

# Bivariate models for time series of counts: a comparison study between PBINAR models and dynamic factor models

Magda Monteiro<sup>1,3\*</sup>; Isabel Pereira<sup>2,3</sup> and Manuel G. Scotto<sup>4</sup>

<sup>1</sup> *ESTGA - Águeda School of Technology and Management, University of Aveiro, Portugal,*

<sup>2</sup> *DMat - Mathematics Department, University of Aveiro, Portugal*

<sup>3</sup> *CIDMA - Center for Research & Development in Mathematics and Applications, University of Aveiro, Portugal*

<sup>4</sup> *Departamento de Matemática and CEMAT, IST, Universidade de Lisboa, Portugal*

*Preprint submitted to Communication in Statistics - Simulation and Computation - March 19, 2019*

## Abstract

The aim of this work is to assess the modeling performance of two bivariate models for time series of counts, within the context of a forest fires analysis in two counties of Portugal. The first model is a periodic bivariate integer-valued autoregressive (PBINAR), easily interpreted due to the PINAR description of each component. The alternative model is a bivariate dynamic factor (BDF) that has a flexible structure, with the dynamics described through the mean value of each component that is a function of latent factors. The results reveal that BDF model exhibits a better ability to

---

\*Magda Monteiro, ESTGA, Universidade de Aveiro. Rua Comandante Pinho e Freitas, n. 28 3750-127 Águeda, Portugal. Email: msvm@ua.pt.

capture the dependence structure.

*Keywords:* bivariate count processes, INAR processes, dynamic factor models, thinning operation, simulation-based estimation methods.

# 1 Introduction

While in the last three decades the development of univariate count time series models has been a research topic of active interest (Al-Osh and Alzaid, 1987, 1993; McKenzie, 2003; Ferland et al., 2006; Fokianos, 2012; Davis and Liu, 2016), the research activity in modeling multivariate time series of counts is now beginning to grow (Jung et al., 2011; Pedeli and Karlis, 2011; Scotto et al., 2015; Davis et al., 2016). The existing models for multivariate time series of counts can be classified into two main groups, namely the parameter-driven (PD) class in which serial correlation is introduced via some latent dynamic process (MacDonald and Zucchini, 1997; Kedem and Fokianos, 2002), and the observation-driven (OD) class that uses the information collected from previous observations to induce time correlation (Fokianos et al., 2009; Ferland et al., 2006; Khan et al., 2016). Models belonging to the OD class include the integer-valued autoregressive moving average (INARMA) models based on thinning operators (e.g., Scotto et al., 2015; Weiß, 2018), and the integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) models (e.g., Tjøstheim, 2012). In the class of INGARCH models, the conditional distribution of observed counts given past outcomes or latent process is usually assumed to come from some well-known discrete distributions, such as Poisson, negative binomial (NB), generalized Poisson (GP), double Poisson (DP) and ConwayMaxwellPoisson (COM-Poisson) distributions (Ferland et al., 2006; Zhu, 2011a,b; Fokianos et al., 2009; Zhu, 2012). The class of univariate PD extends the generalized linear models (GLM) by incorporating into the conditional mean function a latent process which evolves independently of the past observed counts. The autocorrelation as well as over-dispersion are introduced in the model via such process (see Fokianos (2012) and references therein).

Turning now to bivariate and multivariate count time series, Pedeli and Karlis (2011) introduced the BINAR(1) model in which the cross-correlation is induced through the innovation process by assuming either a bivariate Poisson distribution or a bivariate negative binomial distribution. The periodic version of Pedeli and Karlis' model, referred to as PBINAR, was first proposed by Monteiro et al. (2015). A full BINAR model with bivariate

Poisson innovations was introduced and studied by Pedeli and Karlis (2013). Furthermore, multivariate INGARCH-type models have also been proposed in the literature. Heinen and Rengifo (2007) introduced the so-called multivariate autoregressive conditional double Poisson in which it is assumed that, conditionally on past observations, the means follow an autoregressive vector using copulas to introduce contemporaneous correlation. Liu (2012) proposed a bivariate Poisson integer-valued GARCH (BINGARCH) model which has the capability of modeling the serial dependence between two time series of counts. More recently, Bracher and Held (2017) introduced a class of periodic INGARCH-type models.

Within the PD class, Jorgensen et al. (1999) proposed a state-space multivariate Poisson model while Wedel et al. (2003) developed a general class of models with analytic factors, for the analysis of truncated multivariate count data. An autoregressive structure for conditional means was proposed by Held et al. (2005) and applied to the multivariate infectious disease surveillance data. Jung et al. (2011) proposed a model with latent factor for multivariate count time series which generalizes previous works (Jorgensen et al., 1999; Wedel et al., 2003) and applied it to the number of trades in 5-min intervals for five New York Stock Exchange (NYSE) stocks from two industrial sectors. Serhiyenko et al. (2015) developed zero-inflated Poisson models for multivariate time series of counts. Finally, Ravishanker et al. (2015) studied finite mixtures of multivariate Poisson time series.

The rest of the paper is organized as follows: Section 2 provides models description and the parameter estimation approach used in each model. Section 3 presents the comparative analysis in the context of monthly number of fires modeling with the discussion of the models adequacy. Finally, some concluding remarks are given in Section 4.

## 2 Models specification

In this section, the PBINAR model and the periodic BDF model, both with period  $S$ , are introduced and explained in detail. The PBINAR model exhibits a structure of easy interpretation, allowing cross-correlation only through the bivariate innovation process. Fur-

thermore, it is an over-parameterized model containing  $5 \times S$  parameters. The BDF model presented can be viewed as an alternative model, having the advantage of containing far less parameters.

## 2.1 PBINAR(1) model

The periodic bivariate integer-valued autoregressive of order one, say  $(\mathbf{Y}_t)$ , proposed by Monteiro et al. (2015) is an extension of the bivariate model introduced by Pedeli and Karlis (2011), assuming periodic time-varying parameters and periodic bivariate sequences of innovations, where  $\mathbf{Y}_t$  admits the representation

$$\mathbf{Y}_t = \mathbf{A}_t \circ \mathbf{Y}_{t-1} + \mathbf{Z}_t \equiv \begin{bmatrix} \phi_{1,t} & 0 \\ 0 & \phi_{2,t} \end{bmatrix} \circ \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix}, \quad t \in \mathbb{Z} \quad (1)$$

with  $\phi_{j,t} = \alpha_{j,i}$ , for  $t = i + kT$  ( $i = 1, \dots, T$ ),  $j = 1, 2$ , and  $k \in \mathbb{N}_0$ . Usually, the process  $(\mathbf{Y}_t)$  is observed during  $N$  years in periods of  $S$  seasons. Thus, in order to facilitate models' interpretation, we will denote throughout  $\mathbf{Y}_t \equiv \mathbf{Y}_{s,n}$  for  $t = 1, 2, \dots, T$ ,  $n = 1, 2, \dots, N$  and  $s = 1, 2, \dots, S$ , where  $n$  is the year associated with time  $t$  and  $s$  is the respective season. Note that, when  $t$  corresponds to the first season of the year  $n$  then the previous time, which corresponds to the season  $S$  of the year  $n - 1$ , will be denoted as season 0 of the year  $n$ , i.e.,  $\mathbf{Y}_{0,n} \equiv \mathbf{Y}_{S,n-1}$ . Therefore, the model in (1) can be rewritten as

