Estimation and forecasting in SUINAR(1) model

Nélia Silva[1,3], Isabel Pereira[*,1,3], M. Eduarda Silva[2,3]

(nelia@mat.ua.pt, isabelp@mat.ua.pt, mesilva@fep.up.pt)

[1] Departamento de Matemática, Universidade de Aveiro, Portugal

[2] Faculdade de Economia, Universidade do Porto, Portugal

[3] UI&D Matemática e Aplicações, Universidade de Aveiro, Portugal

Abstract

This work considers a generalization of the INAR(1) model to the panel data first order Seemingly Unrelated INteger AutoRegressive Poisson model, SUINAR(1). It presents Bayesian and classical methodologies to estimate the parameters of Poisson SUINAR(1) model and to forecast future observations of the process. In particular, prediction intervals for forecasts - classical approach - and HPD prediction intervals - Bayesian approach - are derived. A simulation study is provided to give additional insight into the finite sample behaviour of the parameter estimates and forecasts.

Keywords: forecasts, Gibbs sampling, INAR model, panel data.
AMS: 62CF15, 62M10, 62M20

1 Introduction

The usual linear models for time series are suitable for modelling stationary dependent sequences under the assumption of Gaussianity, which is inappropriate for modelling counting processes. Motivated by the need of modelling correlated series counts, the INteger-valued AutoRegressive (INAR) process was proposed by Al-Osh and Alzaid (1987) and Mckenzie (1985). Generalizing the Poisson INAR(1) model, PoINAR(1), for a r units panel with n time periods where the parameters are constant along the time but different from individual to individual, we have the expression

$$X_{k,t} = \alpha_k \circ X_{k,t-1} + \epsilon_{k,t}, \qquad k = 1, \ldots, r; t = 2, \ldots, n, \tag{1}$$

where $x_{k,1}$ is known, $\alpha_k \circ X_{k,t-1} | X_{k,t-1} \sim B(X_{k,t-1}, \alpha_k)$, $\alpha_k \in (0,1)$, $\epsilon_{k,t}$, are, for each $k = 1, \ldots, r$ Poisson random variables with parameter μ_k and, moreover, $\epsilon_{k,t}$ and $X_{k,t-1}$ are independent, for all k and t. In many data sets individuals are not independent and here this dependence is modeled in (1) through the innovations term by

$$\epsilon_{k,t} = \epsilon^*_{k,t} + \zeta_t, \qquad k = 1, \ldots, r;\ t = 2, \ldots, n.$$

[*]Addressing author: Departamento de Matemática, Aveiro, Portugal

Thus, equation (1) takes the form

$$X_{k,t} = \alpha_k \circ X_{k,t-1} + \epsilon^*_{k,t} + \zeta_t, \qquad k = 1,\dots,r;\ t = 2,\dots,n, \tag{2}$$

with $\epsilon^*_{k,t} \sim P(\lambda_k)$ i.i.d., $k = 1,\dots,r$; $\zeta_t \sim P(\delta)$ i.i.d., $t = 2,\dots,n$; $\epsilon^*_{k,t}$ and ζ_t are independent for $k = 1,\dots,r$, $t = 2,\dots,n$.

The model defined in (2) is called *Seemingly Unrelated INteger AutoRegressive*, SUINAR, since the individuals appear independent from each other.

Particular situations of the model defined in (1) were studied by Silva *et al.* (2005b) - PoRINAR(1) model - where the parameters are constant along the time and from individual to individual, i.e., considering independent replicates of the PoINAR(1) model.

Other authors, Berglund and Brännäs (2001), Blundell *et al.* (1999) and Böckenholt (1999), considered a generalization of this model in which the parameters depend on exogenous variables and vary with time and from individual to individual.

In time series analysis we are usually interested in estimating the underlying model and in the predictive capabilities of that model. Thus, the aim of this study is to establish a comparison between classical and Bayesian approaches in order to conduct inference for model parameters and obtain predictions for future values.

The remaining of the paper is organized as follows: Section 2, the SUINAR process is introduced and some properties of the model are derived. In Section 3, the estimation of the parameters is studied under several classical methods and Bayesian methodology. This is analysed using an MCMC algorithm - ARMS - for which we give full details. In Section 4, forecasts of future observations and prediction intervals are derived, under both approaches. In Section 5, the results are illustrated through a simulation study. Finally, in Section 6 we some concluding remarks are given.

