	2.058	0.449
		(0.165)	(0.230)	(0.191)	(0.392)	(0.159)
$\boldsymbol{\theta} = (0.3, 1.0, 0.3, 2.0, 0.5)$	500	0.302	1.023	0.305	2.024	0.484
		(0.104)	(0.161)	(0.121)	(0.259)	(0.113)
	1000	0.302	1.005	0.304	2.007	0.496
		(0.075)	(0.119)	(0.085)	(0.181)	(0.082)
	144	0.285	0.209	0.685	0.173	0.065
		(0.209)	(0.056)	(0.224)	(0.076)	(0.031)
	200	0.265	0.209	0.709	0.164	0.069
$\boldsymbol{\theta} = (0.221, 0.181, 0.740, 0.109, 0.109)^{a}$		(0.173)	(0.047)	(0.193)	(0.067)	(0.028)
	500	0.239	0.206	0.755	0.147	0.078
		(0.097)	(0.030)	(0.129)	(0.038)	(0.018)
	1000	0.232	0.205	0.769	0.141	0.082
		(0.057)	(0.021)	(0.096)	(0.017)	(0.013)

[a]This set of parameters corresponds to the one obtained in the real data illustration and given in Table 4.

present large variability as indicated by the standard errors specially for small values of β and small sample sizes.

BINMA(1,1) *model with bivariate negative binomial innovations*

The minimization in (6) is subject to $\beta_j \in]0,1[$, $\lambda_j > 0$ and $\tau > 0$. The results are summarized in Table 2 by the (mean) estimates and corresponding standard errors. Once again, the sample bias and standard errors decrease as the sample size increases, indicating that the distribution of the estimators is consistent and symmetric. The results show that, in general, $\hat{\lambda}_1$ and $\hat{\tau}$ are underestimated. Also for this model, even in small samples, the distribution of the estimators is fairly symmetric. The estimates present a larger variability than expected, specially for small sample sizes.

Table 2. Sample mean and standard errors (in brackets) of the estimates for BINMA(1, 1) model with bivariate negative binomial innovations.

$\boldsymbol{\theta} = (\beta_1, \lambda_1, \beta_2, \lambda_2, \tau)$	n	β_1	λ_1	β_2	λ_2	τ
	200	0.116	2.935	0.501	0.997	0.478
		(0.085)	(0.308)	(0.180)	(0.152)	(0.110)
$\boldsymbol{\theta} = (0.1, 3.0, 0.5, 1.0, 0.5)$	500	0.103	2.978	0.495	1.001	0.486
		(0.052)	(0.203)	(0.115)	(0.099)	(0.071)
	1000	0.101	2.987	0.498	1.000	0.490
		(0.037)	(0.141)	(0.078)	(0.069)	(0.050)
	200	0.117	2.864	0.863	1.996	0.937
		(0.121)	(0.373)	(0.171)	(0.290)	(0.217)
$\boldsymbol{\theta} = (0.1, 3.0, 0.9, 2.0, 1.0)$	500	0.110	2.924	0.892	1.981	0.977
		(0.089)	(0.287)	(0.104)	(0.175)	(0.144)
	1000	0.108	2.950	0.901	1.983	0.988
		(0.066)	(0.224)	(0.080)	(0.130)	(0.100)
	200	0.620	1.940	0.685	1.991	0.915
		(0.163)	(0.256)	(0.195)	(0.295)	(0.199)
$\boldsymbol{\theta} = (0.6, 2.0, 0.7, 2.0, 1.0)$	500	0.618	1.958	0.705	1.985	0.962
		(0.104)	(0.168)	(0.142)	(0.201)	(0.130)
	1000	0.610	1.978	0.702	1.992	0.981
		(0.067)	(0.114)	(0.103)	(0.147)	(0.092)
	200	0.618	1.970	0.685	2.020	0.469
		(0.172)	(0.252)	(0.193)	(0.281)	(0.108)
$\boldsymbol{\theta} = (0.6, 2.0, 0.7, 2.0, 0.5)$	500	0.621	1.972	0.701	2.003	0.486
		(0.118)	(0.167)	(0.148)	(0.198)	(0.069)
	1000	0.610	1.985	0.700	2.003	0.491
		(0.076)	(0.114)	(0.111)	(0.146)	(0.050)
	200	0.321	0.990	0.320	1.970	0.508
		(0.148)	(0.147)	(0.147)	(0.263)	(0.133)
$\boldsymbol{\theta} = (0.3, 1.0, 0.3, 2.0, 0.5)$	500	0.311	0.996	0.306	1.991	0.519
		(0.078)	(0.085)	(0.084)	(0.164)	(0.079)
	1000	0.310	0.996	0.299	2.006	0.517
		(0.055)	(0.060)	(0.065)	(0.122)	(0.053)
	144	0.223	0.175	0.764	0.140	0.254
		(0.211)	(0.053)	(0.208)	(0.042)	(0.386)
	200	0.194	0.183	0.756	0.146	0.218
$\boldsymbol{\theta} = (0.137, 0.203, 0.687, 0.165, 0.228)$[a]		(0.172)	(0.043)	(0.197)	(0.037)	(0.309)
	500	0.152	0.196	0.727	0.157	0.207
		(0.090)	(0.027)	(0.150)	(0.025)	(0.212)
	1000	0.146	0.200	0.708	0.160	0.208
		(0.053)	(0.019)	(0.114)	(0.018)	(0.166)

[a]This set of parameters corresponds to the one obtained in the real data illustration and given in Table 4.