$$\mathbf{Y}_{s,n} = \mathbf{A}_{s,n} \circ \mathbf{Y}_{s-1,n} + \mathbf{Z}_{s,n} = \begin{bmatrix} \alpha_{s,1} & 0 \\ 0 & \alpha_{s,2} \end{bmatrix} \circ \begin{bmatrix} Y_{s-1,n;1} \\ Y_{s-1,n;2} \end{bmatrix} + \begin{bmatrix} Z_{s,n;1} \\ Z_{s,n;2} \end{bmatrix}, \quad (2)$$

where  $s = 1, \dots, S$ , emphasizing the periodic nature of the model. Within this framework, the (binomial) thinning operator “ $\circ$ ” is defined as

$$\alpha_{s,j} \circ Y_{s-1,n;j} \stackrel{d}{=} \sum_{m=1}^{Y_{s-1,n;j}} U_{m;s,j},$$

where  $(U_{m;s,j})$  is a periodic sequence of independent Bernoulli random variables with success probability  $P(U_{m;s,j} = 1) = \alpha_{s,j} \in (0; 1)$ . Attending to the properties of the binomial

thinning operator, each component of the model can be written as

$$Y_{s,n;j} = \alpha_{s,j} \circ Y_{s-1,n;j} + Z_{s,n;j}, \quad j = 1, 2. \quad (3)$$

Moreover, it is assumed that  $\mathbf{Z}_{s,n} = (Z_{s,n;1}, Z_{s,n;2})$  is a periodic sequence of independent random vectors with mean  $E[\mathbf{Z}_{s,n}] = \boldsymbol{\lambda}_s := [\lambda_{s,1} \ \lambda_{s,2}]'$  and covariance matrix  $\boldsymbol{\Sigma}_s$  with  $\sigma_{j,s}^2 = v_{s,j} \lambda_{s,j}$ , for  $v_{s,j} > 0$ , and  $\sigma_{12,s} =: \varphi_s$ . Furthermore,  $Z_{s,n;j}$  is independent of  $Y_{s-1,n;j}$  and of  $\alpha_{s,j} \circ Y_{s-1,n;j}$ . The distributions under consideration for the innovations bivariate process are the bivariate Poisson,  $\mathbf{Z}_{s,n} \sim \text{BPoi}(\lambda_{s,1}, \lambda_{s,2}, \varphi_s)$  with  $v_{s,j} = 1$ , and the bivariate negative binomial  $\mathbf{Z}_{s,n} \sim \text{BNegBin}(\lambda_{s,1}, \lambda_{s,2}, \beta_s)$  with  $v_{s,j} = 1 + \beta_s \lambda_{s,j}$ . Further properties of this model can be found in Monteiro et al. (2015).

The most common approach to estimate models' parameters is the conditional maximum likelihood (CML). Consider  $N$  complete years  $(\mathbf{y}_{1,1}, \mathbf{y}_{2,1}, \dots, \mathbf{y}_{S,N})$  and  $\boldsymbol{\theta}$  the vector of unknown parameters. Without loss of generality, it is assumed that  $\mathbf{Y}_0 = \mathbf{y}_0$ . Note that the transition probabilities takes the form

$$\begin{aligned} p_s(\mathbf{b}|\mathbf{a}) &:= P(\mathbf{Y}_{s,n} = \mathbf{b} | \mathbf{Y}_{s-1,n} = \mathbf{a}) \\ &= \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_2} P(\alpha_{1,s} \circ Y_{s-1,n;1} = m_1, \alpha_{2,s} \circ Y_{s-1,n;2} = m_2 | \mathbf{Y}_{s-1,n} = \mathbf{a}) \times \\ &\quad \times P(Z_{s,n;1} = b_1 - m_1, Z_{s,n;2} = b_2 - m_2), \end{aligned}$$

where  $\mathbf{b} := [b_1 \ b_2]'$ ,  $\mathbf{a} := [a_1 \ a_2]'$ ,  $M_j := \min(a_j, b_j)$ ,  $j = 1, 2$ . Hence, the CML-estimator  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  is obtained by maximizing the conditional log-likelihood function

$$\ell(\boldsymbol{\theta}) := \ln(L(\boldsymbol{\theta})) = \sum_{n=1}^N \sum_{s=1}^S \ln(p_s(\mathbf{y}_{s,n} | \mathbf{y}_{s-1,n})). \quad (4)$$

For periodic bivariate Poisson innovations the expression  $p_s(\mathbf{b}|\mathbf{a})$  becomes

$$p_s(\mathbf{b}|\mathbf{a}) = \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_2} \sum_{l=0}^L \frac{\varphi_s^l}{l!} e^{-(\lambda_{s,1} + \lambda_{s,2} - \varphi_s)} \prod_{j=1}^2 \left( C_{m_j}^{a_j} \alpha_{s,j}^{m_j} (1 - \alpha_{s,j})^{a_j - m_j} \times \frac{(\lambda_{s,j} - \varphi_s)^{b_j - m_j - l}}{(b_j - m_j - l)!} \right)$$

with  $\lambda_{i,j} > 0$ ,  $j = 1, 2$ ,  $\varphi_s \in [0, \min(\lambda_{1,s}, \lambda_{2,s})]$ , and  $L := \min(b_1 - m_1, b_2 - m_2)$ .

In the case of periodic bivariate negative binomial innovations case,  $p_s(\mathbf{b}|\mathbf{a})$  is given by

$$p_s(\mathbf{b}|\mathbf{a}) = \left(\frac{\beta_s^{-1}}{\eta_s}\right)^{\beta_s^{-1}} \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_2} \prod_{j=1}^2 \left[ \frac{\Gamma(\beta_s^{-1} + a_1 - m_1 + a_2 - m_2)}{\Gamma(\beta_s^{-1})\Gamma(a_j - m_j + 1)} C_{m_j}^{a_j} \alpha_{s,j}^{m_j} (1 - \alpha_{s,j})^{a_j - m_j} \left(\frac{\lambda_{s,j}}{\eta_s}\right)^{b_j - m_j} \right],$$

with  $\lambda_{s,j}, \beta_s > 0$ ,  $j = 1, 2$ ,  $\eta_s = \lambda_{s,1} + \lambda_{s,2} + \beta_s^{-1}$ .

## 2.2 Bivariate dynamic factor model

In this section, the bivariate version of the dynamic factor (BDF) model of Jung et al. (2011) with the inclusion of a periodic component is presented. Consider a bivariate vector of counts  $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2})$  observed at time  $t$  ( $t = 1, \dots, T$ ), where the dynamics is introduced at the level of the latent factors. Accordingly, counts are assumed to be conditionally independently distributed with an appropriate discrete distribution, with means  $\lambda_{t,j}$  considered as latent random variables. Furthermore, it will be assumed that the logarithm of the mean vector  $\boldsymbol{\lambda}_t = (\lambda_{t,1}, \lambda_{t,2})$  admits the form

$$\ln(\boldsymbol{\lambda}_t) = \boldsymbol{\mu} + \mathbf{S}_t \boldsymbol{\beta} + \boldsymbol{\Gamma} \mathbf{X}_t, \quad (5)$$

where  $\boldsymbol{\mu}$  is an intercept to which the seasonal contribution is added (through  $\mathbf{S}_t \boldsymbol{\beta}$ ) where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{12})$ , and  $\mathbf{S}_t$  is a  $2 \times 12$  design matrix with  $S_{t,1} = S_{t,2}$ , and  $\mathbf{X}_t$  is a 3-dimensional vector of latent random factors which are assumed to be independent from each other, and  $\boldsymbol{\Gamma}$  is the  $(2 \times 3)$  loading factors matrix. The choice of the  $\ln(\cdot)$  function is justified by the fact that it guarantees the positiveness of the mean vector without imposing further restrictions on the parameters included in the linear function.

