A bivariate INAR(1) process with application

Xanthi Pedeli and Dimitris Karlis*
Department of Statistics
Athens University of Economics and Business

Abstract

The study of time series models for count data has become a topic of special interest during the last years. However, while research on univariate time series for counts now flourish, the literature on multivariate time series models for count data is notably more limited. In the present paper, a bivariate integer-valued autoregressive process of order 1 (BINAR(1)) is introduced. Emphasis is placed on model with bivariate Poisson and bivariate negative binomial innovations. We discuss properties of the BINAR(1) model and propose the method of conditional maximum likelihood for the estimation of its unknown parameters. Issues of diagnostics and forecasting are considered and predictions are produced by means of the conditional forecast distribution. Estimation uncertainty is accommodated by taking advantage of the asymptotic normality of maximum likelihood estimators and constructing appropriate confidence intervals for the h-step-ahead conditional probability mass function. The proposed model is applied to a bivariate data series concerning daytime and nighttime road accidents in the Netherlands.
Keywords: BINAR; count data; Poisson; negative binomial; bivariate time series.

1 Introduction

Multivariate count data occur in several different disciplines like epidemiology, marketing, criminology and engineering just to name a few. In many cases the data are observed across time leading to multivariate time series data as, for example, when one studies the purchases of different products across time, or the occurrence of different diseases across time.

*Corresponding author

In the literature there are several models to fit univariate count time series models (see Davis et al., 1999). A commonly used class of such models consists of the so-called integer autoregressive time series models, introduced by McKenzie (1985) and Al-Osh and Alzaid (1987). The interested reader is referred to McKenzie (2003) and Jung and Tremayne (2006) for a brief but detailed review of such models. The literature on multivariate time series models for count data is less developed. Some interesting attempts have been made during the last decade but most of them do not arise in the context of INAR processes. Among the models that have been built in the aforementioned setting are those of Latour (1997); Brännäs and Nordström (2000); Heinen and Rengifo (2007); Silva et al. (2008) and Quoreshi (2006).

The aim of this paper is to introduce and examine in detail a bivariate integer-valued autoregressive model of order 1 (BINAR(1)). To motivate the model consider the case of road traffic accidents. Accident analysis assumes that even if the behavior of crashes differs between day and night, both types of accidents share some common hazards. Weather conditions, the road's quality and characteristics and "human error", i.e. fallible perception, attention and/or memory, count between the factors that introduce correlation. On the other hand, serial correlation between successive daily crash counts, i.e. autocorrelation, is reported as an important challenge for all accident models (Brijs et al., 2008). Thus, appropriate time series models are needed to handle the presence of correlation both between and within the series of daytime and nighttime crash counts.

The remaining of the paper is structured as follows. A general specification of the BINAR(1) process and alternative methods for the estimation of its unknown parameters are given in section 2. In sections 3 we concentrate on the special cases of bivariate Poisson and bivariate negative binomial innovations respectively. In section 4 we give a specification of the model residuals as a diagnostic tool while issues of forecasting are discussed in section 5. An application to real data concerning daytime and nighttime road accidents follows in section 6. Some concluding remarks are presented in section 7.

2 The BINAR(1) Process

2.1 Model

Let $\mathbf{X}$ and $\mathbf{R}$ be non-negative integer-valued random 2-vectors. Let $\mathbf{A}$ be a 2×2 diagonal matrix with independent elements $\{\alpha_{jj}\}_{j=1,2}$. The bivariate integer-valued autoregressive process of order 1 can be defined as

$$\mathbf{X}_t = \mathbf{A} \circ \mathbf{X}_{t-1} + \mathbf{R}_t = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} R_{1t} \\ R_{2t} \end{bmatrix}, \quad t \in \mathbb{Z} \tag{2.1}$$

where "$\circ$"is the binomial thinning operator defined as $\alpha \circ X = \sum_{i=1}^{X} Y_i = Y$, where $\{Y_i\}_{i=1}^{X}$ is a sequence of *iid* Bernoulli random variables such that $P(Y_i = 1) = \alpha = 1 - P(Y_i = 0)$ and $\alpha \in [0, 1]$ (Steutel and van Harn, 1979). In the bivariate case, the $\mathbf{A}\circ$ operation is a matricial operation which acts as the usual matrix multiplication keeping in the same time the properties of the binomial thinning operation. One can see that with the above definition the jth element, $j = 1, 2$ is given by $X_{jt} = \alpha_j \circ X_{j,t-1} + R_{jt}$. The elements $\mathbf{R}_t$ which entered the system in the interval $(t - 1, t]$ are usually called as innovations.

Assuming independence between and within the thinning operations and $\{R_{jt}\}$ an *iid* sequence with mean λ_j and variance $\sigma_j^2 = \upsilon_j \lambda_j$, $\upsilon_j > 0, j = 1, 2$, the unconditional first and second order moments based on second order stationarity conditions are:

$$E(X_{jt}) = \mu_{X_j} = \frac{\lambda_j}{1 - \alpha_j} \tag{2.2}$$

$$Var(X_{jt}) = \sigma_{X_j}^2 = \frac{(\alpha_j + \upsilon_j)\lambda_j}{1 - \alpha_j^2} \tag{2.3}$$

$$\text{Cov}\,(X_{jt}, X_{j,t+h}) = \gamma_{X_j}(h) = \alpha_j^h \sigma_{X_j}^2\ ; \quad h = 1, 2, \ldots \tag{2.4}$$

$$\text{Corr}\,(X_{jt}, X_{j,t+h}) = \rho_{X_j}(h) = \alpha_j^h\ ; \quad h = 1, 2, \ldots \tag{2.5}$$

Obviously, the mean, variance and autocovariance functions can take only positive values, since λ_j,σ_j^2 and α_j are all positive. Depending on whether $\upsilon_j > 1$, $\upsilon_j \in (0, 1)$, or $\upsilon_j = 1$, the variance may be larger than the mean (overdispersion), smaller than the mean (underdispersion), or equal to the mean (equidispersion) respectively.

Dependence between the two series that comprise the BINAR(1) process is introduced by allowing for dependence between R_{1t} and R_{2t} while retaining all the previous assumptions fixed. Whatever the underlying joint distribution of $\{R_{1t}, R_{2t}\}$ is, it can be shown that the covariance between the innovations of the two series at time t, totally determines the covariance between the current value of the one process and the innovations of the other process at the same point in time t and vice versa (see Appendix):

$$\mathrm{Cov}(X_{1t}, R_{2t}) = \mathrm{Cov}(R_{1t}, R_{2t}) \tag{2.6}$$

As expected, the covariance between the sequences $\{X_{1t}\}$ and $\{X_{2t}\}$ at time t is also affected by the corresponding "survival"parts of the two processes. More specifically it can be shown that,

$$\mathrm{Cov}(X_{1,t+h}, X_{2t}) = \frac{\alpha_1^h}{(1-\alpha_1\alpha_2)}\mathrm{Cov}(R_{1t}, R_{2t}) \ ; \quad h = 0, 1, \ldots \text{ and} \tag{2.7}$$

$$\mathrm{Corr}(X_{1,t+h}, X_{2t}) = \frac{\alpha_1^h\sqrt{(1-\alpha_1^2)(1-\alpha_2^2)}}{(1-\alpha_1\alpha_2)\sqrt{(\alpha_1+\upsilon_1)(\alpha_2+\upsilon_2)\lambda_1\lambda_2}}\mathrm{Cov}(R_{1t}, R_{2t}) \ ; \quad h = 0, 1, \ldots \tag{2.8}$$

Covariances and correlations between X_{1t} and $X_{2,t+h}$, $h = 0, 1, \ldots$, can be defined analogously.

Note that (2.7) presumes that $\{\mathbf{X}_t\}$ is a strictly stationary process, i.e. that the joint distribution of $\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}$ is the same as that of $\begin{pmatrix} X_{1,t+h} \\ X_{2,t+h} \end{pmatrix}$, for all h. Using the analytical representations

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} R_{1t} \\ R_{2t} \end{bmatrix} \tag{2.9}$$

and

$$\begin{pmatrix} X_{1,t+h} \\ X_{2,t+h} \end{pmatrix} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} X_{1,t+h-1} \\ X_{2,t+h-1} \end{bmatrix} + \begin{bmatrix} R_{1,t+h} \\ R_{2,t+h} \end{bmatrix} \tag{2.10}$$

it is easy to see that strict stationarity does indeed hold for $\{\mathbf{X}_t\}$ since the variables involved in the right-hand sides of (2.9) and (2.10) have identical distributions (see also Latour, 1997).

2.2 Estimation

As already noted, the structure of the BINAR(1) model implies that the two innovation series $\{R_{1t}, R_{2t}\}$ follow jointly a bivariate distribution. Let $G_{R_1,R_2}(s_1, s_2)$ be the joint probability generating function (*jpgf*) of $\{R_{1t}, R_{2t}\}$. Then, the *jpgf* of $\mathbf{X}_t = \{X_{1t}, X_{2t}\}$ is given by

$$\begin{aligned} G_{\mathbf{X}_t}(\mathbf{s}) = G_{X_{1t},X_{2t}}(s_1, s_2) &= G_{X_{1,0}}(1-\alpha_1^t+\alpha_1^t s_1)G_{X_{2,0}}(1-\alpha_2^t+\alpha_2^t s_2) \\ &\times \prod_{i=0}^{t-1} G_{R_1,R_2}((1-\alpha_1^i+\alpha_1^i s_1), (1-\alpha_2^i+\alpha_2^i s_2)) \end{aligned} \tag{2.11}$$

which reduces to

$$G_{\mathbf{X}_t}(\mathbf{s}) = G_{X_{1t},X_{2t}}(s_1,s_2) = \prod_{i=0}^{\infty} G_{R_1,R_2}((1-\alpha_1^i+\alpha_1^i s_1),(1-\alpha_2^i+\alpha_2^i s_2)) \quad (2.12)$$

The moment generating function $M_{\mathbf{X}_t}(\mathbf{s}) = G_{\mathbf{X}_t}(e^{\mathbf{s}})$ can then be used to obtain appropriate sample moments for the estimation of the unknown model parameters. However, when the definition of a full density function is feasible, maximum-likelihood (ML) estimation is generally preferable. In the remaining of the present section we describe a general setting for ML estimation of the unknown parameters involved in the conditional mean and covariance functions.