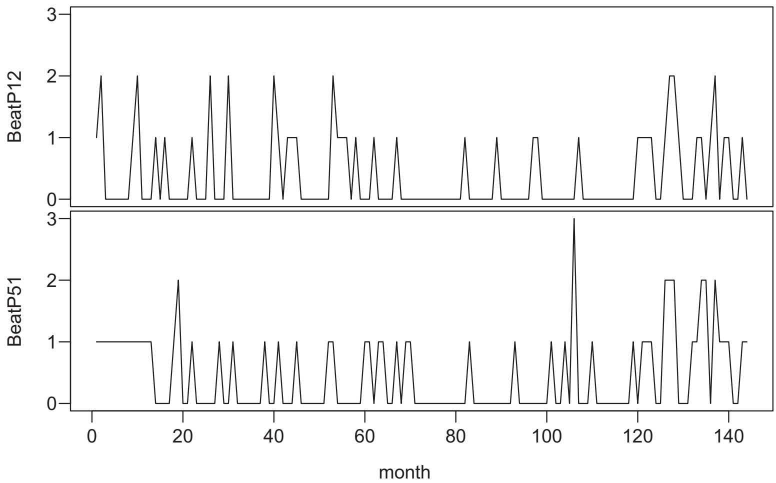

Figure 1. Time series of the monthly number of vagrancy offences, from 1991 to 2001, registered in Pittsburgh.

Table 3. Sample measures for the vagrancy dataset.

	Mean		Variance		ACF(1)		CCF(0)
	BeatP12	BeatP51	BeatP12	BeatP51	BeatP12	BeatP51	
Sample	0.347	0.431	0.354	0.387	0.215	0.253	0.255
BP innov.	0.354	0.379	0.354	0.379	0.181	0.425	0.346
BNB innov.	0.230	0.278	0.240	0.287	0.137	0.687	0.032

5. Real data illustration

The model is now used to fit a bivariate dataset consisting on the aggregated monthly number of vagrancy offences registered by two police car beats, called beats plus (number 12 and 51), in Pittsburgh (Pennsylvania, USA), from January of 1990 to December of 2001, in a total of $n = 144$ observations per series (see Figure 1). Vagrancy can be defined as the state of wandering from place to place without permanent home or regular employment. The dataset is available from the Forecasting Principles site http://www.forecastingprinciples.com/index.php/crimedata.

A preliminary analysis of the sample mean, variance and cross-correlation of the data, presented in Table 3, indicates that marginally the Poisson distribution might be suitable. Furthermore, the values of the sample autocorrelation function (ACF) in Figure 2, which are nearly zero after the first lag, suggest that a first-order model is appropriate to the dataset, while the sample cross-correlation indicates dependence between the two series (Figure 3). Therefore, we opt for a BINMA$(1, 1)$ process with Bivariate Poisson innovations to model these data. On the other hand, it is expected that vagrants stay a limited time in the system and that these individuals can come and go several times during their life times in the system.

The GMM estimates were obtained numerically as in (6), with starting value $\boldsymbol{\theta}^* = (\beta_1, \lambda_1, \beta_2, \lambda_2, \phi) = (0.4, 0.5, 0.5, 0.5, 0.2)$. The estimates and corresponding standard errors are given in Table 4.

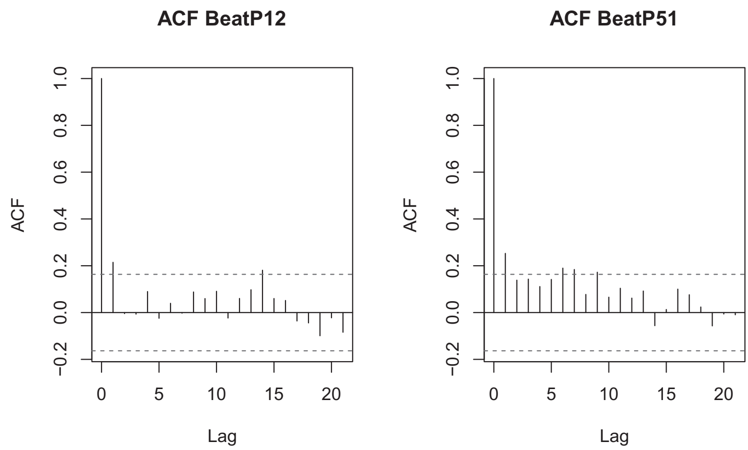

Figure 2. Sample autocorrelation for the vagrancy dataset.

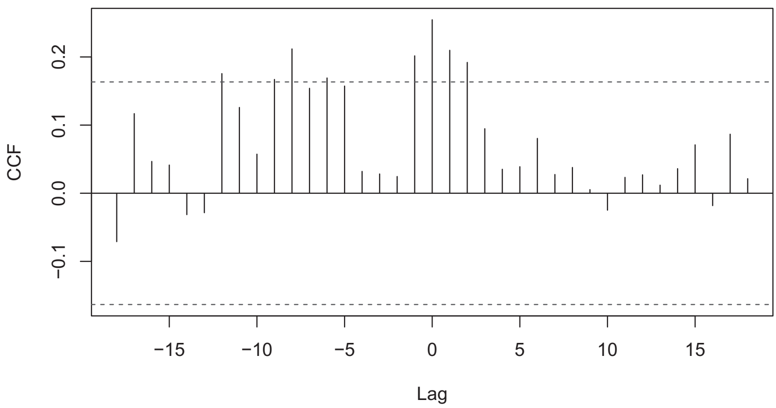

Figure 3. Sample cross-correlation for the vagrancy dataset.