Turning now to the the bivariate context, it will be assumed that  $\mathbf{X}_t = (Z_t, W_{t,1}, W_{t,2})$ , where  $Z_t$  represents a common latent factor which incorporates temporal correlation and cross-correlation and two specific factors  $(W_{t,1}, W_{t,2})$  related with each component which also incorporate temporal correlation. Furthermore, the matrix of loading factors is assumed to be of the form

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 & 1 & 0 \\ \gamma_2 & 0 & 1 \end{bmatrix}.$$

Serial and cross-correlation in the counts are introduced through the structure of the factors components. Following Jung et al. (2011), it will be considered that the components of  $\mathbf{X}_t$  follow independent Gaussian AR(1) processes, that is  $\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t$  with  $\mathbf{\Phi} = \text{diag}(\phi, \rho_1, \rho_2)$  and  $\boldsymbol{\epsilon}_t$  follows a multivariate Gaussian distribution with zero-mean vector and covariance-matrix  $\boldsymbol{\Sigma}_\epsilon = \text{diag}(\sigma_\epsilon^2, \sigma_1^2, \sigma_2^2)$ . In order to ensure the stationarity of the factors, it is assumed that  $|\phi| < 1$ ,  $|\rho_j| < 1$ , for  $j = 1, 2$ .

The BDF model is also a state-space model also known as partial observed Markov processes (POMP), being  $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2})$  the observation process and  $\mathbf{X}_t = (Z_t, W_{t,1}, W_{t,2})$  the state vector at time  $t$ . For this type of non-linear and non-Gaussian state-space models, the analytical study of the likelihood function is, in general, intractable since it is necessary to use high-dimension integration. Note that for the current model, by setting  $\mathbf{Y}_{1:T} = (\mathbf{Y}_1, \dots, \mathbf{Y}_T)$  and  $\mathbf{X}_{0:T} = (\mathbf{X}_0, \dots, \mathbf{X}_T)$ ,

$$f_{\mathbf{X}_{0:T}, \mathbf{Y}_{1:T}}(\mathbf{x}_{0:T}, \mathbf{y}_{1:T}; \boldsymbol{\theta}) = f_{\mathbf{X}_0}(\mathbf{x}_0; \boldsymbol{\theta}) \prod_{t=1}^T f_{\mathbf{X}_t | \mathbf{X}_{t-1}}(\mathbf{x}_t | \mathbf{x}_{t-1}; \boldsymbol{\theta}) f_{\mathbf{Y}_t | \mathbf{X}_t}(\mathbf{y}_t | \mathbf{x}_t; \boldsymbol{\theta}), \quad (6)$$

where  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \mathbf{\Phi}, \boldsymbol{\Sigma}_\epsilon)$ ,  $\mathbf{X}_t | \mathbf{X}_{t-1} \sim \mathcal{N}(\mathbf{\Phi}\mathbf{X}_{t-1}, \boldsymbol{\Sigma})$ ,  $\mathbf{X}_0 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$  and

$$\boldsymbol{\Sigma}_0 = \text{diag} \left( \frac{\sigma_\epsilon^2}{1 - \phi^2}, \frac{\sigma_1^2}{1 - \rho_1^2}, \frac{\sigma_2^2}{1 - \rho_2^2} \right),$$

being the marginal probability function of  $\mathbf{Y}_{1:T}$  obtained by multiple integration as

$$f_{\mathbf{Y}_{1:T}}(\mathbf{y}_{1:T}; \boldsymbol{\theta}) = \int f_{\mathbf{X}_{0:T}, \mathbf{Y}_{1:T}}(\mathbf{x}_{0:T}, \mathbf{y}_{1:T}; \boldsymbol{\theta}) d\mathbf{x}_{0:T}.$$

Note that, for  $f_{\mathbf{Y}_t | \mathbf{X}_t}(\mathbf{y}_t | \mathbf{x}_t; \boldsymbol{\theta})$  in (6) we can either consider the Poisson distribution or a more flexible distribution such as the negative binomial distribution, i.e.,

$$f_{\mathbf{Y}_t | \mathbf{X}_t}(\mathbf{y}_t | \mathbf{x}_t; \boldsymbol{\theta}) = \begin{cases} \prod_{j=1}^2 e^{-\lambda_{t,j}} \frac{\lambda_{t,j}^{y_{t,j}}}{y_{t,j}!}, & \text{Poisson} \\ \prod_{j=1}^2 \frac{\Gamma(y_{t,j} + \frac{1}{v_j^2})}{\Gamma(\frac{1}{v_j^2}) \Gamma(y_{t,j} + 1)} \left( \frac{1}{1 + v_j^2 \lambda_{t,j}} \right)^{\frac{1}{v_j^2}} \left( \frac{\lambda_{t,j}}{\lambda_{t,j} + \frac{1}{v_j^2}} \right)^{y_{t,j}}, & \text{Neg. Bin.} \end{cases}.$$

The difficulty in obtaining a closed-form expression for the likelihood lead to several developments of simulation-based methods in order to perform likelihood-based inferences (Liesenfeld and Richard, 2003; Richard and Zhang, 2007; Scharth and Kohn, 2016). Such algorithm is based on the Sequential Monte Carlo (SMC) algorithm (Andrieu et al., 2010) also known as particle filter. This procedure consists in extending the model of interest by allowing its (time-invariant) parameters to become artificially random, following a random walk in time ( $\theta_t$ ). From an initial  $\theta_0$  and successively diminishing the parameters artificial random perturbations the algorithm generates a sequence of updated parameter estimates  $\theta_1, \theta_2, \dots$ , converging to the maximum likelihood estimates, under some regularity conditions on the likelihood surface and on the algorithmic parameters (for further details see Ionides et al., 2011, 2006). Iterated filtering are implemented in this work using the software package POMP (Partially Observed Markov Processes) described in King et al. (2016, 2017)) and available at the R statistical computing environment (R Development Core Team, 2008). In order to apply the algorithm, it is necessary to define the number of iterations, the initial magnitude of the parameter perturbations and the correspondent decrease rate from one iteration to the next one, and the number of particles to use in the embedded SMC algorithm. In practice, the convergence of the IF algorithm is assessed via convergence diagnostic plots, since the conditions on the likelihood surface are rather technical (Bretó, 2014).

### **3 PBINAR and BDF models: a comparative study for the analysis of bivariate time series of fire activity**

In this section, monthly records of number of forest fires in O. Bairro and Vagos counties (Aveiro's district), collected from 1985 to 2014, are analyzed through PBINAR and BDF models. The bivariate time series were obtained from the Instituto de Conservação da Natureza e das Florestas (ICNF, 1016). Figure 1 displays both time series. Note that both series exhibit a periodic structure in the sample autocorrelation function and in the sample cross-correlation (see Figure 2).

(Figures 1 and 2 about here)

The PBINAR (1) model of period 12 is initially fitted to the pair of bivariate time series by considering (i) a periodic bivariate Poisson (BPois.) distribution for the innovation process, and (ii) a bivariate periodic negative binomial (BNBin) as defined in (6). Table 1 resumes the analysis of Pearson residuals for both models. In both cases the sample means of Pearson residuals are close to zero although in the PBINAR model the sample variance differs quite significantly from one. Furthermore, the Ljung-Box test rejects the absence of autocorrelation at different lags. Hence, such model does not accurately captures the dependence structure exhibited by both time series.