The conditional density for the BINAR(1) model can be expressed as the convolution of two binomials, namely

$$f_1(x_1) = \binom{X_{1,t-1}}{x_1} \alpha_1^{x_1}(1-\alpha_1)^{X_{1,t-1}-x_1} \quad (2.13)$$

$$f_2(x_2) = \binom{X_{2,t-1}}{x_2} \alpha_2^{x_2}(1-\alpha_2)^{X_{2,t-1}-x_2}, \quad (2.14)$$

and a bivariate distribution of the form $f_3(k,s) = P(R_{1t}=k, R_{2t}=s)$. Thus the conditional density becomes

$$f(\mathbf{x}_t|\mathbf{x}_{t-1},\boldsymbol{\theta}) = \sum_k \sum_s f_1(x_{1t}-k) f_2(x_{2t}-s) f_3(k,s) \quad (2.15)$$

where $\boldsymbol{\theta}$ is the vector of unknown parameters.
The conditional likelihood function is then given by

$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{t=1}^{T} f(\mathbf{x}_t|\mathbf{x}_{t-1},\boldsymbol{\theta}) \quad (2.16)$$

for some initial value $\mathbf{x}_0$ and hence maximization provides with the ML estimates. Numerical maximization is straightforward with standard statistical packages.

3 Parametric Cases

In this section we discuss two specific BINAR(1) models. The first one comes from the assumption that the innovations of the two series follow jointly a bivariate Poisson distribution. The second model assumes a bivariate negative binomial distribution for the two innovation processes. The two representations can be viewed as appropriate tools for modeling equidispersed and overdispersed bivariate time series respectively. Some additional specifications for negative correlation time series data are also briefly considered.

3.1 The Poisson BINAR(1) Process

3.1.1 Model

Let assume that the joint probability mass function (*jpmf*) of the two innovation processes $\{R_{1t}, R_{2t}\}$ is a bivariate Poisson distribution given by

$$P(R_{1t} = x, R_{2t} = y) = \\ e^{-(\lambda_1+\lambda_2-\phi)} \frac{(\lambda_1 - \phi)^x}{x!} \frac{(\lambda_2 - \phi)^y}{y!} \sum_{i=0}^{s} \binom{x}{i} \binom{y}{i} i! \left(\frac{\phi}{(\lambda_1 - \phi)(\lambda_2 - \phi)} \right)^i \quad (3.1)$$

where $s = min(x, y)$, $\lambda_1, \lambda_2 > 0$ and $\phi \in [0, min(\lambda_1, \lambda_2))$. We will denote this distribution as $BP(\lambda_1, \lambda_2, \phi)$. The bivariate Poisson distribution defined in (3.1) allows for dependence between the two random variables. Marginally each random variable follows a Poisson distribution with parameters λ_1 and λ_2 respectively. Parameter ϕ is the covariance between the two random variables. If $\phi = 0$ then the two variables are independent and the bivariate Poisson distribution reduces to the product of two independent Poisson distributions. For a comprehensive treatment of the bivariate Poisson distribution and its multivariate extensions the reader can refer to the books of Kocherlakota and Kocherlakota (1992) and Johnson et al. (1997).

The above assumption leads to the equidispersion case, i.e. $\upsilon_j = 1$, or equivalently assume that R_{jt} are *iid* Poisson sequences with $\sigma_j^2 = \lambda_j$, $j = 1, 2$. Obviously, in this case the covariance function (2.7) remains unaffected while the correlation function (2.8) is simplified due to the simplification of the variances of the two processes. Hence, the Poisson BINAR(1) model is characterized by the vector of expectations $\boldsymbol{\mu}_{\mathbf{X}_t} = E(\mathbf{X}_t)$ with elements

$$\mu_{X_{jt}} = \frac{\lambda_j}{1 - \alpha_j} \; ; \quad j = 1, 2 \quad (3.2)$$

the variance-covariance matrix $\boldsymbol{\gamma}_{\mathbf{X}_t}(h)$ with diagonal elements

$$\text{Cov}(X_{j,t+h}, X_{jt}) = \frac{\alpha_j^h \lambda_j}{1 - \alpha_j} \; ; \quad j = 1, 2, \quad h = 0, 1, \ldots \quad (3.3)$$

and off-diagonal elements

$$\mathrm{Cov}(X_{j,t+h}, X_{it}) = \frac{\alpha_j^h \phi}{1-\alpha_1\alpha_2}\ ; \quad j \neq i, \quad h = 0, 1, \dots \tag{3.4}$$

and the correlation matrix $\boldsymbol{\rho}_{\mathbf{X}_t}(h)$ with diagonal and off-diagonal elements equal to

$$\mathrm{Corr}(X_{j,t+h}, X_{jt}) = \alpha_j^h\ ; \quad j = 1, 2, \quad h = 0, 1, \dots \tag{3.5}$$

and

$$\mathrm{Corr}(X_{j,t+h}, X_{it}) = \frac{\alpha_j^h \sqrt{(1-\alpha_1)(1-\alpha_2)}\phi}{(1-\alpha_1\alpha_2)\sqrt{\lambda_1\lambda_2}}\ ; \quad j \neq i, \quad h = 0, 1, \dots \tag{3.6}$$

respectively. Note also that conditionally on the previous observations $\mathbf{X}_{t-1} = \{X_{1,t-1}, X_{2,t-1}\}$, the vector of conditional means $\boldsymbol{\mu}_{\mathbf{X}_{t|t-1}} = E(\mathbf{X}_{t|t-1})$ has elements

$$\mu_{X_{jt|t-1}} = \alpha_j X_{j,t-1} + \lambda_j, \quad j = 1, 2.$$

For $h = 0$, the conditional variance-covariance matrix $\boldsymbol{\gamma}_{\mathbf{X}_{t|t-1}}(h)$ has diagonal and off-diagonal elements equal to

$$\mathrm{Cov}(X_{j,t+h}, X_{jt}|X_{j,t-1}) = \alpha_j(1-\alpha_j)X_{j,t-1} + \lambda_j \quad \text{and}$$

$$\mathrm{Cov}(X_{j,t+h}, X_{it}|X_{j,t-1}, X_{i,t-1}) = \phi$$

respectively, while otherwise it is the zero matrix.

3.1.2 Estimation

The conditional density for the Poisson BINAR(1) model can be obtained by substituting

$$f_3(k,s) = \frac{e^{-(\lambda_1+\lambda_2-\phi)} \sum_{m=0}^{min(k,s)} (\lambda_1-\phi)^{k-m}(\lambda_2-\phi)^{s-m}\phi^m}{(k-m)!(s-m)!m!} \tag{3.7}$$

in (2.15). Then we get

$$f(\mathbf{x}_t|\mathbf{x}_{t-1}, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi) = e^{-(\lambda_1+\lambda_2-\phi)} \sum_{k=0}^{g_1}\sum_{s=0}^{g_2} \frac{\sum_{m=0}^{min(k,s)} (\lambda_1-\phi)^{k-m}(\lambda_2-\phi)^{s-m}\phi^m}{(k-m)!(s-m)!m!}$$
$$\times \binom{x_{1,t-1}}{x_{1t}-k} \alpha_1^{x_{1t}-k}(1-\alpha_1)^{x_{1,t-1}-x_{1t}+k} \binom{x_{2,t-1}}{x_{2t}-s} \alpha_2^{x_{2t}-s}(1-\alpha_2)^{x_{2,t-1}-x_{2t}+s} \tag{3.8}$$

where $g_1 = min(x_{1t}, x_{1,t-1})$ and $g_2 = min(x_{2t}, x_{2,t-1})$.

3.2 A BINAR(1) Process with BVNB Innovations

3.2.1 Model

Assume that the *jpmf* of the innovations $\{R_{1t}, R_{2t}\}$ is a bivariate negative binomial distribution of the following form (Marshall and Olkin, 1990; Boucher et al., 2008; Cheon et al., 2009):

$$P(R_{1t} = x, R_{2t} = y) = \frac{\Gamma(\beta^{-1} + x + y)}{\Gamma(\beta^{-1})\Gamma(x+1)\Gamma(y+1)} \times \left(\frac{\lambda_1}{\lambda_1 + \lambda_2 + \beta^{-1}}\right)^x \left(\frac{\lambda_2}{\lambda_1 + \lambda_2 + \beta^{-1}}\right)^y \left(\frac{\beta^{-1}}{\lambda_1 + \lambda_2 + \beta^{-1}}\right)^{\beta^{-1}} \tag{3.9}$$

where $\lambda_1, \lambda_2, \beta > 0$. We will denote this distribution as BVNB$(\lambda_1, \lambda_2, \beta)$. Note that the marginal distribution of R_{jt} is univariate negative binomial with parameters β^{-1} and $p_j = \beta^{-1}/(\lambda_j + \beta^{-1})$, $j = 1, 2$ and that the correlation between the two count variables R_{1t} and R_{2t}

$$\text{Corr}(x, y) = \sqrt{\frac{\lambda_1 \lambda_2 \beta^2}{(1 + \lambda_1 \beta)(1 + \lambda_2 \beta)}} \tag{3.10}$$

must be positive. This assumption allows for more flexibility than the Poisson BINAR(1) model does, due to the involvement of the overdispersion parameter β in the model's specification.