2 The SUINAR (1) model and its properties

Equation (2) is written in matrix form as

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_r \end{bmatrix}_t = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_r \end{bmatrix} \circ \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_r \end{bmatrix}_{t-1} + \begin{bmatrix} \epsilon^*_1 \\ \epsilon^*_2 \\ \vdots \\ \epsilon^*_r \end{bmatrix}_t + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_t \zeta_t,$$

or alternatively

$$\mathbf{x}_{.t} = \mathbf{A} \circ \mathbf{x}_{.(t-1)} + \boldsymbol{\epsilon}_{.t} + \mathbf{1}_r \zeta_t, \qquad t = 2,\dots,n,$$

with

$$\mathbf{A} \circ \mathbf{x}_{.(t-1)} = \left(\alpha_1 \circ X_1 = \sum_{i=1}^{X_1} B_{i1}, \cdots, \alpha_r \circ X_r = \sum_{i=1}^{X_r} B_{ir} \right)'_{t-1},$$

where $\mathbf{x}_{.t} = (X_{1,t}, X_{2,t}, \dots, X_{r,t})$, B_{ik} are i.i.d. Bernoulli random variables with α_k as the success probability and independent of $\mathbf{x}_{t-1}$ and ϵ_t, $t = 1, 2, \dots, n$.

The following properties are important for the remainder of the paper.

1. Let $\boldsymbol{\epsilon}_{.t} = \boldsymbol{\epsilon}^*_{.t} + \zeta_t \mathbf{1}_r -$ The covariance matrix of $(\boldsymbol{\epsilon}_{.t})$ at lag j is given by,

$$\gamma_\epsilon(j) = Cov(\boldsymbol{\epsilon}_{.t}, \boldsymbol{\epsilon}_{.(t+j)})$$
$$= \begin{bmatrix} cov(\epsilon_{1,t}, \epsilon_{1,t+j}) & cov(\epsilon_{1,t}, \epsilon_{2,t+j}) & \dots & cov(\epsilon_{1,t}, \epsilon_{r,t+j}) \\ cov(\epsilon_{2,t}, \epsilon_{1,t+j}) & cov(\epsilon_{2,t}, \epsilon_{2,t+j}) & \dots & cov(\epsilon_{2,t}, \epsilon_{r,t+j}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\epsilon_{r,t}, \epsilon_{1,t+j}) & cov(\epsilon_{r,t}, \epsilon_{2,t+j}) & \dots & cov(\epsilon_{r,t}, \epsilon_{r,t+j}) \end{bmatrix}.$$

When $j = 0$, it follows that

$$\gamma_\epsilon(0) = \begin{bmatrix} \lambda_1 + \delta & \delta & \cdots & \delta \\ \delta & \lambda_2 + \delta & \cdots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \cdots & \lambda_r + \delta \end{bmatrix}.$$

If $j \geq 1$, then $\gamma_\epsilon(j) = 0$, due to the independence between $\epsilon^*_{k,t}$ and ζ_t for $k = 1, \ldots, r$, $t = 2, \ldots, n$.

2. The mean value of the process $\mathbf{x}_{.t}$ is given by

$$E(\mathbf{x}_{.t}) = (\mathbf{I}_r - \mathbf{A})^{-1}(\boldsymbol{\lambda} + \delta \mathbf{1}_r),$$

where $\mathbf{x}_{.t} = (X_{1,t}, X_{2,t}, \ldots, X_{r,t})$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)$ and $\mathbf{I}_r$ is the $(r \times r)$ identity matrix.

For the kth individual, we have

$$\mathrm{E}[X_{k,t}] = (\lambda_k + \delta)/(1 - \alpha_k), \qquad k = 1, \ldots, r.$$

3. The covariance matrix of the process, $\mathbf{x}_{.t}$, is defined by

$$\gamma_X(0) = \begin{bmatrix} (\lambda_1 + \delta)/(1 - \alpha_1) & \delta/(1 - \alpha_1\alpha_2) & \cdots & \delta/(1 - \alpha_1\alpha_r) \\ \delta/(1 - \alpha_2\alpha_1) & (\lambda_2 + \delta)/(1 - \alpha_2) & \cdots & \delta/(1 - \alpha_2\alpha_r) \\ \vdots & \vdots & \ddots & \cdots \\ \delta/(1 - \alpha_r\alpha_1) & \delta/(1 - \alpha_r\alpha_2) & \cdots & (\lambda_r + \delta)/(1 - \alpha_r) \end{bmatrix}. \tag{3}$$

4. The covariance matrix $\mathbf{x}_{.t}$ at lag j is given by

$$\gamma_X(j) = E\left[(\mathbf{x}_{.t} - E(\mathbf{x}_{.t}))\left(\mathbf{x}_{.(t-j)} - E(\mathbf{x}_{.(t-j)})\right)\right]' = \mathbf{A}^j \gamma_X(0),\ j = 1, 2, \ldots.$$

3 Parameter Estimation

In this section we consider the estimation of the $2r + 1$ unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \delta) = (\alpha_1, \alpha_2, \ldots, \alpha_r; \lambda_1, \lambda_2, \ldots, \lambda_r; \delta)$of the SUINAR(1) process from the sample $\mathbf{x}_{r,n} = \{X_{k,t}; k = 1, 2, \ldots, r; t = 1, 2, \ldots, n\}$. The methods under study are the Conditional Maximum Likelihood, Conditional Minimum Square, method of moments and Bayesian methodology.