In fitting the BDF model with innovations being either Poisson or negative binomially distributed, the IF algorithm was primary applied to two hundred initial candidates of the parameters from which were selected three initial parameter vectors that presented the highest values of the likelihood estimates. This search was computational time-consuming, using parallelization of the cores of a i7-4500U intel core computer with 8 GB of RAM. For these chosen three initial parameter vector values, the IF algorithm was applied with the following algorithmic parameters: 300 iterations, an exponential decay of perturbations of  $\alpha = 0.70$  and 20000 particles. Furthermore, 70,000 particles are used to obtain the point estimates of the log-likelihood  $\ell(\cdot)$ , which result from taking the average of ten likelihood evaluations from which we calculate the Monte Carlo standard error. Parameters standard errors are obtained via profile likelihoods (see Ionides et al. (2017)).

The convergence of the maximum likelihood estimates is assessed through convergence diagnostic plots (see Figure 3 and Figure 4 for the Poisson case in the appendix). Such plots suggest an appropriate choice of the algorithmic parameters. Moreover, having into account the first search and also Figure 3 the IF method better identifies the parameters associated to the latent factors than the parameters in the  $\boldsymbol{\mu}$  and  $\boldsymbol{\beta}$  vectors, since plausible likelihoods are found with widely differing parameter values. Furthermore, Pearson residuals shows that both sample means and sample variances are close to zero and one, respectively. Additionally, the Pearson residuals did not presented significant autocorrelation (see Table

1).

(Table 1 about here)

In order to assess the adequacy of the distributional assumption, the probability integral transform (PIT) histogram for a discrete cumulative predictive distribution proposed by Czado et al. (2009) is computed (Figure 5). In PIT histograms approximate 95 % confidence intervals, obtained from a standard  $\chi^2$  goodness-of-fit test (being the null hypothesis that the  $J = 10$  bins of the histogram are drawn from a uniform distribution) were incorporated as in Jung et al. (2015). Figure 5 also shows cumulative predictive distribution against the uniform distribution.

(Figure 5 about here)

Note that, for the Poisson distribution the  $p$ -values of the  $\chi^2$  goodness-of-fit test do not reject, at any conventional significance level, the null hypotheses of uniform PIT histograms for both components of the series. On the contrary, these hypotheses are rejected for the negative binomial distribution, leading to conclude that Poisson distributional assumption is more suitable than the negative binomial distribution. From visual inspection of the PIT histograms it can be seen that for the Poisson case, with the exception of the last bin, the bins lie inside of the confidence lines, while in the negative binomial case, only a few bins are within such range due to the U-shape of the PIT histograms. Furthermore, the deviations from the identity function on the plots of the discrete cumulative predictive distribution against  $u \in [0; 1]$  become obvious in the case of the negative binomial.

Having into account that in most real case situations the negative binomial distribution outperforms the Poisson distribution, the parametric resampling procedure proposed by Tsay (1992) was applied to better understand why the results in this situation point in the opposite direction. To this end, 5000 realizations of each model (Poisson and negative binomial) were simulated being the true (unknown) parameters replaced by their corresponding point estimates obtained by the IF algorithm, in order to assess the overall reproducibility of the fitted models and also to check the autocorrelation and the cross-correlation of the processes.

For each time  $t$ , the 10% and 90% quantiles of the 5000 realizations were computed to confront with the actual data (see Figure 6). Although data generally lie inside the bounds of the acceptance envelope for both models, the Poisson model exhibits 90% quantile boundaries considerably lower than those for the negative binomial model, which helps to justify the previous results. The resampling procedure was also applied to the autocorrelation and cross-correlation function and both models generally capture the structure of these functions with a slight advantage of the Poisson model.

Taking into account that Poisson distributional assumption provides a better fit, in Table 2 are presented the parameters' point estimates and also the log-likelihood estimate (complete model). Note that, point estimates associated with the specific factor of O. Bairro are not significantly different from zero, pointing to a simplification of the model by dropping the factor  $W_{t,1}$  in O.Bairro component.

(Table 2 about here)

For the simplified model, the analysis of the convergence of the IF algorithm, the Pearson residuals analysis as well as the Poisson distribution adequacy analysis was also conducted. Parameter point estimates and the estimate of the log-likelihood are presented in Table 2. The log-likelihood estimate is close to the obtained in the complete model with the advantage of incorporating lesser parameters.

## 4 Conclusions

In this paper, a comparative study between two different types of bivariate time series models suitable for modeling periodic time series of counts has been presented. The PBINAR model is the periodic counterpart of the observation-driven BINAR model of Pedeli and Karlis (2011) in which the cross-correlation is only introduced through the periodic bivariate innovation process. In terms of estimation, it can be applied standard estimation methods as the conditional maximum likelihood with the application of standard numerical optimization packages. Although has, on the one hand, a large number of parameters to be estimated

which makes difficult its generalization to higher order vectors and, on the other hand, the inability, for the considered case, to capture all the existing correlation in the bivariate data under study. The alternative BDF model presents a more flexible structure that incorporates temporal and cross-correlation through the latent factors. This model, being a non-linear and non-Gaussian state space model, poses some challenges in terms of parameter estimation, that are overcome using simulation-based methods that include the IF method used in this work. This model seems to have captured the dynamics of the underlying process of the bivariate count data under analysis and has the advantage of being easier to extend to vectors of higher orders and also is more parsimonious in terms of the number of parameters in the model. For the monthly number of fires in the counties of O.Bairro and Vagos, the bivariate factor model with Poisson distribution revealed to be adequate since it was able to capture the main features of the bivariate count data. Having into account that the district of Aveiro has 19 counties, we can build a BDF model that describes the phenomenon under study for the entire district using common counties characteristics.

The IF method reveals that the BDF model might have some parameter identifiability issues associated to  $\boldsymbol{\mu}$  and  $\boldsymbol{\beta}$  vectors since there are several different values of these parameters with close values of log-likelihood estimates. Before using BDF model in a forecasting step or extending it to higher dimensions it is important to clarify the existence of this identifiability problem. The comparison with other estimation methods such as efficient importance sampling described in Richard and Zhang (2007) or a Bayesian approach, will allow to clarify this issue. Alternatively, it can also be included additional assumptions, scientifically reasonable, so that the vectors  $\boldsymbol{\mu}$  and  $\boldsymbol{\beta}$  can be better identified or, alternatively, the inclusion of covariates containing periodic features, such as monthly temperature and humidity. This will be a topic for future research.

As a final note we would like to stress that, although in this work the PINAR and the BDF were considered for comparison purposes, such comparison is certainly not limited to these specific families of models, and can be enlarged for other competitors. An instance of

that is the bivariate integer-valued GARCH model proposed by Cui and Zhu (2018), with the necessary adaptations to the periodic case, can be an alternative since it has a flexible cross-correlation structure allowing for the presence of negative correlations.

## Acknowledgments

The authors thank the referee for carefully reading the article and for the comments, which greatly improved the article.

We acknowledge the financial support of Portuguese Foundation for Science and Technology ("FCT– Fundação para a Ciência e a Tecnologia"), through the CIDMA - Center for Research and Development in Mathematics and Applications, within project UID/MAT/04106/2019, and through the project UID/Multi/04621/2019 of CEMAT/IST-ID, Center for Computational and Stochastic Mathematics, Instituto Superior Técnico, University of Lisbon.