Table 4. GMM estimates for the vagrancy dataset (standard errors in brackets).

BINMA (1, 1)	BeatP12		BeatP51		Cross-Corr.
BP innov.	$\hat{\beta}_1$	$\hat{\lambda}_1$	$\hat{\beta}_2$	$\hat{\lambda}_2$	$\hat{\phi}$
	0.221 (0.165)	0.181 (0.058)	0.740 (0.386)	0.109 (0.058)	0.109 (0.032)
BNB innov.	$\hat{\beta}_1$	$\hat{\lambda}_1$	$\hat{\beta}_2$	$\hat{\lambda}_2$	$\hat{\tau}$
	0.137 (0.156)	0.203 (0.050)	0.687 (0.414)	0.165 (0.059)	0.228 (0.566)

In order to assess the adequacy of the fitted model, we use several model diagnostic tools, namely validation based on residual analysis, parametric resampling, predictive distributions (PIT histogram) [15] and the dispersion test of [1].

The Pearson residuals (or standardized residuals) are defined as $R_{j,t} = \frac{X_{j,,t} - \mathrm{E}[X_{j,t}|\mathcal{F}_{j,t-1}]}{\mathrm{Var}[X_{j,t}|\mathcal{F}_{j,t-1}]^{1/2}}$, for $j = 1, 2$, where $\mathcal{F}_{j,t-1}$ is the σ-algebra generated by $\{X_{j,1}, \ldots, X_{j,t-1}\}$, and the conditional mean and variance are defined in (5). Note that in practice population parameters are replaced by their estimated counterparts. If the model is well-chosen then the Pearson

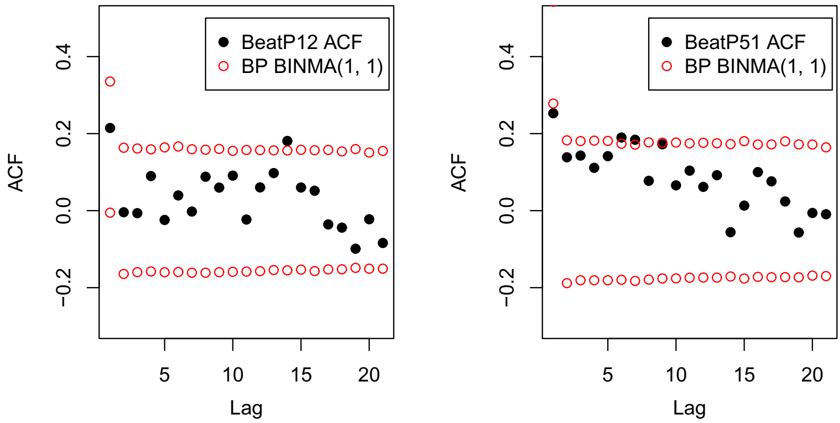

Figure 4. Acceptance envelope for the autocorrelation function for the BINMA(1, 1) model with BP innovations.

residuals should exhibit zero mean and unit variance and no (significant) serial correlation. In this case, the sample mean and variance for the residuals are $\bar{R}_1 = -0.023$, $\bar{R}_2 = -0.084$, $\hat{s}^2_{R_1} = 0.979$ and $\hat{s}^2_{R_2} = 1.267$. Additionally, the analysis of the sample autocorrelation and sample partial autocorrelation of the residuals, as well as the usual tests of randomness, do not reject the hypothesis of uncorrelated random variables for the residual series. For the Ljung-Box test, the values of the test statistic and the corresponding p-values are 5.667 and 0.842, respectively, for the BeatP12 residuals and 8.063 and 0.623, respectively, for the BeatP51 residuals.

To assess the adequacy of the model to represent specific features of interest of the data, in this case auto- and cross-correlation (ACF and CCF) we use parametric bootstrap: the fitted model is used to generate 5000 (bivariate) time series samples, all with the same number of observations as the original data set which are then used to construct empirical distributions for the ACF and CCF. Figures 4 and 5 represent the acceptance envelopes computed from the 2.5% and 97.5% quantiles of the empirical distribution for the ACF and CCF. It is clear that the model represents adequately both the autocorrelation and the cross-correlation.

Finally, to check the adequacy of the distributional assumptions we construct the PIT (probability integral transform) histogram, which can be defined (in the discrete context) as the conditional cumulative distribution function given the observed count. Figure 6 represents the (non-randomized) PIT histogram and the approximate $100(1-\alpha)\%$ confidence intervals ($\alpha = 0.05$) obtained from a standard χ^2 goodness-of-fit test of the null hypothesis that the bins of the histogram are drawn from a uniform distribution, as in [15]. As we can see, the PIT histogram is close to an uniform distribution, specially for the first series. Besides, the uniformity test does not reject the hypothesis of uniform distribution ($p-$values larger than .99 for each series). Additionally, the dispersion indexes are $\hat{I}_{\text{BeatP12}} = 1.020$ and $\hat{I}_{\text{BeatP51}} = 0.898$, respectively. Using the dispersion test of [1], the Poisson distribution is not refuted ($\hat{I}_{\text{BeatP12}} \in]0.752, 1.229[$ and $\hat{I}_{\text{BeatP51}} \in]0.718, 1.257[$).