Recall that in section 2.1, $\{R_{jt}\}$ was generally defined as an *iid* sequence with mean λ_j and variance $\sigma_j^2 = \upsilon_j \lambda_j$, $\upsilon_j > 0$, $j = 1, 2$. For the BVNB model, $\sigma_j^2 = \lambda_j(1 + \beta\lambda_j)$ implying that $\upsilon_j = 1 + \beta\lambda_j, \quad \lambda_j, \beta > 0$. Consequently $\upsilon_j > 1$ which indicates the overdispersion case. However, the resulting model is not a BINAR model with negative binomial marginals but a model that effectively accounts for overdispersion. In specific, the statistical properties of the BINAR(1) model with BVNB innovations are encompassed in the vector of expectations $\boldsymbol{\mu}_{\mathbf{X}_t} = E(\mathbf{X}_t)$ with elements

$$\mu_{X_{jt}} = \frac{\lambda_j}{1 - \alpha_j} \; ; \quad j = 1, 2 \tag{3.11}$$

the variance-covariance matrix $\boldsymbol{\gamma}_{\mathbf{X}_t}(h)$ with diagonal and off-diagonal elements equal to

$$\text{Cov}(X_{j,t+h}, X_{jt}) = \frac{\alpha_j^h \lambda_j (1 + \beta\lambda_j + \alpha_j)}{1 - \alpha_j^2} \; ; \quad j = 1, 2, \quad h = 0, 1, \ldots \tag{3.12}$$

and

$$\mathrm{Cov}(X_{j,t+h}, X_{it}) = \frac{\alpha_j^h \beta \lambda_1 \lambda_2}{1-\alpha_1\alpha_2} \; ; \quad j \neq i, \quad h = 0, 1, \ldots \tag{3.13}$$

respectively, and the correlation matrix $\boldsymbol{\rho}_{\mathbf{X}_t}(h)$ with diagonal elements

$$\mathrm{Corr}(X_{j,t+h}, X_{jt}) = \alpha_j^h \; ; \quad j = 1, 2, \quad h = 0, 1, \ldots \tag{3.14}$$

and off-diagonal elements

$$\mathrm{Corr}(X_{j,t+h}, X_{it}) = \frac{\alpha_j^h \beta}{(1-\alpha_1\alpha_2)} \sqrt{\frac{(1-\alpha_1^2)(1-\alpha_2^2)\lambda_1\lambda_2}{(1+\beta\lambda_1+\alpha_1)(1+\beta\lambda_2+\alpha_2)}} \; ; \quad j \neq i, \quad h = 0, 1, \ldots \tag{3.15}$$

Conditionally on the previous observations $\mathbf{X}_{t-1} = \{X_{1,t-1}, X_{2,t-1}\}$, the vector of conditional means $\boldsymbol{\mu}_{\mathbf{X}_{t|t-1}} = E(\mathbf{X}_{t|t-1})$ has elements

$$\mu_{X_{jt|t-1}} = \alpha_j X_{j,t-1} + \lambda_j, \quad j = 1, 2.$$

For $h = 0$, the conditional variance-covariance matrix $\boldsymbol{\gamma}_{\mathbf{X}_{t|t-1}}(h)$ has diagonal and off-diagonal elements equal to

$$\mathrm{Cov}(X_{j,t+h}, X_{jt}|X_{j,t-1}) = \alpha_j(1-\alpha_j)X_{j,t-1} + \lambda_j(1+\beta\lambda_j) \;\text{ and}$$
$$\mathrm{Cov}(X_{j,t+h}, X_{it}|X_{j,t-1}, X_{i,t-1}) = \beta\lambda_1\lambda_2$$

respectively, while otherwise it is the zero matrix.

3.2.2 Estimation

For the BINAR(1) model with BVNB innovations it holds that

$$f_3(k,s) = \frac{\Gamma(\beta^{-1}+k+s)}{\Gamma(\beta^{-1})k!s!} \left(\frac{\lambda_1}{\lambda_1+\lambda_2+\beta^{-1}}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2+\beta^{-1}}\right)^s \left(\frac{\beta^{-1}}{\lambda_1+\lambda_2+\beta^{-1}}\right)^{\beta^{-1}} \tag{3.16}$$

Thus the conditional density (2.15) becomes

$$\begin{aligned} f(\mathbf{x}_t|\mathbf{x}_{t-1}, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \beta) = \\ \sum_{k=0}^{g_1}\sum_{s=0}^{g_2} \frac{\Gamma(\beta^{-1}+k+s)}{\Gamma(\beta^{-1})k!s!} \left(\frac{\lambda_1}{\lambda_1+\lambda_2+\beta^{-1}}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2+\beta^{-1}}\right)^s \left(\frac{\beta^{-1}}{\lambda_1+\lambda_2+\beta^{-1}}\right)^{\beta^{-1}} \\ \times \binom{x_{1,t-1}}{x_{1t}-k} \alpha_1^{x_{1t}-k}(1-\alpha_1)^{x_{1,t-1}-x_{1t}+k} \binom{x_{2,t-1}}{x_{2t}-s} \alpha_2^{x_{2t}-s}(1-\alpha_2)^{x_{2,t-1}-x_{2t}+s} \end{aligned} \tag{3.17}$$

where $g_1 = min(x_{1t}, x_{1,t-1})$ and $g_2 = min(x_{2t}, x_{2,t-1})$.

3.3 Other distributional choices

As mentioned before, the choice of the joint distribution for R_{1t} and R_{2t} determines the properties of the underlying process. While the bivariate negative binomial provides overdispersion, it is interesting to note that a selection of a distribution with negative correlation can also produce negative correlation between the two series (see (2.7)). The literature on bivariate count distributions with negative correlation is limited. One of the reasons is that negative correlation in bivariate counts occurs rather infrequently. However there are such models in the literature, as for example the bivariate Poisson-lognormal model of Aitchinson and Ho (1989) (see also Chib and Winkelmann, 2001), the finite mixture model developed in Karlis and Meligkotsidou (2007) and models based on copulas see, e.g. Nikoloulopoulos and Karlis (2009) and the references therein. Finally, noted that while we used a certain bivariate negative binomial distribution, there are certain other alternatives in the literature which could have been used. We have selected this one mainly because of its relative simplicity.

4 Diagnostics

In this section we describe diagnostics for assessing the goodness of fit. Usually, in model fitting, this is accomplished by means of residual analysis. However, due to the structural distinctiveness of INAR-type models, the classical definition of residuals as differences between the observed and fitted values, may prove to be inadequate as a diagnostic tool. We follow Freeland and McCabe (2004a) by introducing a definition for residuals for count data that distinguishes between a set of residuals for the continuation process $r_{1t} = \alpha \circ X_{t-1} - \alpha X_{t-1}$ and another for the arrival component $r_{2t} = R_t - \lambda$. In this section we attempt to extend the ideas of Freeland and McCabe (2004a) to the BINAR(1) model.

For each one of the two series $\{X_{1t}, X_{2t}\}$, we define two sets of residuals; one for each random component. So, for the continuation components we let $r_{1t}^{(j)} = \alpha_j \circ X_{j,t-1} - \alpha_j X_{j,t-1}$ and for the arrival components we let $r_{2t}^{(j)} = R_{jt} - \lambda_j$, $j = 1, 2$. In order to arrive at a sensible and practical form of the above definitions, the unobservable quantities $\alpha_j \circ X_{j,t-1}$ and R_{jt} should be replaced with $E_t[\alpha_j \circ X_{j,t-1}]$ and $E_t[R_{jt}]$ respectively, i.e. with their conditional expectations given the observed values of X_{jt} and $X_{j,t-1}$.

PROPOSITION 1. Let $E_t[\cdot]$ denote the conditional expectation to the sigma field, $\Im_t = \sigma(X_{j0}, X_{j1}, ..., X_{jt})$, $j = 1, 2$. For the BINAR(1) model with bivariately distributed innovations the following equalities hold:

$$E_t[\alpha_1 \circ X_{1,t-1}] = \frac{\alpha_1 X_{1,t-1} P(x_{1t} - 1 | X_{1,t-1} - 1, X_{2,t-1})}{P(x_{1t} | X_{1,t-1}, X_{2,t-1})} \tag{4.1}$$

$$E_t[\alpha_2 \circ X_{2,t-1}] = \frac{\alpha_2 X_{2,t-1} P(x_{2t} - 1 | X_{1,t-1}, X_{2,t-1} - 1)}{P(x_{2t} | X_{1,t-1}, X_{2,t-1})} \tag{4.2}$$

$$E_t[R_{1t}] = \frac{\sum_{x_{2t}} \sum_{k=0}^{g_1} \sum_{s=0}^{g_2} k f_1(x_{1t} - k) f_2(x_{2t} - s) f_3(k, s)}{P(x_{1t} | X_{1,t-1}, X_{2,t-1})} \tag{4.3}$$

$$E_t[R_{2t}] = \frac{\sum_{x_{1t}} \sum_{k=0}^{g_1} \sum_{s=0}^{g_2} s f_1(x_{1t} - k) f_2(x_{2t} - s) f_3(k, s)}{P(x_{2t} | X_{1,t-1}, X_{2,t-1})} \tag{4.4}$$

where the densities $f_1(\cdot)$ and $f_2(\cdot)$ are given in (2.13) and (2.14), $f_3(k, s) = P(R_{1t} = k, R_{2t} = s)$, $g_1 = min(x_{1t}, x_{1,t-1})$ and $g_2 = min(x_{2t}, x_{2,t-1})$. Using Proposition 1 we can now define the residuals as

$$r_{1t}^{(j)\star} = E_t[r_{1t}^{(j)}] = E_t[\alpha_j \circ X_{j,t-1}] - \alpha_j X_{j,t-1}, \quad \text{and} \tag{4.5}$$

$$r_{2t}^{(j)\star} = E_t[r_{2t}^{(j)}] = E_t[R_{jt}] - \lambda_j, \quad \text{for} \quad j = 1, 2 \tag{4.6}$$

Regarding separately each one of the two series that comprise the BINAR(1) model, it is noted that adding the components of the two new sets of residuals gives the usual definition of residuals, i.e.

$$\begin{aligned} r_{1t}^{(j)\star} + r_{2t}^{(j)\star} &= E_t[\alpha_j \circ X_{j,t-1}] - \alpha_j X_{j,t-1} + E_t[R_{jt}] - \lambda_j \\ &= E_t[\alpha_j \circ X_{j,t-1} + R_{jt}] - \alpha_j X_{j,t-1} - \lambda_j \\ &= X_{jt} - \alpha_j X_{j,t-1} - \lambda_j = r_t^{(j)}. \end{aligned} \tag{4.7}$$

Thus, the adequacy of each component of the model may by assessed by plotting the aformentioned sets of residuals.

5 Forecasting

The usual way to produce forecasts in time series models is via the conditional forecast distribution. Freeland and McCabe (2004b) established the h-step-ahead conditional distribution of the Poisson INAR(1) model, based on the remark of Al-Osh and Alzaid (1987) that

$$(X_t, X_{t-h}) \overset{d}{=} \left(\alpha^h \circ X_{t-h} + \sum_{i=0}^{h-1} \alpha^i \circ R_{t-i}, X_{t-h} \right) \tag{5.1}$$

where R_t is a sequence of uncorrelated non-negative integer-valued random variables with finite mean and variance.