## References

- Al-Osh, M. and Alzaid, A. (1987). First-order integer-valued autoregressive (inar(1)) process. *J. Time Ser. Analysis*, 8:261–275.
- Al-Osh, M. and Alzaid, A. (1993). Some autoregressive moving average process with generalized poisson marginal distribution. *Ann. Inst. Statist. Math.*, 45:223–232.
- Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society B*, 72(3):269–342.
- Bracher, J. and Held, L. (2017). Periodically stationary multivariate autoregressive models. Technical report, The institution that published. Available at.
- Bretó, C. (2014). On idiosyncratic stochasticity of financial leverage effects. *Statistics and Probability Letters*, 91:20–26.

- Cui, Y. and Zhu, F. (2018). A new bivariate integer-valued garch model allowing for negative cross-correlation. *Test*, 27 (2):428–452.
- Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65:1254–1261.
- Davis, R., Holan, S., R., L., and Ravishanker, N. (2016). *Handbook of Discrete-valued Time Series*. Chapman and Hall.
- Davis, R. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, 26:1673–1707.
- Ferland, R., Latour, A., and D., O. (2006). Integer-valued garch process. *Journal of Time Series Analysis*, 27(6):923–942.
- Fokianos, K. (2012). Count time series models. In Subba Rao, T., Subba Rao, S., and Rao, C. R., editors, *Handbook in Statistics. Time Series Analysis: Methods and Applications*, chapter 12, pages 923–942. Elsevier, 1 edition.
- Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439.
- Heinen, A. and Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas. *J Empir Fin*, 14:564–583.
- Held, L., Hohle, M., and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, 5:187–199.
- ICNF (21016). Instituto de Conservação da Natureza e das Florestas. <http://www.icnf.pt/portal/florestas/dfci/inc/estat-sgif#tip>. last checked: 2016–04–08.
- Ionides, E., Bretó, C., J., P., Smith, R., and King, A. (2017). Monte carlo profile confidence intervals for dynamic systems. *J. R. Soc. Interface*, 14.

- Ionides, E. L., Bhadra, A., Atchad, Y., and King, A. (2011). Iterated filtering. *Ann. Statist.*, 39(3):1776–1802.
- Ionides, E. L., Breto, C., and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the USA*, 103:18438–18443.
- Jorgensen, B., Lundbye-Christense, S., Song, P.-K., and Sun, L. (1999). State space model for multivariate longitudinal count data. *Biometrika*, 86(1):169–181.
- Jung, R., Liesenfeld, R., and Richard, J. (2011). Dynamic factor models for multivariate count data: An application to stock-market trading activity. *Journal of Business & Economic Statistics*, 29(1):73–85.
- Jung, R., McCabe, B., and Tremaayne, A. (2015). Model validation and diagnostics. In Davis, R. A., Holan, H., Lund, R., and Ravishanker, N., editors, *Handbook of Discrete-Valued Time Series*, pages 189–218. Chapman & Hall.
- Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*. John Wiley & Sons.
- Khan, N., Sunecher, Y., and Jowaheer, V. (2016). Modelling a non-stationary binar(1) poisson process. *Journal of Statistical Computation and Simulation*, 86(15):3106–3126.
- King, A. A., Ionides, E. L., Bretó, C. M., Ellner, S. P., Ferrari, M. J., Kendall, B. E., Lavine, M., Nguyen, D., Reuman, D. C., Wearing, H., and Wood, S. N. (2017). *pomp: Statistical Inference for Partially Observed Markov Processes*. R package, version 1.16.
- King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed markov processes via the R package pomp. *Journal of Statistical Software*, 69(12):1–43.
- Liesenfeld, R. and Richard, J.-F. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance*, 10:505–531.
- Liu, H. (2012). *Some models for time series of counts*. PhD thesis, Columbia University.

- MacDonald, I. and Zucchini, W. (1997). *Hidden Markov Models and Other Models for Discrete-valued Time Series*. Chapman & Hall.
- McKenzie, E. (2003). Discrete variate time series. In Shanbhag, D. N. and Rao, C. R., editors, *Handbook of Statistics on Stochastic Processes: Modelling and Simulation*, volume 21, pages 573–606. Elsevier.
- Monteiro, M., Pereira, I., and Scotto, M. G. (2015). A periodic bivariate integer-valued autoregressive model. In Bourguignon, J.-P., Jelstch, R., Pinto, A. A., and Viana, M., editors, *Dynamics, Games and Science, CIM Series in Mathematical Sciences 1*, pages 455–477, Lisbon. Springer.
- Pedeli, X. and Karlis, D. (2011). A bivariate inar(1) process with application. *Statistical Modelling*, 11(4):325–349.
- Pedeli, X. and Karlis, D. (2013). Some properties of multivariate inar(1) processes. *Computational Statistics and Data Analysis*, 67:213–225.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Ravishanker, N., Venkatesan, R., and Hu, S. (2015). Dynamic models for time series of counts with a marketing application. In Davis, R. A., Holan, H., Lund, R., and Ravishanker, N., editors, *Handbook of discrete-valued time series*, pages 423–445. Chapman & Hall.
- Richard, J.-F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141:1385–1411.
- Scharth, M. and Kohn, R. (2016). Particle efficient importance sampling. *Journal of Econometrics*, 190:133–147.
- Scotto, M., Weiss, C., and Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Stat. Model.*, 15:590–618.

- Serhiyenko, V., Ravishanker, N., and Venkatesan, R. (2015). Approximate bayesian estimation for multivariate count time series models. In Choudhary, P. K., Nagaraja, C. H., and Ng, H. K. T., editors, *Ordered Data Analysis, Modeling and Health Research Methods*, chapter 9, pages 155–167. Springer.
- Tjøstheim, D. (2012). Some recent theory for autoregressive count time series. *Test*, 21:413–438.
- Tsay, R. (1992). Model checking via parametric bootstraps in time series analysis. *Applied Statistics*, 41:1–15.
- Wedel, M., Bockenholt, U., and Wagner, A. K. (2003). Factor models for multivariate count data. *Journal of Multivariate Analysis*, 87:356–369.
- Weiß, C. (2018). *An Introduction to Discrete-valued Time Series*. John Wiley & Sons.
- Zhu, F. (2011a). Modeling time series of counts with com-poisson ingarch models. *Mathematical and Computer Modelling*, 56:191–203.
- Zhu, F. (2011b). A negative binomial integer-valued garch model. *Journal of Time Series Analysis*, 32:54–67.
- Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *Journal of Mathematical Analysis and Applications*, 389:58–71.

## A Convergence plots

(plots convergence, Figures 3 and 4 about here )

Table 1: Pearson Residual summary for models under consideration

Model	Distr.	County	Mean	Var.	Ljung-Box Test (p-value)			
					$Q_1$	$Q_5$	$Q_{10}$	$Q_{15}$
PBINAR	BPois.	O. Bairro	-0.018	2.844	0.024	0.003	0.003	0.004
		Vagos	-0.020	3.298	0.01	0.0015	0.006	0.02
	BNBin	O. Bairro	-0.001	1.608	0.039	0.0068	0.004	0.015
		Vagos	-0.004	1.982	0.044	0.004	0.014	0.062
DF	Poisson	O. Bairro	0.023	1.128	0.607	0.850	0.926	0.990
		Vagos	0.009	1.368	0.628	0.042	0.153	0.445
	NBin.	O. Bairro	0.002	1.02	0.654	0.831	0.890	0.983
		Vagos	-0.011	1.21	0.525	0.045	0.148	0.449

Table 2: Parameter estimates and log-likelihood ( $\ell$ ) estimates and standard errors (in parenthesis) for bivariate DF model with Poisson distribution.