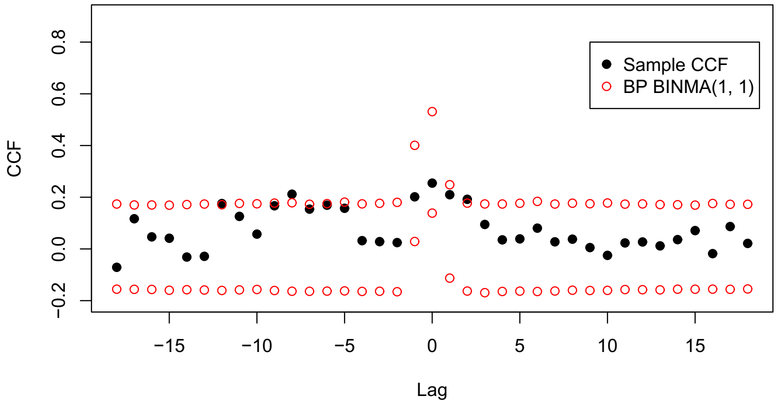

Figure 5. Acceptance envelope for the cross-correlation function for the vagrancy dataset for the BINMA(1, 1) model with BP innovations.

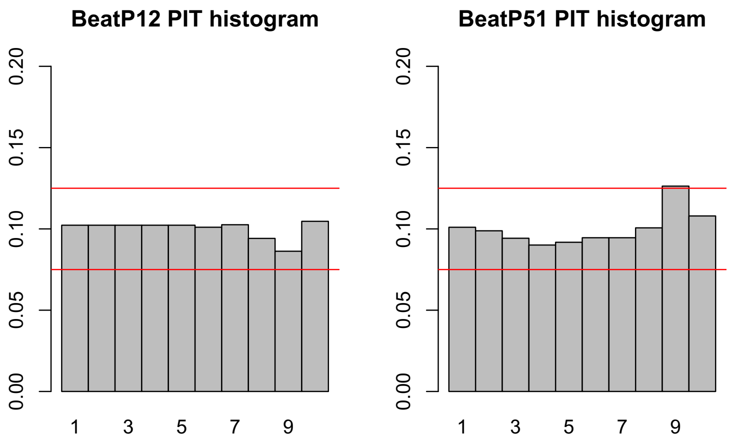

Figure 6. Non-randomized PIT histogram for the vagrancy dataset for the BINMA(1, 1) model with BP innovations.

Note that the standard errors of the estimates indicate that the β's parameters may be zero, a characteristic of the GMM estimates also observed in the simulation study, as can be seen for instance at the bottom part of Table 1, where 5000 replicates of BINMA(1, 1) model with BP innovations with 144 observations and parameter values given in Table 4 were used. However, the detailed model assessment suggests that the fitted BINMA(1, 1) model with BP innovations suitably describes the dependence structure. Furthermore, Table 3 indicates that the model approximates well the empirical values for the marginal means, variances and ACF(1)'s.

The residual analysis indicates that for the second series, BeatP51, the Pearson residual variance is larger than one and the PIT histogram shows a U-shape, indicating that the fitted model does not completely explain the slight under-dispersion present in this series. Although beyond of the scope of this work, this issue might be addressed using copulas for the specification of a joint distribution for the innovations with equidispersion in the first series and overdispersion in the second series, as in [17].

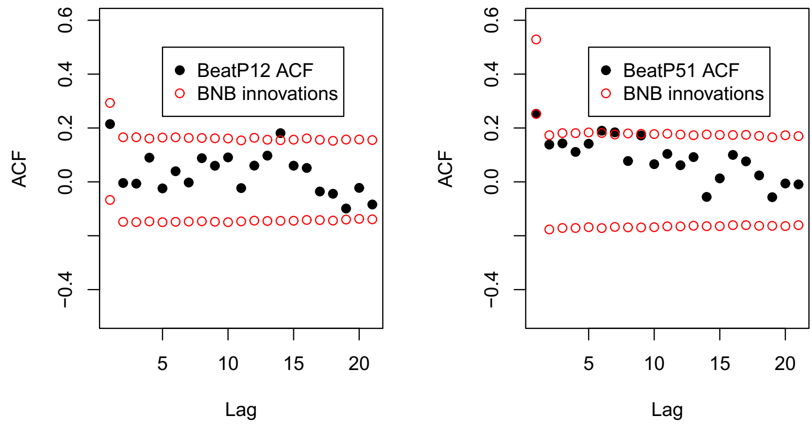

Figure 7. Acceptance envelope for the autocorrelation function for the vagrancy dataset for the BINMA(1,1) model with BNB innovation process.