The above result holds also for the marginal distribution of each one of the two series (X_{1t}, X_{2t}) that consist a BINAR(1) model. As in the univariate case, $\alpha_j^h \circ X_{j,t-h} \mid X_{j,t-h}$, $j = 1, 2$, has a binomial distribution with parameters $(\alpha_j^h, X_{j,t-h})$. Moreover, the joint and marginal distributions of $\sum_{i=0}^{h-1} \alpha_1^i \circ R_{1,t-i}$ and $\sum_{i=0}^{h-1} \alpha_2^i \circ R_{2,t-i}$ are determined by the joint and marginal distributions of X_{1t} and X_{2t}. This relation can be described in terms of the *jpgf* of $\left\{\sum\limits_{i=0}^{h-1} \alpha_1^i \circ R_{1,t-i}, \sum\limits_{i=0}^{h-1} \alpha_2^i \circ R_{2,t-i}\right\}$. Denote by S_j the quantity $\sum\limits_{i=0}^{h-1} \alpha_j^i \circ R_{1,t-j}$, $j = 1, 2$. Then,

$$G_{S_1,S_2}(s_1, s_2) = \prod_{i=0}^{h-1} G_{R_1,R_2}((1 - \alpha_1^i + \alpha_1^i s_1), (1 - \alpha_2^i + \alpha_2^i s_2)) \tag{5.2}$$

Hence, the joint distribution of $\{X_{1t}, X_{2t}\}$ given $\{X_{1,t-h}, X_{2,t-h}\}$ is a convolution of two binomial distributions with parameters $(\alpha_1^h, X_{1,t-h})$ and $(\alpha_2^h, X_{2,t-h})$ respectively, and a bivariate distribution with *jpgf* of the form (5.2). Obviously, if (5.2) has not a closed-form expression, then neither the h-step-ahead forecast distribution can be specified in closed-form. However, it is straightforward to evaluate it numerically.

For the Poisson BINAR(1) model, it can be proved that

$$\begin{aligned} G_{S_1,S_2}(s_1, s_2) &= \exp\left[\left(\frac{1-\alpha_1^h}{1-\alpha_1}\right)\lambda_1(s_1 - 1) + \left(\frac{1-\alpha_2^h}{1-\alpha_2}\right)\lambda_2(s_2 - 1)\right. \\ &+ \left.\left(\frac{1-\alpha_1^h\alpha_2^h}{1-\alpha_1\alpha_2}\right)\phi(s_1 - 1)(s_2 - 1)\right] \end{aligned} \tag{5.3}$$

while the corresponding *jpgf* for the BINAR(1) model with BVNB innovations is given by

$$G_{S_1,S_2}(s_1, s_2) = \prod_{i=0}^{h-1} \left[1 - \beta\lambda_1\alpha_1^i(s_1 - 1) - \beta\lambda_2\alpha_2^i(s_2 - 1)\right]^{-\beta^{-1}} \tag{5.4}$$

which is not of a convenient form.

Note however that irrespective of the *jpgf* of $\left\{\sum\limits_{i=0}^{h-1} \alpha_1^i \circ R_{1,t-i}, \sum\limits_{i=0}^{h-1} \alpha_2^i \circ R_{2,t-i}\right\}$, closed-form expressions are available for their conditional expectations and variances (conditional on X_{1t}, X_{2t}). More specifically, it can be proved that

$$E\left(\sum_{i=0}^{h-1} \alpha_j^i \circ R_{j,t-i}\right) = \left(\frac{1-\alpha_j^h}{1-\alpha_j}\right)\lambda_j \tag{5.5}$$

and

$$Var\left(\sum_{i=0}^{h-1} \alpha_j^i \circ R_{j,t-i}\right) = \left(\frac{1-\alpha_j^{2h}}{1-\alpha_j^2}\right)\upsilon_j\lambda_j + \left(\frac{1-\alpha_j^h}{1-\alpha_j} - \frac{1-\alpha_j^{2h}}{1-\alpha_j^2}\right)\lambda_j \tag{5.6}$$

THEOREM 1. The *jpmf* of $\{X_{1,T+h}, X_{2,T+h}\}$ given $\{x_{1T}, x_{2T}\}$ is given by

$$\begin{aligned}
& P_h(X_{1,T+h} = x_1, X_{2,T+h} = x_2 | x_{1T}, x_{2T}) = \\
& \sum_{k=0}^{min(x_1,x_{1T})} \sum_{s=0}^{min(x_2,x_{2T})} \left(\begin{array}{c} x_{1T} \\ x_1 - k \end{array}\right) (\alpha_1^h)^{x_1-k}(1-\alpha_1^h)^{x_{1T}-x_1+k} \\
\times & \left(\begin{array}{c} x_{2T} \\ x_2 - s \end{array}\right) (\alpha_2^h)^{x_2-s}(1-\alpha_2^h)^{x_{2T}-x_2+s} \\
\times & f\left(\sum_{i=0}^{h-1} \alpha_1^i \circ R_{1,T+h-i} = k, \sum_{i=0}^{h-1} \alpha_2^i \circ R_{2,T+h-i} = s | x_{1T}, x_{2T}\right)
\end{aligned}$$

with means,

$$E(x_{j,T+h}|x_{2T}, x_{2T}) = \alpha_j^h x_{jT} + \left(\frac{1-\alpha_j^h}{1-\alpha_j}\right) E(R_{jt})\ ; \quad j = 1,2, \quad h = 1,2,\ldots \tag{5.7}$$

and variances,

$$\begin{aligned}
Var(x_{j,T+h}|x_{1T}, x_{2T}) \;\; = \;\; & \alpha_j^h(1-\alpha_j^h)x_{jT} + \left(\frac{1-\alpha_j^{2h}}{1-\alpha_j^2}\right) Var(R_{jt}) \\
& + \left(\frac{1-\alpha_j^h}{1-\alpha_j} - \frac{1-\alpha_j^{2h}}{1-\alpha_j^2}\right) E(R_{jt})\ ; \quad j = 1,2, \quad h = 1,2,\ldots
\end{aligned} \tag{5.8}$$

The corresponding *jpgf* of $\{X_{1,T+h}, X_{2,T+h}\}$ given $\{x_{1T}, x_{2T}\}$ is of the form

$$G_{X_{1,T+h},X_{2,T+h}}(s_1,s_2|x_{1T},x_{2T}) = (1-\alpha_1^h+\alpha_1^h s_1)^{X_{1T}}(1-\alpha_2^h+\alpha_2^h s_2)^{X_{2T}} G_{S_1,S_2}(s_1,s_2) \tag{5.9}$$

where $G_{S_1,S_2}(s_1,s_2)$ is given in (5.2).

Corollary 1. For the Poisson BINAR(1) model, the *jpgf* and *jpmf* of $\{X_{1,T+h}, X_{2,T+h}\}$ given $\{x_{1T}, x_{2T}\}$ are given by

$$\begin{aligned} & G_{X_{1,T+h},X_{2,T+h}}(s_1,s_2|x_{1T},x_{2T}) = (1-\alpha_1^h+\alpha_1^h s_1)^{X_{1T}}(1-\alpha_2^h+\alpha_2^h s_2)^{X_{2T}} \\ \times \quad & \exp\left\{\left(\frac{1-\alpha_1^h}{1-\alpha_1}\right)\lambda_1 s_1 + \left(\frac{1-\alpha_2^h}{1-\alpha_2}\right)\lambda_2 s_2 + \left(\frac{1-\alpha_1^h\alpha_2^h}{1-\alpha_1\alpha_2}\right)\phi(s_1-1)(s_2-1)\right\} \end{aligned}$$

and

$$\begin{aligned} & P_h(X_{1,T+h}=x_1, X_{2,T+h}=x_2|x_{1T},x_{2T}) = \\ & \sum_{k=0}^{min(x_1,x_{1T})}\sum_{s=0}^{min(x_2,x_{2T})} \binom{x_{1T}}{x_1-k}(\alpha_1^h)^{x_1-k}(1-\alpha_1^h)^{x_{1T}-x_1+k} \\ \times \quad & \binom{x_{2T}}{x_2-s}(\alpha_2^h)^{x_2-s}(1-\alpha_2^h)^{x_{2T}-x_2+s} \\ \times \quad & \exp\left\{-\left[\left(\frac{1-\alpha_1^h}{1-\alpha_1}\right)\lambda_1 + \left(\frac{1-\alpha_2^h}{1-\alpha_2}\right)\lambda_2 - \left(\frac{1-\alpha_1^h\alpha_2^h}{1-\alpha_1\alpha_2}\right)\phi\right]\right\} \\ \times \quad & \sum_{m=0}^{min(k,s)} \frac{\left[\left(\frac{1-\alpha_1^h}{1-\alpha_1}\right)\lambda_1 - \left(\frac{1-\alpha_1^h\alpha_2^h}{1-\alpha_1\alpha_2}\right)\phi\right]^{k-m}\left[\left(\frac{1-\alpha_2^h}{1-\alpha_2}\right)\lambda_2 - \left(\frac{1-\alpha_1^h\alpha_2^h}{1-\alpha_1\alpha_2}\right)\phi\right]^{s-m}\left[\left(\frac{1-\alpha_1^h\alpha_2^h}{1-\alpha_1\alpha_2}\right)\phi\right]^{m}}{(k-m)!(s-m)!m!} \end{aligned}$$

respectively, with means,

$$E(x_{j,T+h}|x_{1T},x_{2T}) = \alpha_j^h x_{jT} + \frac{1-\alpha_j^h}{1-\alpha_j}\lambda_j\ ; \quad j=1,2, \quad h=1,2,\ldots \tag{5.10}$$

variances,

$$Var(x_{j,T+h}|x_{1T},x_{2T}) = \alpha_j^h(1-\alpha_j^h)x_{jT} + \frac{1-\alpha_j^h}{1-\alpha_j}\lambda_j\ ; \quad j=1,2, \quad h=1,2,\ldots \tag{5.11}$$

and covariance,

$$\text{Cov}(x_{1,T+h}, x_{2,T+h}|x_{1T}, x_{2T}) = \left(\frac{1-\alpha_1^h\alpha_2^h}{1-\alpha_1\alpha_2}\right)\phi\ ; \quad h = 1, 2, \ldots \tag{5.12}$$