	Complete Model				Simple Model			
	$\mu_j$	$\gamma_j$	$\rho_j$	$\sigma_j$	$\mu_j$	$\gamma_j$	$\rho_j$	$\sigma_j$
O.Bairro	-0.874 (0.130)	1 –	0.627 (0.333)	0.069 (0.179)	-1.003 (0.133)	1 –	– –	– –
Vagos	-0.827 (0.131)	0.845 (0.090)	0.508 (0.075)	0.439 (0.047)	-0.955 (0.135)	0.869 (0.085)	0.500 (0.071)	0.458 (0.041)
	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
	-1.961 (0.152)	-0.451 (0.135)	0.497 (0.131)	0.280 (0.129)	-1.659 (0.153)	-0.350 (0.137)	0.529 (0.133)	0.314 (0.124)
Seasonal Param.	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
	0.615 (0.124)	1.676 (0.124)	2.278 (0.123)	2.528 (0.118)	0.657 (0.117)	1.692 (0.129)	2.373 (0.129)	2.717 (0.122)
	$\beta_9$	$\beta_{10}$	$\beta_{11}$	$\beta_{12}$	$\beta_9$	$\beta_{10}$	$\beta_{11}$	$\beta_{12}$
	2.051 (0.120)	0.960 (0.104)	-0.677 (0.114)	-2.090 (0.154)	2.221 (0.100)	1.169 (0.102)	-0.645 (0.123)	-1.978 (0.150)
Common Factor Param.	$\phi$	$\sigma_\epsilon$	log lik		$\phi$	$\sigma_\epsilon$	log lik	
	0.704 (0.036)	0.805 (0.049)	-1132.640 (0.055)		0.718 (0.033)	0.827 (0.048)	-1132.082 (0.048)	

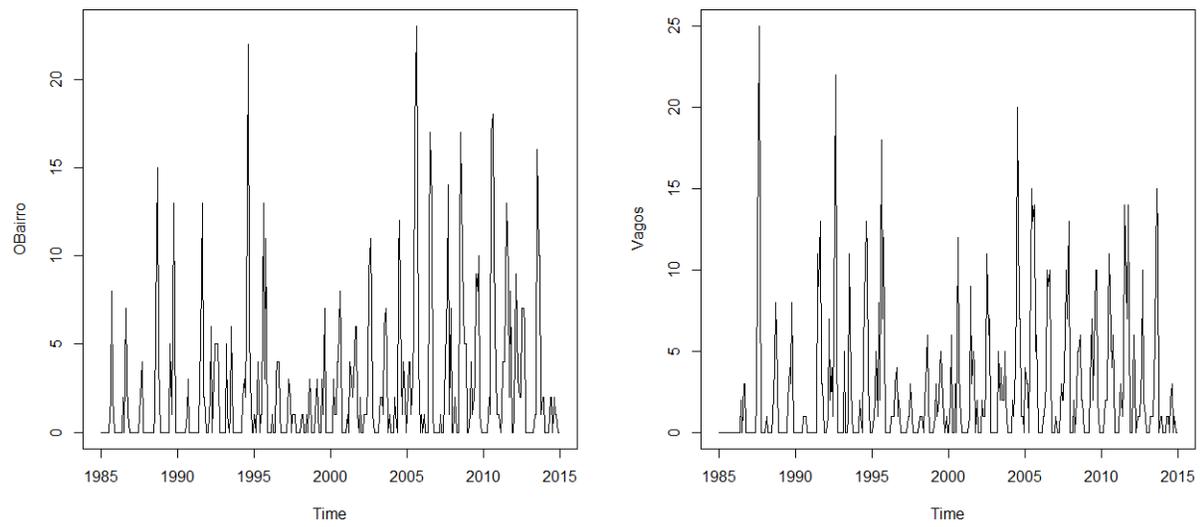


Figure 1: Monthly number of fires in O. Bairro and Vagos counties

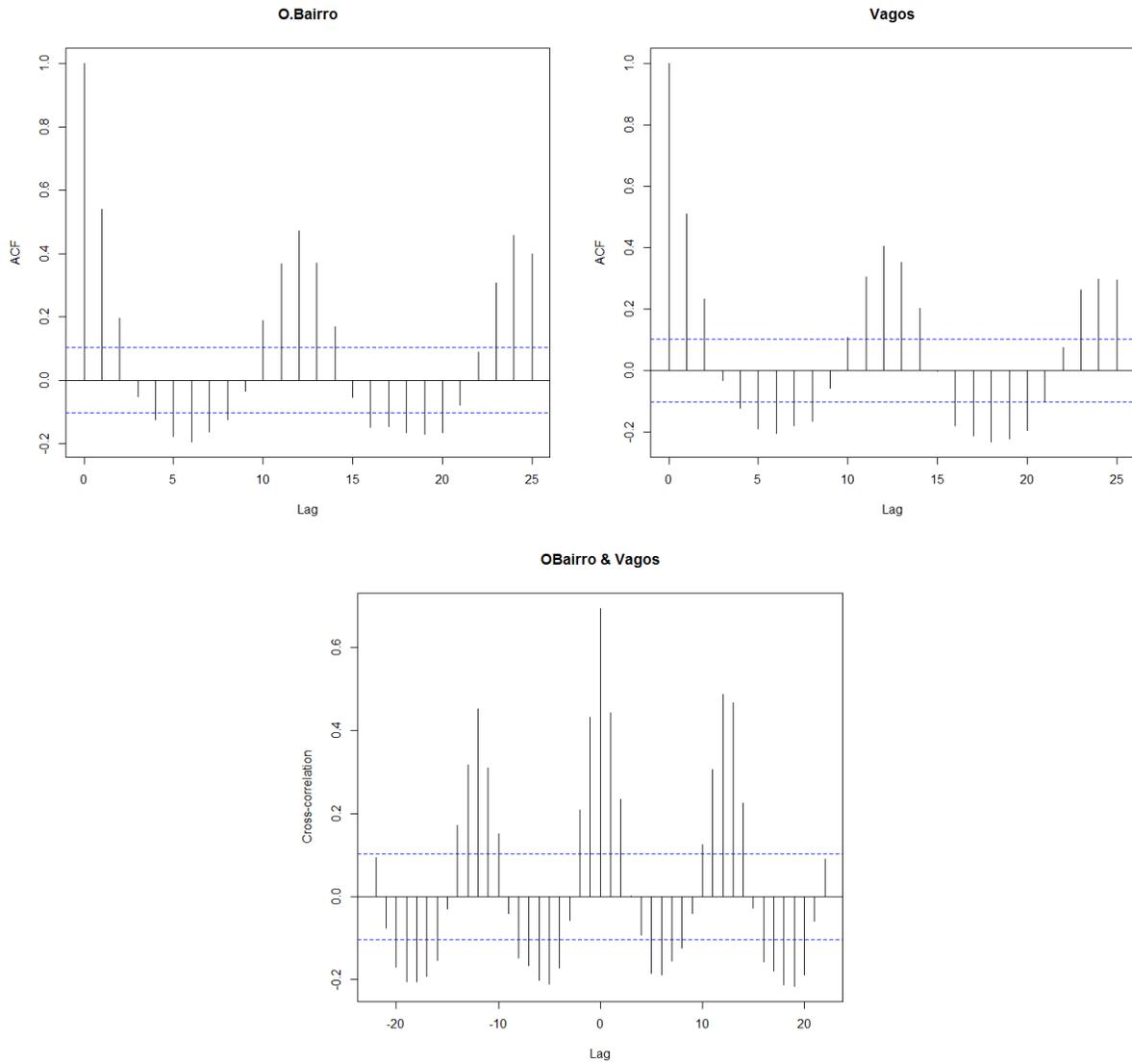


Figure 2: Sample autocorrelation (ACF) and cross-correlation (CCF) of Monthly number of fires in O. Bairro and Vagos Counties