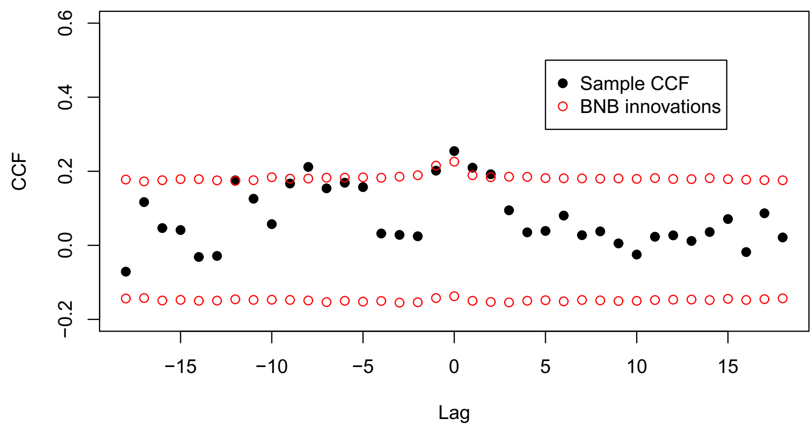

Figure 8. Acceptance envelope for the cross-correlation for the vagrancy dataset for the BINMA(1,1) model with BNB innovation process.

Just as an illustration, we have also fitted a BINMA(1,1) model with BNB innovation distribution. The GMM estimates were obtained numerically, with starting value $\boldsymbol{\theta}^* = (\beta_1, \lambda_1, \beta_2, \lambda_2, \tau) = (0.9, 0.5, 0.2, 1.0, 1.0)$. The estimates and corresponding standard errors are given in Table 4 and the acceptance envelopes computed from the 2.5% and 97.5% quantiles of the empirical distribution for the ACF and CCF are shown in Figures 7 and 8. In this case, as expected, it can be seen that the model does not completely capture both the autocorrelation and the cross-correlation.

6. Final remarks

Thinning-based models, usually denoted as INARMA, have become popular in the literature on univariate time series of counts. These models are, in fact, nonlinear models (the nonlinearity is induced by the random operation) and their extension to the multivariate context is not straightforward. In this work we considered a bivariate INMA, BINMA, model which extends INMA models previously proposed in the literature. This BINMA process is able to model time series negatively correlated. The Bivariate Poisson distribution is characterized not only by equi-dispersion, just as the univariate case, but also by a covariance that is bounded by the marginal means (and variances) hindering its application in presence of over-dispersion and strong covariance. A BINMA model with Bivariate Negative Binomial innovations is also presented. This model can account for overdispersion. Further study of bivariate INMA models will be reported elsewhere.

Acknowledgements

The authors thank to the referee for the helpful comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was partially supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020.

ORCID

Isabel Silva http://orcid.org/0000-0002-6307-3456
Maria Eduarda Silva http://orcid.org/0000-0003-2972-2050
Cristina Torres http://orcid.org/0000-0002-8644-2381

References

[1] B. Aleksandrov and C.H. Weiß, *Parameter estimation and diagnostic tests for INMA(1) processes*, TEST (2019). doi:10.1007/s11749-019-00653-7.
[2] M. Alosh, *The impact of missing data in a generalized integer-valued autoregression model for count data*, J. Biopharm. Stat. 19 (2009), pp. 1039–1054.
[3] A.A. Alzaid and M.A. Al-Osh, *First order integer-valued autoregressive INAR(1) process*, J. Time Ser. Anal. 8 (1987), pp. 261–275.
[4] M.A. Al-Osh and A.A. Alzaid, *Integer-valued moving average (INMA) process*, Statist. Papers 29 (1988), pp. 281–300.
[5] K. Brännäs and A. Hall, *Estimation in integer-valued moving average models*, Appl. Stoch. Models Bus. Ind. 17 (2001), pp. 277–291.
[6] K. Brännäs and J. Nordström, *Tourist accommodation effects of festivals*, Tourism Econ. 12 (2006), pp. 291–302.
[7] S. Cheon, S.H. Song, and B.C. Jung, *Tests for independence in a bivariate negative binomial model*, J. Korean Statist. Soc. 2 (2009), pp. 185–190.
[8] Y. Cui and R. Lund, *A new look at time series of counts*, Biometrika 96 (2009), pp. 781–792.
[9] L.P. Hansen, *Large sample properties of generalized method of moments estimators*, Econometrica 50 (1982), pp. 1029–1054.
[10] L.P. Hansen, J. Heaton, and A. Yaron, *Finite-sample properties of some alternative GMM estimators*, J. Bus. Econ. Stat. 14 (1996), pp. 262–280.
[11] M. Iwasaki and H. Tsubaki, *Bivariate negative binomial generalized linear models for environmental count data*, J. Appl. Stat. 33 (2006), pp. 909–923.
[12] J. Jesus and R.E. Chandler, *Estimating functions and the generalized method of moments*, Interface. Focus. 1 (2011), pp. 871–885.