Corollary 2. For the BINAR(1) model with BVNB innovations, the *jpgf* and the *jpmf* of $\{X_{1,T+h}, X_{2,T+h}\}$ given $\{x_{1T}, x_{2T}\}$ are given by

$$G_{X_{1,T+h},X_{2,T+h}}(s_1, s_2|x_{1T}, x_{2T}) = (1-\alpha_1^h+\alpha_1^h s_1)^{X_{1T}}(1-\alpha_2^h+\alpha_2^h s_2)^{X_{2T}}$$
$$\times \prod_{i=0}^{h-1}\left[1-\beta\lambda_1\alpha_1^i(s_1-1)-\beta\lambda_2\alpha_2^i(s_2-1)\right]^{-\beta^{-1}}$$

and

$$\begin{aligned} &P_h(X_{1,T+h} = x_1, X_{2,T+h} = x_2|x_{1T}, x_{2T}) = \\ &\sum_{k=0}^{min(x_1,x_{1T})}\sum_{s=0}^{min(x_2,x_{2T})}\left(\begin{array}{c} x_{1T} \\ x_1-k \end{array}\right)(\alpha_1^h)^{x_1-k}(1-\alpha_1^h)^{x_{1T}-x_1+k} \\ \times &\left(\begin{array}{c} x_{2T} \\ x_2-s \end{array}\right)(\alpha_2^h)^{x_2-s}(1-\alpha_2^h)^{x_{2T}-x_2+s} \\ \times &\ f\left(\sum_{i=0}^{h-1}\alpha_1^i\circ R_{1,T+h-i} = k, \sum_{i=0}^{h-1}\alpha_2^i\circ R_{2,T+h-i} = s|x_{1T}, x_{2T}\right) \end{aligned}$$

respectively, where $f\left(\sum_{i=0}^{h-1}\alpha_1^i\circ R_{1,T+h-i} = k, \sum_{i=0}^{h-1}\alpha_2^i\circ R_{2,T+h-i} = s|x_{1T}, x_{2T}\right)$ can be numerically calculated.
The means and variances of this process are given by,

$$E(x_{j,T+h}|x_{1T}, x_{2T}) = \alpha_j^h x_{jT} + \frac{1-\alpha_j^h}{1-\alpha_j}\lambda_j\ ; \quad j = 1, 2, \quad h = 1, 2, \ldots \tag{5.13}$$

and

$$\begin{aligned} Var(x_{j,T+h}|x_{1T}, x_{2T}) &= \alpha_j^h(1-\alpha_j^h)x_{jT} + \left(\frac{1-\alpha_j^{2h}}{1-\alpha_j^2}\right)(1+\beta\lambda_j)\lambda_j \\ &+ \left(\frac{1-\alpha_j^h}{1-\alpha_j} - \frac{1-\alpha_j^{2h}}{1-\alpha_j^2}\right)\lambda_j\ ; \quad j = 1, 2, \quad h = 1, 2, \ldots \end{aligned} \tag{5.14}$$

whereas the covariance function is not of a closed-form.

The marginal probabilities $P_h(x_1|x_{1T}, x_{2T})$ and $P_h(x_2|x_{1T}, x_{2T})$ can be calculated directly as, $P_h(x_1|x_{1T}, x_{2T}) = \sum_{x_2} P_h(x_1, x_2|x_{1T}, x_{2T})$ and $P_h(x_2|x_{1T}, x_{2T}) = \sum_{x_1} P_h(x_1, x_2|x_{1T}, x_{2T})$ respectively.

Given the fact that the vector of parameters $\boldsymbol{\theta}$ is unknown, in practice we are only able to compute $P_h(x_1, x_2|x_{1T}, x_{2T}; \hat{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ are typically the maximum likelihood estimators introduced in section 2.2. Lack of knowledge about the true values of the model parameters and the need to estimate them introduce uncertainty in the estimation of the h-step-ahead *jpmf*'s. Estimation uncertainty, i.e. the error made in estimating these probabilities, can be assessed by taking advantage of the asymptotic normality of ML estimators. Under standard regularity conditions, the ML estimator $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, is asymptotically normally distributed around the true parameter value, i.e. $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{a}{\sim} N(0, \boldsymbol{i}^{-1})$, where $\boldsymbol{i}^{-1}$ is the inverse of the Fisher information matrix (Bu and McCabe, 2008). The δ-method can then be used for finding the asymptotic distribution of a random variable $g(\hat{\boldsymbol{\theta}})$. An application of the δ-method to $g(\hat{\boldsymbol{\theta}}) = P_h(\mathbf{x}|\mathbf{x}_T; \hat{\boldsymbol{\theta}})$ provides us with a confidence interval for the probability associated with any fixed value of $\mathbf{x} = (x_1, x_2)$ in the forecast distribution. Obviously, these intervals may be truncated outside [0,1] (Freeland and McCabe, 2004b).

THEOREM 2 (Freeland and McCabe, 2004b)**:** The quantity $P_h(\mathbf{x}|\mathbf{x}_T; \hat{\boldsymbol{\theta}})$ has an asymptotically normal distribution with mean $P_h(\mathbf{x}|\mathbf{x}_T; \boldsymbol{\theta}_{\mathbf{0}})$ and variance

$$\sigma_h^2(\mathbf{x}; \boldsymbol{\theta}_0) = T^{-1} \left\{ \left(\left. \frac{\partial P_h}{\partial \hat{\boldsymbol{\theta}}'} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \boldsymbol{i}^{-1} \left(\left. \frac{\partial P_h}{\partial \hat{\boldsymbol{\theta}}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \right\} \qquad (5.15)$$

It is apparent that analytical expressions for (5.15) are only available in cases where $P_h(\mathbf{x}|\mathbf{x}_T)$ has a closed-form expression as is the case for the Poisson BINAR(1) model (see Appendix).

Closing this section, it is worth noting that summarizing the forecast distribution by means of conditional expectations, while ensures a minimum mean square error, it has drawbacks with respect to data coherency since the integer-valued property of the time series is not taken into account. Freeland and McCabe (2004b) suggest instead the use of the median of this distribution which always lies in the support of the series and is therefore coherent. Also, Pavlopoulos and Karlis (2008) propose a parametric bootstrap approach which guarantees both integer-valued predictions and prediction intervals with integer-valued ends. In our case, since the distributions are discrete, it is relatively easy to find the median to use as prediction instead of the mean, as the median will satisfy the discrete nature of the data.

6 Application

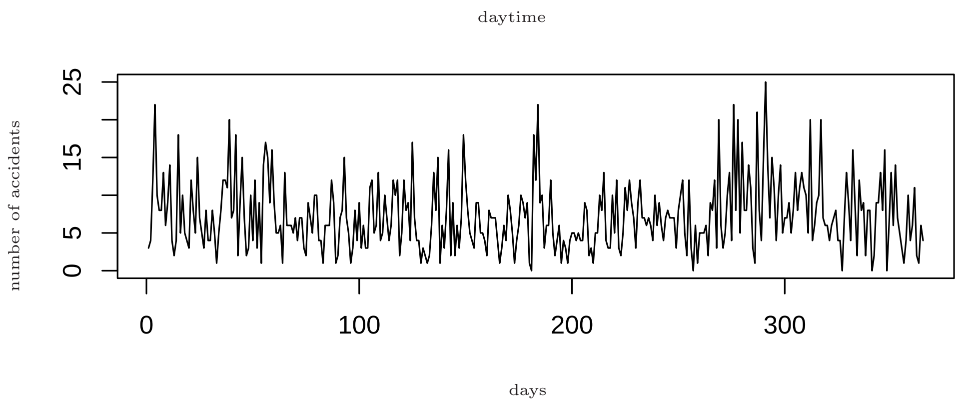
The data used in this application refer to the joint modelling of daytime and nighttime road accidents in Schiphol area, in the Netherlands for the year 2001. As nighttime accidents we refer to accidents happened between 10.00am-06.00pm, while the rest were considered as daytime accidents. In accident analysis those types of accidents are considered to have different behavior. During nighttimes the traffic is of different nature (e.g. more people travel for entertainment). On the other hand since both types share the same environment, like weather conditions, characteristics of the road, they are correlated. The data are daily observations.

Data from successive days are typically correlated as they refer to similar conditions. Ignoring this time series nature of the data can lead to incorrect inference (see Brijs et al., 2008). Hence joint modelling of the two time series can be very useful. For such data typically the autocorrelations are relatively small and thus AR(1) models are enough to capture the time dependence.

The data can be seen in figure 1. The daytime and nighttime accidents have mean values (variances) equal to 7.27 (20.93) and 1.50 (1.87) respectively implying ovedispersion. The autocorrelation functions for both series present a rather exponential decay with a few exceptions. The first order autocorrelation coefficient is 0.12 for daytime accidents and 0.13 for nighttime accidents. Finally the correlation between the two time series is 0.145 revealing a short of correlation between the series.

In order to model the data we considered both the bivariate Poisson INAR(1) model and the INAR(1) model with BVNB innovations. The results can be seen in Table 1. Comparing the log-likelihood one can see that both the time series context and the correlation between the series are needed. The negative binomial BINAR(1) model can also model the overdispersion and thus it provides the better fit. We have also fitted a model with bivariate Poisson lognormal innovations which having two more parameters offered very small improvement being much more computationally demanding (results are not presented here).

It is clear that the time series models are better than the models that neglect this. In addition the BVNB INAR(1) model is much better as it captures the overdispersion in the dataset together with the correlation between the two series but also the autocorrelation within each series. The standard



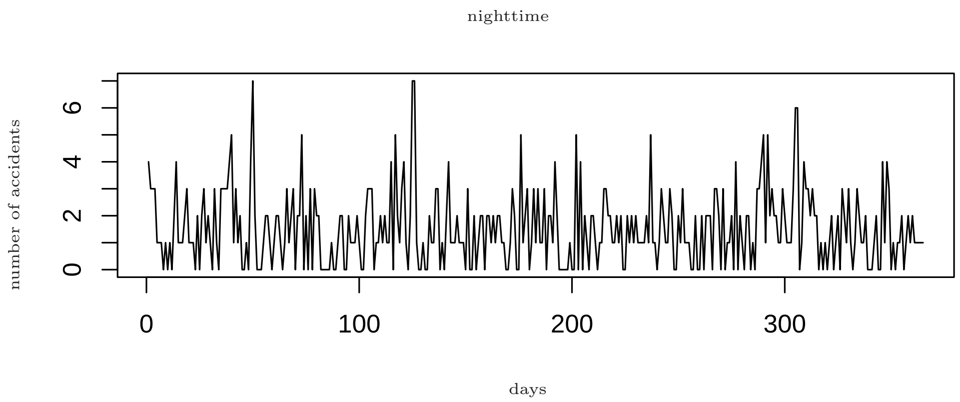


Figure 1: Time series plots and acf plots for the series of daytime and night-time accidents.

Table 1: Maximum Likelihood Estimates from fitting alternatively a BINAR(1), two independent INAR(1) models and a simple bivariate Poisson model.