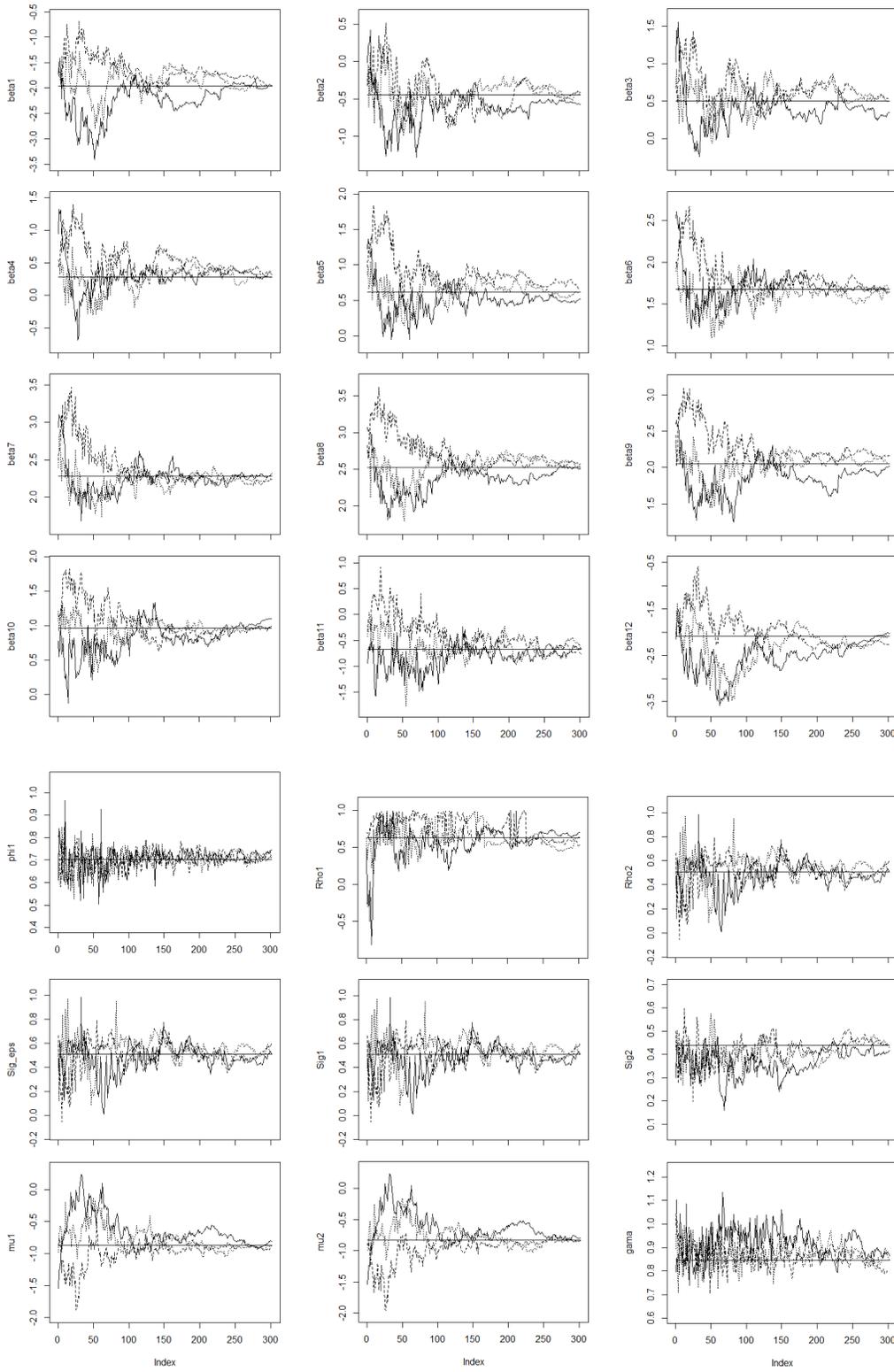


Figure 3: Trajectories of the algorithm updates showing  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_{300}$  convergence to the maximum likelihood estimate (horizontal line) starting from three different starting points  $\theta_0$

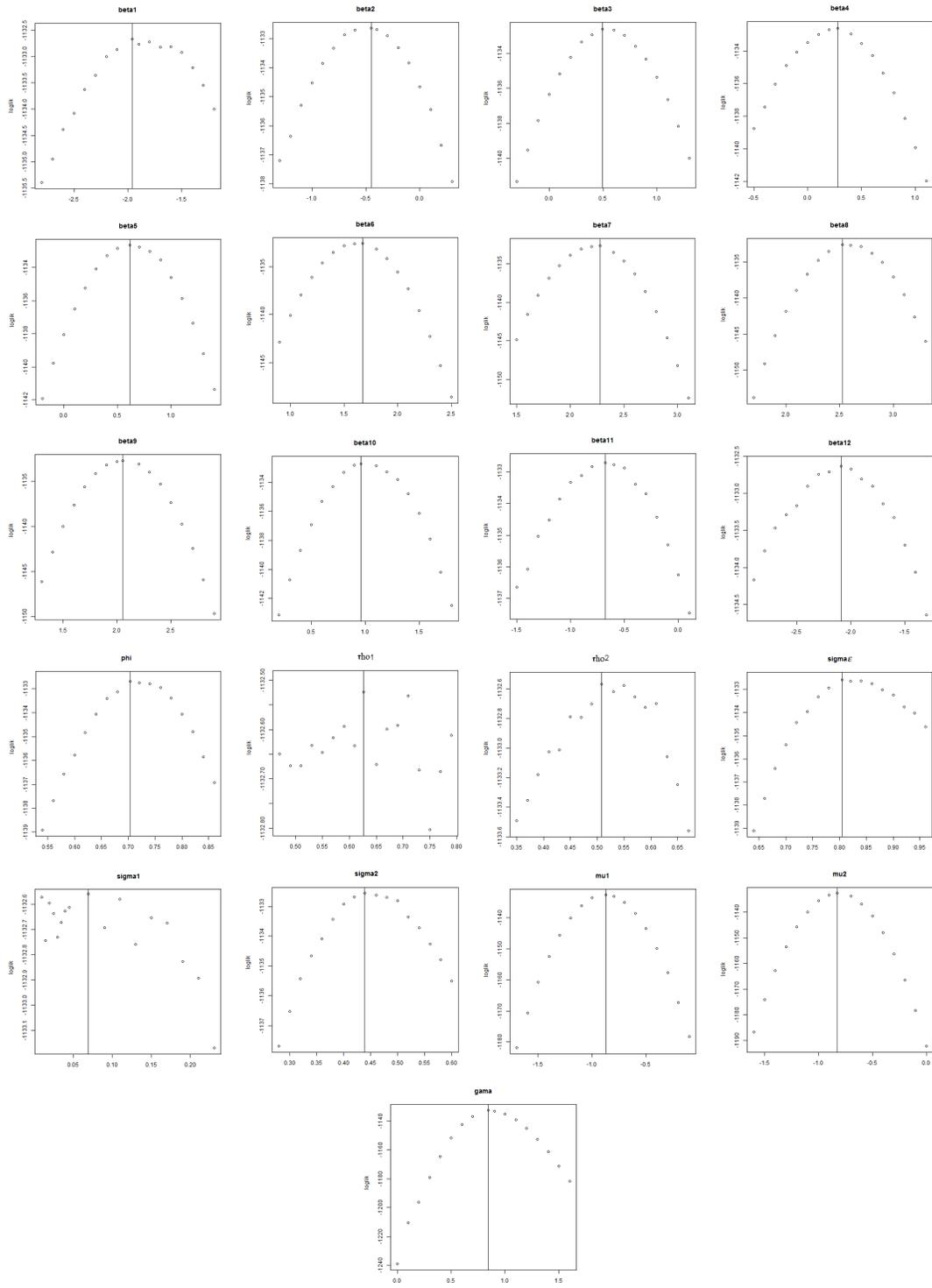


Figure 4: Sliced likelihoods for the corresponding parameters, in which the likelihood surface is explored along one of the parameters, keeping the other parameters fixed at the point which IF converges to. Each filled circle shows the likelihood estimate obtained with 70,000 particles.