[13] V. Jowaheer, N. Mamode Khan, and Y. Sunecher, *A non-stationary BINAR(1) process with negative binomial innovations for modeling the number of goals in the first and second half: the case of arsenal football club*, Commun. Stat. Case Stud. Data Anal. Appl. 2 (2016), pp. 21–33.
[14] V. Jowaheer, N. Mamode Khan, and Y. Sunecher, *A BINAR(1) time series model with cross-correlated COM-Poisson innovations*, Commun. Stat. Theory Methods 47 (2017), pp. 1133–1154.
[15] R.C. Jung, B.P.M. McCabe, and A.R. Tremayne, *Model validation and diagnostics*, in *Handbook of Discrete-Valued Time Series*, R.A. Davis, S.H. Holan, R. Lund, and N. Ravishanker, eds., Chapman & Hall/CRC Press, Boca Raton, 2015, pp. 189–218.
[16] R.C. Jung and A.R. Tremayne, *Convolution-closed models for count time series with applications*, J. Time Ser. Anal. 32 (2011), pp. 268–280.
[17] D. Karlis and X. Pedeli, *Flexible bivariate INAR(1) processes using copulas*, Commun. Stat. Theory Methods 42 (2013), pp. 723–740.
[18] S. Kocherlakota and K. Kocherlakota, *Bivariate Discrete Distributions*, Markel Dekker, New York, 1992.
[19] A. Latour, *The multivariate GINAR(p)process*, Adv. Appl. Probab. 29 (1997), pp. 228–248.
[20] A.W. Marshall and I. Olkin, *Multivariate distributions generated from mixtures of convolution and product families*, Lect. Notes Monogr. Ser. (1990), pp. 371–393.
[21] D. Ma, D. Wang, and J. Cheng, *Bidimensional discrete-time risk models based on bivariate claim count time series*, J. Inequal. Appl. 2015 (2015), pp. 1–22.
[22] N. Mamode Khan, Y. Sunecher, and V. Jowaheer, *Analyzing the full BINMA time series process using a robust GQL approach*, J. Time Ser. Econ. 9 (2016).
[23] N. Mamode Khan, Y. Sunecher, and V. Jowaheer, *Inferential methods for an unconstrained non-stationary BINMA time series process with Poisson innovations*, J. Stat. Theory Pract. 11 (2017), pp. 76–106.
[24] N. Mamode Khan, Y. Sunecher, and V. Jowaheer, *Modelling a non-stationary BINAR(1) Poisson process*, J. Stat. Comput. Simul. 86 (2016), pp. 3106–3126.
[25] B.P.M. McCabe and G.M. Martin, *Bayesian predictions of low count time series*, Int. J. Forecast 21 (2005), pp. 315–330.
[26] E. McKenzie, *Autoregressive moving-average processes with negative-binomial and geometric marginal distributions*, Adv. Appl. Probab. 18 (1986), pp. 679–705.
[27] E. McKenzie, *Some ARMA models for dependent sequences of Poisson counts*, Adv. Appl. Probab. 20 (1988), pp. 822–835.
[28] H.B. Nielsen, *Generalized method of moments estimation* (2007) Available at http://www.econ.ku.dk/metrics/Econometrics2_07_I/LectureNotes/gmm.pdf (accessed 21 December 2015).
[29] X. Pedeli and D. Karlis, *A bivariate INAR(1) process with application*, Stat. Model. 11 (2011), pp. 325–349.
[30] P. Popović, *Random coefficient bivariate INAR(1) process*, Facta Universitatis, Ser. Math. Inform. 30 (2015), pp. 263–280.
[31] A.M.M.S. Quoreshi, *Bivariate time series modeling of financial count data*, Commun. Stat. Theory Methods 35 (2006), pp. 1343–1358.
[32] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008, ISBN 3-900051-07-0. Available at http://www.R-project.org.
[33] M.M. Ristić, A.S. Nasti, K. Jayakumar, and H.S. Bakouch, *A bivariate INAR(1) time series model with geometric marginals*, Appl. Math. Lett. 25 (2012), pp. 481–485.
[34] M. Scotto, C.H. Weiß, M.E. Silva, and I. Pereira, *Bivariate binomial autoregressive models*, J. Multivariate Anal. 125 (2014), pp. 233–251.
[35] M.E. Silva and V.L. Oliveira, *Difference equations for the higher-order moments and cumulants of the INAR(1) model*, J. Time Ser. Anal. 25 (2004), pp. 317–333.
[36] M.E. Silva, I. Silva, and C. Torres, *Modelling overdispersion with integer-valued moving average processes*, in *Stochastic Models, Statistics and their Application – SMSA 2019*, A. Steland, E. Rafajłowicz and O. Okhrin, eds., Springer Proceedings in Mathematics & Statistics, Springer, Cham. 2019, pp. 291–303.