	BINAR(1)		Independent INAR(1)		Biv. Poisson		Neg.Bin BINAR(1)	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
$\hat{\alpha}_1$	0.0769	0.0253	$0.0805^{\dagger}$	$0.0253^{\dagger}$			0.0427	0.0380
$\hat{\alpha}_2$	0.0867	0.0430	$0.0991^{\ddagger}$	$0.0431^{\ddagger}$			0.1144	0.0444
$\hat{\lambda}_1$	6.7282	0.2261	$6.7022^{\dagger}$	$0.2258^{\dagger}$	7.2885	0.1415	6.9771	0.3537
$\hat{\lambda}_2$	1.3667	0.0873	$1.3481^{\ddagger}$	$0.0868^{\ddagger}$	1.4973	0.0641	1.3250	0.0937
	$\hat{\phi}$ =0.2682	0.1052			0.3098	0.1048	$\hat{\beta}$ =0.2248	0.0297
Log-Lik	-1744.7200		-1748.1490		-1751.5670		-1619.3770	
AIC	3499.44		3504.298		3509.134		3248.754	

† *daytime accidents*

‡ *nighttime accidents*

errors of the estimates obtained by the two approaches (standard errors are derived numerically from the Hessian) show that fitting a BINAR(1) model to the data generally improves the precision of the produced estimates. On the other hand it is apparent that ignoring any form of the correlation (either within or between) or the overdispersion leads to incorrect standard errors and hence incorrect inferences.

Figures 2-4 and the related inference concern results obtained from the BINAR(1) model with BVNB innovations. Figure 2 includes the plots of the residuals of the two series. Since these residuals have not been standardized, the survival and arrival residuals add up to the Pearson residuals. Moreover, a large Pearson residual is comprised by a large survival and arrival residual, while a small Pearson residual consists of a small survival and arrival residual. The signs of survival and arrival residuals may also differ in some cases. However, they still keep their similarity in pattern.

Another interesting point is the reflection of the model structure in the correlation between different pairs of residuals. More specifically, the sample correlations between the survival and arrival residuals of each series are very high: 0.72 for the daytime series and 0.67 for the nighttime series. The arrival residuals of the two series are also significantly correlated at 0.16 depicting the structural assumption underlying the BINAR(1) model that the correlation between the two series has been introduced by using correlated innovation terms.

Figure 3 shows the one-step-ahead marginal predictive distributions $P(x_{1,n+1}|x_{1n}, x_{2n})$ and $P(x_{2,n+1}|x_{1n}, x_{2n})$ where X_1 corresponds to the daytime series and X_2 corresponds to the nighttime series. The last observation was equal to 4 for the former series and equal to 1 for the latter series. As one can see in Figure 3, both distributions are skewed to the right which is in accordance with the shape of the negative binomial distribution. The most probable one-step-ahead predictive value is equal to 7 for the daytime series and equal to 1 for the nighttime series. The larger dispersion of the series of daytime accidents compared with the nighttime accidents series is also reflected in the plot of its predictive distribution.

Figure 4 shows the observed values of the series of daytime and nighttime accidents with the corresponding one-step-ahead predictions. The divergence between real data and forecasts is also portrayed. The horizontal lines correspond to the observed mean values of the two series. Obviously, divergence is larger for observations that lie far away from the mean. This seems to be expected since the one-step-ahead predictions have the same mean but are less dispersed than the original series. Note also that the correlation coefficient of the two series of forecasts is equal to the correlation coefficient of the real data series.

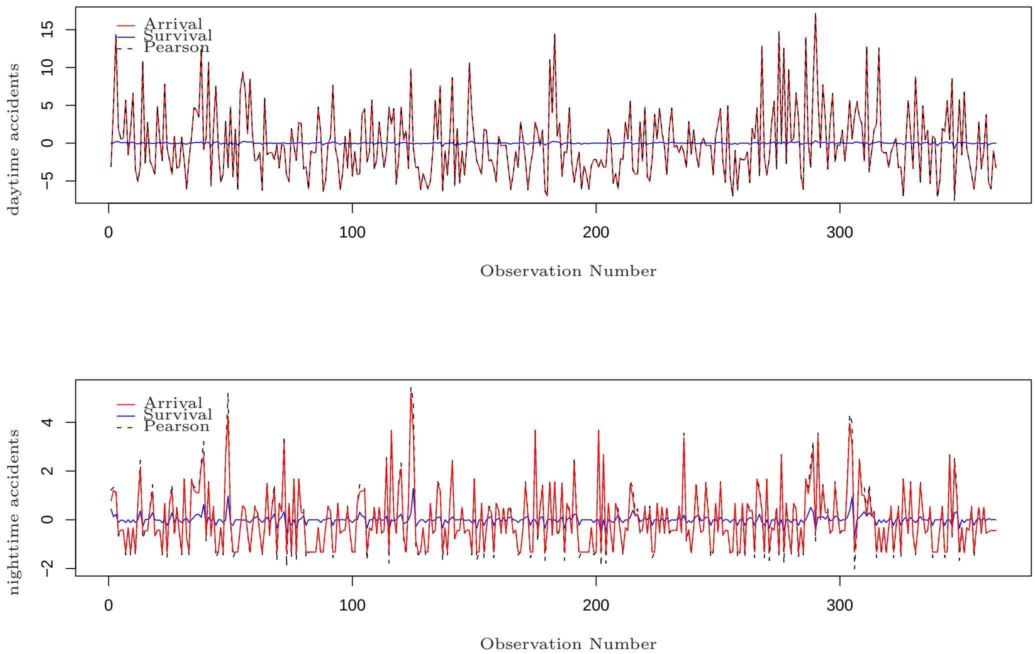


Figure 2: Non-standardized residuals of the daytime and nighttime accidents' series.

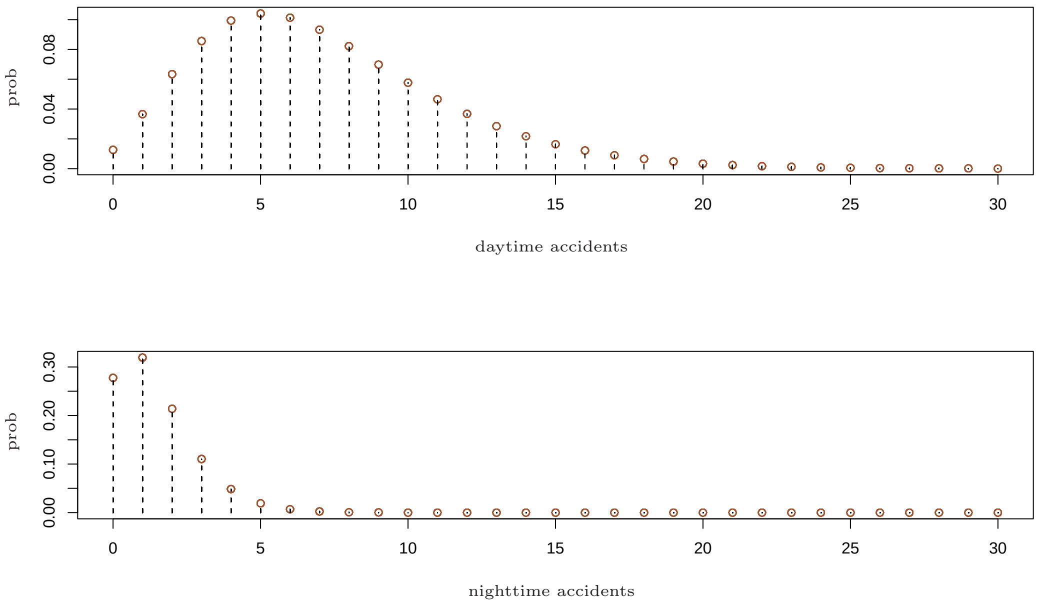


Figure 3: The one-step-ahead predictive distributions $P(x_{1,T+1}|x_{1T}, x_{2T})$ $P(x_{2,T+1}|x_{1T}, x_{2T})$ of the series of daytime and nighttime accidents respectively. The last observed values ($n = 365$) are equal to 4 and 1 accordingly.

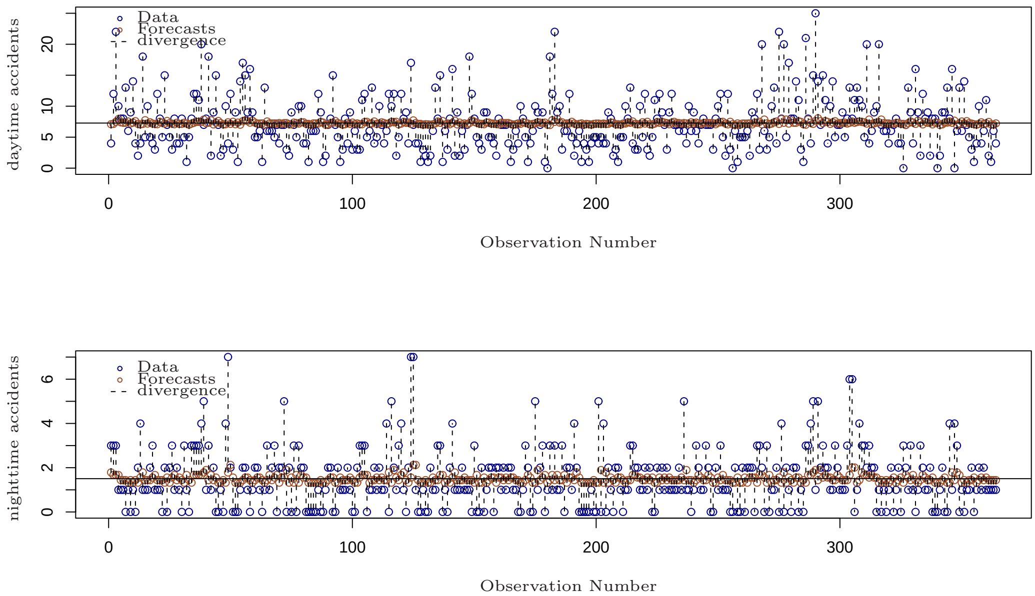


Figure 4: Observed values of the series of daytime and nighttime accidents and the corresponding one-step-ahead predictions.The horizontal lines correspond to the observed mean values of the two series.

7 Concluding Remarks

The main focus in this paper is on bivariate time series for count data. Generally, the desired BINAR(1) model can be constructed in two different ways: The first approach prespecifies the form of the marginal distributions and subsequently identifies the required form of the distribution of the innovations in order for stationarity to hold. In the second approach it is the choice of the form of the innovations distribution that leads to the specification of the underlying marginal distributions. The models proposed in this paper have been built following the last approach. In particular, we considered two different BINAR(1) models, one with bivariate Poisson innovations and another one with bivariate negative binomial innovations. The former specification has the useful property that the joint distribution of the two series under consideration is also bivariate Poisson. In the latter case, we don't end up with a bivariate negative binomial INAR(1) process but we obtain a BINAR(1) model that effectively accounts for overdispersion. Deviations from the equidispersion restriction could alternatively accounted for by assuming another distribution for the innovations, e.g. mixed Poisson, or by the inclusion of appropriate regressors. Results on such extensions will be reported elsewhere.