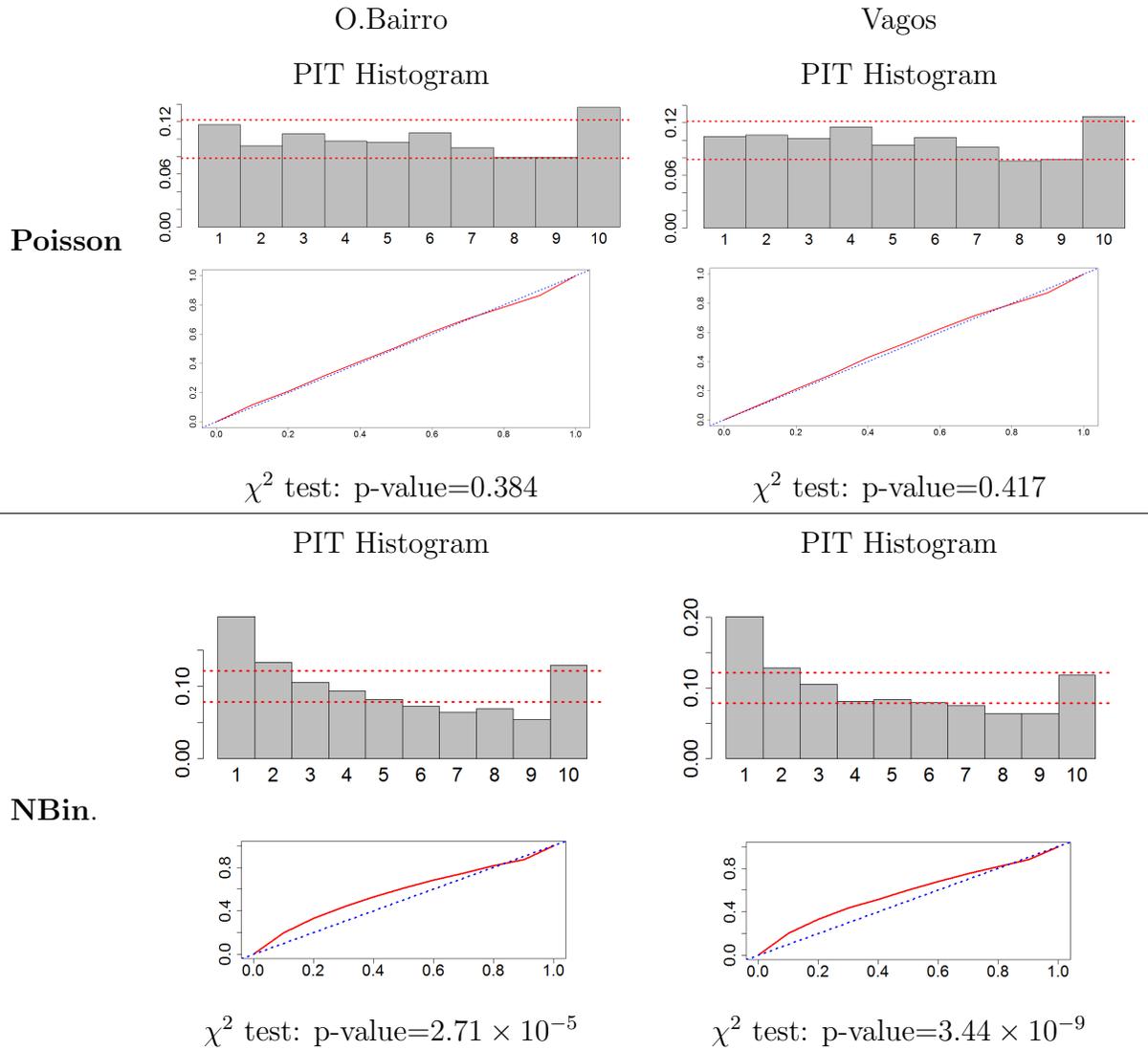


Figure 5: PIT-based diagnostics after fitting DF models (top figures); Cumulative predictive distributions against the uniform distribution (bottom figures); p-values of the goodness of fit chi-squared tests

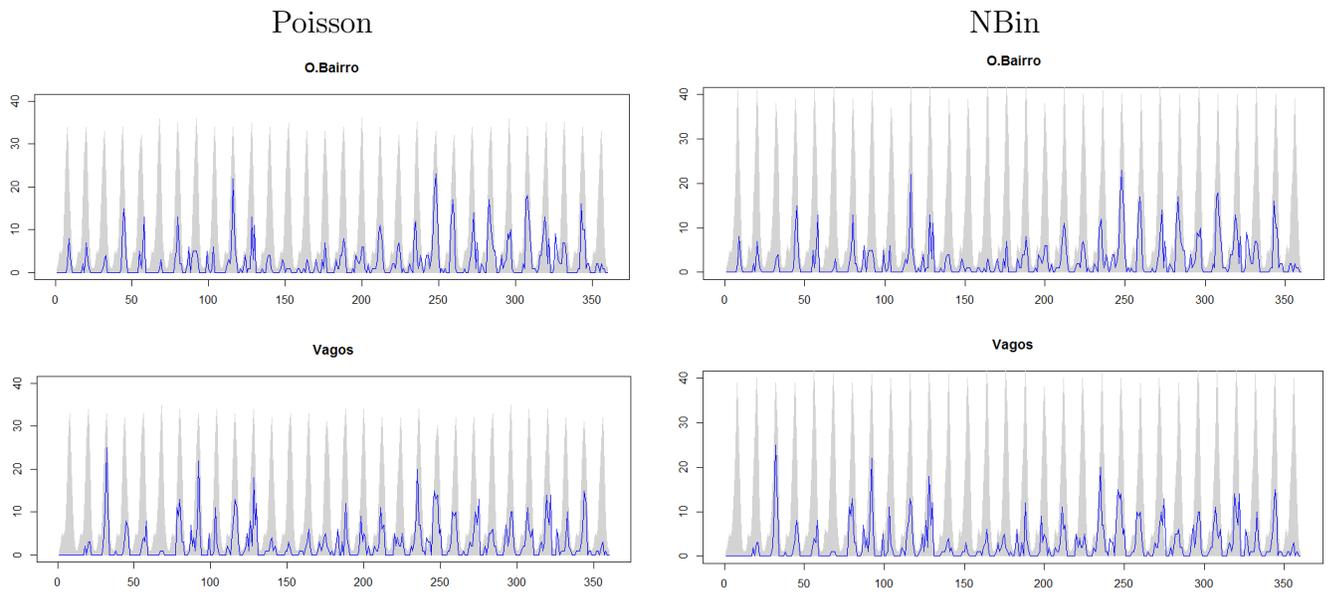


Figure 6: Parametric resampling diagnostics with 80% acceptance bounds for Poisson DF model and for negative binomial DF model.