[37] Y. Sunecher, N. Mamode Khan, and V. Jowaheer, *Estimating the parameters of a BINMA Poisson model for a non-stationary bivariate time series*, Commun. Stat. Simul. Comput. 46 (2017), pp. 6803–6827.
[38] Y. Sunecher, N. Mamode Khan, and V. Jowaheer, *A GQL estimation approach for analysing non-stationary over-dispersed BINAR(1) time series*, J. Stat. Comput. Simul. 87 (2017), pp. 1911–1924.
[39] F.W. Steutel and K. Van Harn, *Discrete analogues of self-decomposability and stability*, Ann. Probab. 7 (1979), pp. 893–899.
[40] P. Thyregod, J. Carstensen, H. Madsen, and K. Arnbjerg-Nielsen, *Integer valued autoregressive models for tipping bucket rainfall measurements*, Environmetrics 10 (1999), pp. 395–411.
[41] C.H. Weiß, *Serial dependence and regression of Poisson INARMA models*, J. Statist. Plan. Inference 138 (2008), pp. 2975–2990.
[42] C.H. Weiß, *An Introduction to Discrete-Valued Time Series*, Chichester, John Wiley & Sons Inc., 2018.
[43] C.H. Weiß, M. Feld, N. Mamode Khan, and Y. Sunecher, *INARMA modeling of count time series*, Stats 2(2) (2019), pp. 284–320.
[44] K. Yu, H. Zou and D. Shi, *Integer-valued moving average models with structural changes*, Math. probl. eng. 2014, Article ID 231592, 5 pages (2014). doi:10.1155/2014/231592.
[45] K. Yu and H. Zou, *The combined Poisson INMA(q) models for time series of counts*, J. Appl. Math. 2015, Article ID 457842, 7 pages (2015). doi:10.1155/2015/457842.
[46] X. Yu, M. Baron and P.K. Choudhary, *Change-point detection in binomial thinning processes, with applications in epidemiology*, Seq. Anal. 32 (2013), pp. 350–367.
[47] H. Zou and K. Yu, *First order threshold integer-valued moving average processes*, Dynam. Cont. Dis. Ser. B 21 (2014), pp. 197–205.

Appendices

Appendix 1. Cross-covariance function of the BINMA(q_1, q_2) process

The cross-covariance function, $\gamma_{X_1,X_2}(k)$, for $k = 0, \ldots, q_1$, of the BINMA(q_1, q_2) process defined in (3), can be written as,

$$\gamma_{X_1,X_2}(k) = \text{Cov}(X_{1,t}, X_{2,t-k}) = \sum_{i=0}^{q_1-k} \text{Cov}\left(\beta_{1,k+i} \circ \varepsilon_{1,t-k-i}, \beta_{2,i} \circ \varepsilon_{2,t-k-i}\right). \tag{A1}$$

By the independence of the random variables $\boldsymbol{\varepsilon}_t$, which follows from the definition of the model (3), it is possible to write

$$\gamma_{X_1,X_2}(k) = \sum_{i=0}^{q_1-k} \text{Cov}\left(\beta_{1,k+i} \circ \varepsilon_{1,t}, \beta_{2,i} \circ \varepsilon_{2,t}\right),$$

since all the other addends in (A1) are zero. Conditioning,

$$\begin{aligned}
\text{Cov}\left(\beta_{1,k+i} \circ \varepsilon_{1,t}, \beta_{2,i} \circ \varepsilon_{2,t}\right) &= \text{E}\left[\text{Cov}\left(\beta_{1,k+i} \circ \varepsilon_{1,t}, \beta_{2,i} \circ \varepsilon_{2,t} | \boldsymbol{\varepsilon}_t\right)\right] + \\
&\quad + \text{Cov}\left[\text{E}\left(\beta_{1,k+i} \circ \varepsilon_{1,t} | \varepsilon_{1,t}\right), \text{E}\left(\beta_{2,i} \circ \varepsilon_{2,t} | \varepsilon_{2,t}\right)\right] \\
&= 0 + \text{Cov}\left(\beta_{1,k+i}\, \varepsilon_{1,t}\, \beta_{2,i}\, \varepsilon_{2,t}\right) \\
&= \beta_{1,k+i}\, \beta_{2,i}\, \text{Cov}\left(\varepsilon_{1,t}, \varepsilon_{2,t}\right) \\
&= \beta_{1,k+i}\, \beta_{2,i}\, \Lambda.
\end{aligned}$$

Thus,

$$\gamma_{X_1,X_2}(k) = \sum_{i=0}^{q_1-k} \beta_{1,k+i}\,\beta_{2,i}\,\Lambda = \Lambda\left(\beta_{1,k} + \sum_{i=1}^{q_1-k} \beta_{1,k+i}\,\beta_{2,i}\right).$$

For $k \geq q_1 + 1$, it follows that $\mathrm{Cov}(X_{1,t}, X_{2,t-k}) = 0$, since all addends in (A1) are zero.

The proofs for $\gamma_{X_2,X_1}(k)$ are analogous.

Appendix 2. Generalized method of moment estimation

Suppose we have an observed sample $\mathbf{X}_n = \{X_t : t = 1, \ldots, n\}$ from which we want to estimate an unknown $q \times 1$ parameter vector $\boldsymbol{\theta}$ with true value $\boldsymbol{\theta}_0$ and consider a vector $\mathbf{T}_n = \mathbf{T}_n(\mathbf{X}_n)$ of $k \geq q$ summary statistics with expectation $\boldsymbol{\alpha}(\boldsymbol{\theta}) = \mathrm{E}[\mathbf{T}_n]$ (where $\boldsymbol{\alpha}(\boldsymbol{\theta})$ are the theoretical counterparts) under the model. The so called *moment condition* is defined by

$$\mathrm{E}[\mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n)] = \mathbf{0}, \tag{A2}$$

where $\mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n)$ is a continuous $k \times 1$ vector function of $\boldsymbol{\theta}$ given by $\mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n) = \mathbf{T}_n - \boldsymbol{\alpha}(\boldsymbol{\theta})$, and $\mathrm{E}[\mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n)]$ exists and is finite for all t and $\boldsymbol{\theta}$. In practice, equation (A2) is replaced by its sample analogous $\frac{1}{n}\sum_{t=1}^{n} \mathbf{h}_n(\boldsymbol{\theta}; X_t) = \mathbf{0}$, and an estimator $\hat{\boldsymbol{\theta}}$ can be obtained as the solution of the last equation.