It is of course self-evident that the proposed model is not a panacea. For example, when significant correlation between the series under consideration is present at lags higher than 1, fitting a BINAR(1) model proves to be rather inadequate. Thus, extensions of the present model to higher orders would be a useful contribution to the improvement of its flexibility. Moreover, the structure of real-life data frequently implies need for the inclusion of both autoregressive and moving average components (when for example seasonal patterns are observed in time series of counts). So, extending the bivariate INAR model to a bivariate INARMA model seems to be another interesting challenge. Finally, generalization of the proposed process to the multivariate case would provide a great opportunity for modelling more than two time series of correlated count data. In this case, the definition of a multivariate discrete distribution for the innovation process is needed. The existing models have certain limitations and they do not lead to models with well specified marginals. Hence inference can be difficult with standard methods like maximum likelihood and some alternatives, like composite likelihood, should be considered.

References

J. Aitchinson and C. Ho. The multivariate Poisson-log normal distribution. *Biometrika*, 75:621–629, 1989.

M.A. Al-Osh and A.A. Alzaid. First-Order Integer-Valued Autoregressive Process. *Journal of Time Series Analysis*, 8(3):261–275, 1987.

J.P. Boucher, D. Michel, and G. Montserrat. Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, 2(1):135–162, 2008.

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

T. Brijs, D. Karlis, and G. Weets. Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accident Analysis and Prevention*, 40:1180–1190, 2008.

R. Bu and B. McCabe. Model selection, estimation and forecasting in INAR(p) models: A likelihood-based Markov Chain approach. *International Journal of Forecasting*, 24:151–162, 2008.

S. Cheon, S.H. Song, and B.C. Jung. Tests for independence in a bivariate negative binomial model. *Journal of the Korean Statistical Society*, 38: 185–190, 2009.

S. Chib and R. Winkelmann. Markov Chain Monte Carlo Analysis of Correlated Count Data. *Journal of Business & Economic Statistics*, 19(4): 428–435, 2001.

R.A. Davis, W.T. Dunsmuir, and Y. Wang. *Modelling time series of count data*, pages 63–114. Ed. S. Ghosh, Marcel Dekker, 1999.

R.K. Freeland and B.P.M. McCabe. Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, 25(5):701–722, 2004a.

R.K. Freeland and B.P.M. McCabe. Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20:427–434, 2004b.

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

N. Johnson, S. Kotz, and N. Balakrishnan. *Multivariate Discrete Distributions*. Wiley, New York, 1997.

R.C. Jung and A.R. Tremayne. Binomial thinning models for integer time series. *Statistical Modelling*, 6:21–96, 2006.

D. Karlis and L. Meligkotsidou. Finite multivariate Poisson mixtures with applications. *Journal of Statistical Planning and Inference*, 137:1942–1960, 2007.

S. Kocherlakota and K. Kocherlakota. *Bivariate Discrete Distributions, Statistics: textbooks and monographs*, volume 132. Markel Dekker, New York, 1992.

A. Latour. The Multivariate GINAR(p) Process. *Advances in Applied Probability*, 29(1):228–248, 1997.

A.W. Marshall and I. Olkin. *Multivarirate distribution generated from mixtures of convolution and product families*, pages 371–393. Topics in Statistical Dependence, Block,Sampson y Sanits (Eds), Institute of Mathematical Statistics, 1990.

E. McKenzie. Some Simple Models for Discrete Variate Time Series. *Water Resources Bulletin*, 21(4):645–650, 1985.

E. McKenzie. *Discrete Variate Time Series*, volume 21. Shanbhag, D. and Rao, C. (eds.), Elsevier Science, 2003.

A.K. Nikoloulopoulos and D. Karlis. Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference (In Press)*, 2009.

H. Pavlopoulos and D. Karlis. INAR(1) modeling of overdispersed count series with an environmental application. *Environmetrics*, 19(4):369–393, 2008.

A.M.M.S. Quoreshi. Bivariate Time Series Modeling of Financial Count Data. *Communications in Statistics - Theory and Methods*, 35:1343–1358, 2006.

N. Silva, I. Pereira, and E.S. Siva. Estimation and forecasting in SUINAR(1). *REVSTAT - Statistical Journal*, 6(3):253–277, 2008.

F.W. Steutel and K. van Harn. Discrete Analogues of Self - Decomposability and Stability. *The Annals of Probability*, 7(5):893–899, 1979.

Appendix

I. PROOF OF EQUATION (2.6).

$$
\begin{aligned}
\mathrm{Cov}(X_{1t}, R_{2t}) &= E(X_{1t}R_{2t}) - E(X_{1t})E(R_{2t}) \\
&= E(X_{1t}R_{2t}) - \mu_1\lambda_2 \\
&= E\left[\left(\sum_{i=0}^{\infty} \alpha_1^i \circ R_{1,t-i}\right) R_{2t}\right] - \mu_1\lambda_2 \\
&= \sum_{i=0}^{\infty} \alpha_1^i \left\{E(R_{1,t-i}R_{2t})\right\} - \mu_1\lambda_2 \\
&= E(R_{1t}R_{2t}) + \sum_{i=1}^{\infty} \alpha_1^i \left\{E(R_{1,t-1}R_{2t})\right\} - \mu_1\lambda_2 \\
&= \mathrm{Cov}(R_{1t}, R_{2t}) + \sum_{i=0}^{\infty} \alpha_1^i \left\{\lambda_1\lambda_2\right\} - \mu_1\lambda_2 \\
&= \mathrm{Cov}(R_{1t}, R_{2t})
\end{aligned}
$$

II. ESTIMATION UNCERTAINTY OF THE POISSON BINAR(1) MODEL.

For the Poisson BINAR(1) model, we let $\hat{\boldsymbol{\theta}}_T = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\phi})$ be the ML estimators of $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi)$ based on a sample of size T. Then, (5.15) can be written as

$$
\begin{aligned}
\sigma_h^2(\mathbf{x}; \boldsymbol{\theta}_0) &= T^{-1}\Bigg\{\left(\frac{\partial P_h}{\partial \alpha_1}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^2 i_{1,1}^{-1} + \left(\frac{\partial P_h}{\partial \alpha_2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^2 i_{2,2}^{-1} + \left(\frac{\partial P_h}{\partial \lambda_1}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^2 i_{3,3}^{-1} \\
&+ \left(\frac{\partial P_h}{\partial \lambda_2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^2 i_{4,4}^{-1} + \left(\frac{\partial P_h}{\partial \phi}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^2 i_{5,5}^{-1} + 2\left(\frac{\partial P_h}{\partial \alpha_1}\frac{\partial P_h}{\partial \alpha_2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{1,2}^{-1} \\
&+ 2\left(\frac{\partial P_h}{\partial \alpha_1}\frac{\partial P_h}{\partial \lambda_1}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{1,3}^{-1} + 2\left(\frac{\partial P_h}{\partial \alpha_1}\frac{\partial P_h}{\partial \lambda_2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{1,4}^{-1} + 2\left(\frac{\partial P_h}{\partial \alpha_1}\frac{\partial P_h}{\partial \phi}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{1,5}^{-1} \\
&+ 2\left(\frac{\partial P_h}{\partial \alpha_2}\frac{\partial P_h}{\partial \lambda_1}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{2,3}^{-1} + 2\left(\frac{\partial P_h}{\partial \alpha_2}\frac{\partial P_h}{\partial \lambda_2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{2,4}^{-1} + 2\left(\frac{\partial P_h}{\partial \alpha_2}\frac{\partial P_h}{\partial \phi}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{2,5}^{-1} \\
&+ 2\left(\frac{\partial P_h}{\partial \lambda_1}\frac{\partial P_h}{\partial \lambda_2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{3,4}^{-1} + 2\left(\frac{\partial P_h}{\partial \lambda_1}\frac{\partial P_h}{\partial \phi}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{3,5}^{-1} + 2\left(\frac{\partial P_h}{\partial \lambda_2}\frac{\partial P_h}{\partial \phi}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) i_{4,5}^{-1}\Bigg\}
\end{aligned}
$$

where $i^{-1}_{k,j}$ is the k,j-element of the matrix $\boldsymbol{i}^{-1}$, $k,j = 1,2,...,5$ and

$$\begin{aligned}
\frac{\partial P_h}{\partial \alpha_1} &= \frac{h}{1-\alpha_1^h} x_{1T} \left\{ P_h(x_1-1, x_2|x_{1T}-1, x_{2T}) - \alpha_1^{h-1} P_h(x_1, x_2|x_{1T}, x_{2T}) \right\} \\
&- \frac{\lambda_1(1-h\alpha_1^{h-1}-(1-h)\alpha_1^h)}{(1-\alpha_1^2)} \{P_h(x_1,x_2|x_{1T},x_{2T}) - P_h(x_1-1,x_2|x_{1T},x_{2T})\} \\
&+ \frac{\alpha_2\phi(1-h\alpha_1^{h-1}\alpha_2^{h-1}-(1-h)\alpha_1^h\alpha_2^h)}{(1-\alpha_1\alpha_2)^2} \{P_h(x_1,x_2|x_{1T},x_{2T}) \\
&- P_h(x_1-1,x_2|x_{1T},x_{2T}) - P_h(x_1,x_2-1|x_{1T},x_{2T}) + P_h(x_1-1,x_2-1|x_{1T},x_{2T})\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial P_h}{\partial \alpha_2} &= \frac{h}{1-\alpha_2^h} x_{2T} \left\{ P_h(x_1, x_2-1|x_{1T}, x_{2T}-1) - \alpha_2^{h-1} P_h(x_1, x_2|x_{1T}, x_{2T}) \right\} \\
&- \frac{\lambda_2(1-h\alpha_2^{h-1}-(1-h)\alpha_2^h)}{(1-\alpha_2^2)} \{P_h(x_1,x_2|x_{1T},x_{2T}) - P_h(x_1,x_2-1|x_{1T},x_{2T})\} \\
&+ \frac{\alpha_1\phi(1-h\alpha_1^{h-1}\alpha_2^{h-1}-(1-h)\alpha_1^h\alpha_2^h)}{(1-\alpha_1\alpha_2)^2} \{P_h(x_1,x_2|x_{1T},x_{2T}) \\
&- P_h(x_1-1,x_2|x_{1T},x_{2T}) - P_h(x_1,x_2-1|x_{1T},x_{2T}) + P_h(x_1-1,x_2-1|x_{1T},x_{2T})\}
\end{aligned}$$

$$\frac{\partial P_h}{\partial \lambda_1} = \left(\frac{1-\alpha_1^h}{1-\alpha_1}\right) \{P_h(x_1,x_2|x_{1T},x_{2T}) + P_h(x_1-1,x_2|x_{1T},x_{2T})\}$$

$$\frac{\partial P_h}{\partial \lambda_2} = \left(\frac{1-\alpha_2^h}{1-\alpha_2}\right) \{P_h(x_1,x_2|x_{1T},x_{2T}) + P_h(x_1,x_2-1|x_{1T},x_{2T})\}$$

$$\begin{aligned}
\frac{\partial P_h}{\partial \phi} &= \left(\frac{1-\alpha_1^h\alpha_2^h}{1-\alpha_1\alpha_2}\right) \{P_h(x_1,x_2|x_{1T},x_{2T}) - P_h(x_1-1,x_2|x_{1T},x_{2T}) \\
&- P_h(x_1,x_2-1|x_{1T},x_{2T}) + P_h(x_1-1,x_2-1|x_{1T},x_{2T})\}
\end{aligned}$$

In order to obtain analytical expressions for the elements that comprise the Fisher information matrix $\boldsymbol{i}^{-1}$ we follow the notation of Freeland and McCabe (2004b) and denote by $\ddot{\ell}_{\boldsymbol{\theta}}$ the second derivatives of the log-likelihood of the Poisson BINAR(1) model with respect to $\boldsymbol{\theta} = [\alpha_1, \alpha_2, \lambda_1, \lambda_2, \phi]'$:

$$\ddot{\ell}_{\boldsymbol{\theta}} = \begin{bmatrix}
\ddot{\ell}_{\alpha_1\alpha_1} & \ddot{\ell}_{\alpha_1\alpha_2} & \ddot{\ell}_{\alpha_1\lambda_1} & \ddot{\ell}_{\alpha_1\lambda_2} & \ddot{\ell}_{\alpha_1\phi} \\
 & \ddot{\ell}_{\alpha_2\alpha_2} & \ddot{\ell}_{\alpha_2\lambda_1} & \ddot{\ell}_{\alpha_2\lambda_2} & \ddot{\ell}_{\alpha_2\phi} \\
 & & \ddot{\ell}_{\lambda_1\lambda_1} & \ddot{\ell}_{\lambda_1\lambda_2} & \ddot{\ell}_{\lambda_1\phi} \\
 & & & \ddot{\ell}_{\lambda_2\lambda_2} & \ddot{\ell}_{\lambda_2\phi} \\
 & & & & \ddot{\ell}_{\phi\phi}
\end{bmatrix}$$

Through ordinary algebra it can be shown that

$$\begin{aligned}\ddot{\ell}_{\alpha_1\alpha_1} &= \frac{1}{(1-\alpha_1)^2}\sum_{t=1}^{T}\left\{\frac{2x_{1,t-1}P(x_{1t}-1,x_{2t}|x_{1,t-1}-1,x_{2,t-1})}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})} - x_{1,t-1}\right.\\ &+ \frac{x_{1,t-1}(x_{1,t-1}-1)P(x_{1t}-2,x_{2t}|x_{1,t-1}-2,x_{2,t-1})}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\\ &- \left.\left(\frac{x_{1,t-1}P(x_{1t}-1,x_{2t}|x_{1,t-1}-1,x_{2,t-1})}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\right)^2\right\}\end{aligned}$$

$$\begin{aligned}\ddot{\ell}_{\alpha_2\alpha_2} &= \frac{1}{(1-\alpha_2)^2}\sum_{t=1}^{T}\left\{\frac{2x_{2,t-1}P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1}-1)}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})} - x_{2,t-1}\right.\\ &+ \frac{x_{2,t-1}(x_{2,t-1}-1)P(x_{1t},x_{2t}-2|x_{1,t-1},x_{2,t-1}-2)}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\\ &- \left.\left(\frac{x_{2,t-1}P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1}-1)}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\right)^2\right\}\end{aligned}$$

$$\ddot{\ell}_{\lambda_1\lambda_1} = \sum_{t=1}^{T}\left\{\frac{P(x_{1t}-2,x_{2t}|x_{1,t-1},x_{2,t-1})}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})} - \left(\frac{P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\right)^2\right\}$$

$$\ddot{\ell}_{\lambda_2\lambda_2} = \sum_{t=1}^{T}\left\{\frac{P(x_{1t},x_{2t}-2|x_{1,t-1},x_{2,t-1})}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})} - \left(\frac{P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\right)^2\right\}$$

$$
\begin{aligned}
\ddot{\ell}_{\phi\phi} &= \sum_{t=1}^{T}\left\{\frac{1}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\times\right.\\
&\times\ \{2P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})-2P(x_{1t}-2,x_{2t}-1|x_{1,t-1},x_{2,t-1})\\
&-\ 2P(x_{1t}-1,x_{2t}-2|x_{1,t-1},x_{2,t-1})+P(x_{1t}-2,x_{2t}-2|x_{1,t-1},x_{2,t-1})\\
&+\ P(x_{1t}-2,x_{2t}|x_{1,t-1},x_{2,t-1})+P(x_{1t},x_{2t}-2|x_{1,t-1},x_{2,t-1})\}\\
&+\ \frac{1}{P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\times\\
&\times\ \{2P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})\\
&+\ 2P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})\\
&-\ 2P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})\\
&-\ P^2(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})-P^2(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})\\
&-\ \left.P^2(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\alpha_1\alpha_2} &= \sum_{t=1}^{T}\left\{\frac{x_{1,t-1}x_{2,t-1}}{(1-\alpha_1)(1-\alpha_2)P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\times\right.\\
&\times\ \{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t}-1,x_{2t}-1|x_{1,t-1}-1,x_{2,t-1}-1)\\
&-\ \left.P(x_{1t}-1,x_{2t}|x_{1,t-1}-1,x_{2,t-1})\ P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1}-1)\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\alpha_1\lambda_1} &= \sum_{t=1}^{T}\left\{\frac{x_{1,t-1}}{(1-\alpha_1)P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\times\right.\\
&\times\ \{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t}-2,x_{2t}|x_{1,t-1}-1,x_{2,t-1})\\
&-\ \left.P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t}-1,x_{2t}|x_{1,t-1}-1,x_{2,t-1})\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\alpha_2\lambda_2} &= \sum_{t=1}^{T}\left\{\frac{x_{2,t-1}}{(1-\alpha_2)P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\times\right.\\
&\times\ \{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t},x_{2t}-2|x_{1,t-1},x_{2,t-1}-1)\\
&-\ \left.P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1}-1)\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\alpha_1\lambda_2} &= \sum_{t=1}^{T}\left\{\frac{x_{1,t-1}}{(1-\alpha_1)P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\times\right.\\
&\times \ \{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t}-1,x_{2t}-1|x_{1,t-1}-1,x_{2,t-1})\\
&- \left. P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})P(x_{1t}-1,x_{2t}|x_{1,t-1}-1,x_{2,t-1})\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\alpha_2\lambda_1} &= \sum_{t=1}^{T}\left\{\frac{x_{2,t-1}}{(1-\alpha_2)P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\times\right.\\
&\times \ \{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1}-1)\\
&- \left. P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1}-1)\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\alpha_1\phi} &= \sum_{t=1}^{T}\left\{\frac{x_{1,t-1}}{(1-\alpha_1)}\left\{\frac{1}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\left[P(x_{1t}-2,x_{2t}-1|x_{1,t-1}-1,x_{2,t-1})\right.\right.\right.\\
&- \left. P(x_{1t}-2,x_{2t}|x_{1,t-1}-1,x_{2,t-1})-P(x_{1t}-1,x_{2t}-1|x_{1,t-1}-1,x_{2,t-1})\right]\\
&- \frac{P(x_{1t}-1,x_{2t}|x_{1,t-1}-1,x_{2,t-1})}{P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\left[P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})\right.\\
&- \left.\left.\left. P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})-P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})\right]\right\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\alpha_2\phi} &= \sum_{t=1}^{T}\left\{\frac{x_{2,t-1}}{(1-\alpha_2)}\left\{\frac{1}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\left[P(x_{1t}-1,x_{2t}-2|x_{1,t-1},x_{2,t-1}-1)\right.\right.\right.\\
&- \left. P(x_{1t},x_{2t}-2|x_{1,t-1},x_{2,t-1}-1)-P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1}-1)\right]\\
&- \frac{P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1}-1)}{P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\left[P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})\right.\\
&- \left.\left.\left. P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})-P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})\right]\right\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\lambda_1\lambda_2} &= \sum_{t=1}^{T}\left\{\frac{1}{P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})\right.\\
&- \left. P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\lambda_1\phi} &= \sum_{t=1}^{T}\left\{\frac{1}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\{P(x_{1t}-2,x_{2t}-1|x_{1,t-1},x_{2,t-1})\right.\\
&- P(x_{1t}-2,x_{2t}|x_{1,t-1},x_{2,t-1})-P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})\}\\
&- \frac{P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})}{P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\{P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})\\
&- \left.P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})-P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\ddot{\ell}_{\lambda_2\phi} &= \sum_{t=1}^{T}\left\{\frac{1}{P(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\{P(x_{1t}-1,x_{2t}-2|x_{1,t-1},x_{2,t-1})\right.\\
&- P(x_{1t},x_{2t}-2|x_{1,t-1},x_{2,t-1})-P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})\}\\
&- \frac{P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})}{P^2(x_{1t},x_{2t}|x_{1,t-1},x_{2,t-1})}\{P(x_{1t}-1,x_{2t}-1|x_{1,t-1},x_{2,t-1})\\
&- \left.P(x_{1t}-1,x_{2t}|x_{1,t-1},x_{2,t-1})-P(x_{1t},x_{2t}-1|x_{1,t-1},x_{2,t-1})\}\right\}
\end{aligned}
$$

Note that, in contrast to the univariate case, the information (as well as the scores) of the Poisson BINAR(1) model cannot be decomposed into quantities associated with each component of the model seperately. This barrier is just due to the model's structure, i.e. to its bivariate nature.

The Fisher information matrix $\boldsymbol{i}$ can then be calculated as usual:

$$
\boldsymbol{i} = -E\left[\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}|\boldsymbol{\theta}\right] = -E\left[\ddot{\ell}_{\boldsymbol{\theta}}|\boldsymbol{\theta}\right]
$$

where $\ell(\boldsymbol{\theta})$ is the log-likelihood of the Poisson BINAR(1) model.