Note that when $k = q$, we obtain the Method of Moments (MM) estimator and we say that $\boldsymbol{\theta}$ is just-identified. The *Generalized Method of Moments* estimator is obtained when $k > q$ and then we say that $\boldsymbol{\theta}$ is over-identified. In this case, since we have more equations than parameters in the GMM estimation we cannot guarantee a unique solution to the equation in (A2), for this reason the estimation of the parameters can be done by minimizing the distance from $\mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n)$ to zero. This distance could be measured by the quadratic form

$$Q_n^*(\boldsymbol{\theta}; \mathbf{X}_n) = \mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n)'\,\mathbf{W}_n\,\mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n),$$

where $[\cdot]'$ denotes transpose and $\mathbf{W}_n$ is a $k \times k$ is any symmetric and positive definite weight matrix that may depend on the data but that converges in probability to a positive definite matrix $\mathbf{W}$. Therefore, the GMM estimator of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \left\{\mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n)'\,\mathbf{W}_n\,\mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n)\right\}.$$

The consistency and asymptotic distribution of the GMM estimator is given in the following theorem (see [12] for additional details).

Theorem A.1: *Assume the existence of a sequence* $(\boldsymbol{\eta}_n)$ *such that* $\tilde{\mathbf{h}}_n(\boldsymbol{\theta}; \mathbf{X}_n) = \boldsymbol{\eta}_n\,\mathbf{h}_n(\boldsymbol{\theta}; \mathbf{X}_n)$ *has a limiting covariance matrix* $\mathbf{S}$*, and suppose that there exists a sequence* $(\boldsymbol{\delta}_n)$ *of invertible* $q \times q$ *matrices, that do not depend on* $\boldsymbol{\theta}$*, such that:*

(1) *$\lim_{n\to\infty}\boldsymbol{\delta}_n = \mathbf{0}$, the zero matrix;*
(2) *Within a neighbourhood of $\boldsymbol{\theta}_0$, for all $r \in \{1, \ldots, k\}$, the matrices $\tilde{\mathbf{H}}_n(\boldsymbol{\theta})\boldsymbol{\delta}_n = \frac{\partial \tilde{\mathbf{h}}_n(\boldsymbol{\theta}; \mathbf{X}_n)}{\partial \boldsymbol{\theta}}\boldsymbol{\delta}_n$ and $\boldsymbol{\Omega}_n^{(r)}(\boldsymbol{\theta})\,\boldsymbol{\delta}_n = \frac{\partial^2 \tilde{h}_{(r)}}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\boldsymbol{\delta}_n$ (where $\tilde{h}_{(r)}$ is the rth element of $\tilde{\mathbf{h}}_n(\boldsymbol{\theta}; \mathbf{X}_n)$) converge to limiting matrices $\mathbf{H}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}^{(r)}(\boldsymbol{\theta})$, respectively, all elements of which are continuous functions of $\boldsymbol{\theta}$ and where $\mathbf{H}(\boldsymbol{\theta})$ is invertible.*

Let $\hat{\boldsymbol{\theta}}_n$ be the value of $\boldsymbol{\theta}$ that minimize

$$Q_n(\boldsymbol{\theta}; \mathbf{X}_n) = \tilde{\mathbf{h}}_n(\boldsymbol{\theta}; \mathbf{X}_n)'\,\mathbf{W}_n\,\tilde{\mathbf{h}}_n(\boldsymbol{\theta}; \mathbf{X}_n). \tag{A3}$$

Then, the expected value of $\hat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}_0$ and the covariance matrix of $\boldsymbol{\delta}_n^{-1}\hat{\boldsymbol{\theta}}_n$ converges to $\mathbf{C} = (\mathbf{M}(\boldsymbol{\theta}_0))^{-1}\tilde{\boldsymbol{\Sigma}}(\mathbf{M}'(\boldsymbol{\theta}_0))^{-1}$, where $\tilde{\boldsymbol{\Sigma}} = \mathbf{H}'(\boldsymbol{\theta}_0)\mathbf{W}_n\,\mathbf{S}\,\mathbf{W}_n'\,\mathbf{H}(\boldsymbol{\theta}_0)$ and $\mathbf{M}(\boldsymbol{\theta}) = \mathbf{H}'(\boldsymbol{\theta})\mathbf{W}\,\mathbf{H}(\boldsymbol{\theta})$.

The variance of the estimator depends on the weight matrix, $\mathbf{W}_n$. In order to obtain the estimator with the smallest possible asymptotic variance, we can use the following theorem.

Theorem A.2: *Under the conditions set in Theorem* A.1, *the smallest attainable limiting covariance matrix of* $\boldsymbol{\delta}_n^{-1}\hat{\boldsymbol{\theta}}_n$ *is* $C_{min} = (\mathbf{H}'(\boldsymbol{\theta}_0)\mathbf{S}^{-1}\,\mathbf{H}(\boldsymbol{\theta}_0))^{-1}$ *and is achieved by setting* $\mathbf{W}_n = \mathbf{S}^{-1}$ *in Equation* (A3).