3.1 Classical Approach

3.1.1 Conditional Maximum Likelihood Estimators

The likelihood function, conditional on $\mathbf{x}_{.1}$=$(x_{1,1}, x_{2,1}, \ldots, x_{r,1})$, is given by the following expression

$$\begin{aligned} &L(\mathbf{x}_{r,n};\boldsymbol{\theta}|\mathbf{x}_{.1}) = \prod_{k=1}^{r}\prod_{t=2}^{n} P(X_{k,t} = x_{k,t}|X_{k,t-1} = x_{k,t-1}) \\ &= \prod_{k=1}^{r}\prod_{t=2}^{n}\sum_{i=0}^{M_{k,t}} \exp[-(\lambda_k+\delta)]\tfrac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!}\tbinom{x_{k,t-1}}{i}\alpha_k^i(1-\alpha_k)^{x_{k,t-1}-i}, \end{aligned} \tag{4}$$

with $M_{k,t} = \min(x_{k,t}, x_{k,t-1})$.

Estimates for δ and λ_k, $k = 1, \ldots, r$ cannot be obtained separatly due to the term $(\lambda_k + \delta)^{x_{k,t}-i}$. Thus, we consider $\mu_k = \lambda_k + \delta$ in the expression (4), and we obtain the conditional maximum likelihood (CML) estimates of α_k and μ_k.

The CML estimates satisfy the following system, where the equations were obtained by canceling the derivatives of the logarithm of expression (4)

$$\begin{cases} \dfrac{\partial \log L(\mathbf{x}_{r,n};\boldsymbol{\theta}|\mathbf{x}_{.1})}{\partial \mu_k} = 0 \Leftrightarrow \displaystyle\sum_{t=2}^{n}\frac{P_t(x_{k,t}-1)}{P_t(x_{k,t})} = (n-1) \\ \\ \dfrac{\partial \log L(\mathbf{x}_{r,n};\boldsymbol{\theta}|\mathbf{x}_{.1})}{\partial \alpha_k} = 0 \Leftrightarrow \displaystyle\sum_{t=2}^{n} x_{k,t} - \alpha_k\sum_{t=2}^{n} x_{k,t-1} - \mu_k\sum_{t=2}^{n}\frac{P_t(x_{k,t}-1)}{P_t(x_{k,t})} = 0 \end{cases},$$

where

$$P_t(y) = \exp[-(\lambda_k+\delta)]\sum_{i=0}^{M_{k,t}}\frac{(\lambda_k+\delta)^{y-i}}{(y-i)!}\binom{x_{k,t-1}}{i}\alpha_k^i(1-\alpha_k)^{x_{k,t-1}-i}.$$

These equations do not yield explicit forms for the estimators of μ_k and α_k, therefore iterative methods are used to solve the system. We use the bisection method, halving the amplitude of the interval which contains the zero of the function until the required precision is obtained.

3.1.2 Conditional Minimum Square Estimators

To obtain the Conditional Least Squares (CLS) estimators, we proceed similarly as Al-Osh and Alzaid (1987) analysis of PoINAR(1) model. Thus, the Conditional Least Squares (CLS) estimator of the parameter is obtained by minimizing

$$Q = \sum_{k=1}^{r}\sum_{t=2}^{n}[X_{k,t} - \mathrm{E}(X_{k,t}|X_{k,t-1})]^2 = \sum_{k=1}^{r}\sum_{t=2}^{n}[X_{k,t} - \alpha_k X_{k,t} - \lambda_k - \delta]^2. \tag{5}$$

Therefore, calculating the derivatives of the previous expression in order to α_k, λ_k and δ, we obtain respectively

$$\begin{cases} \partial Q/\partial\alpha_k = \sum_{t=2}^{n} -2X_{k,t-1}[X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta] \\ \\ \partial Q/\partial\lambda_k = \sum_{t=2}^{n} -2[X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta], k = 1, \ldots, r \\ \\ \partial Q/\partial\delta = -2\sum_{k=1}^{r}\sum_{t=2}^{n}[X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta] \end{cases}. \tag{6}$$

Setting the derivatives to zero, we observe that $\partial Q/\partial\delta$ is multiple of $\partial Q/\partial\lambda_k$. It is easy to check that the normal equations constitute an indeterminate system and, similarly to the maximum likelihood method, it is not possible to estimate the parameters $\delta, \alpha_k, \lambda_k,\ k = 1, \ldots, r$, separately. Therefore, once again we consider $\mu_k = \lambda_k + \delta$ in expression (5).
After some simple algebraic operations the estimators are given by

$$\hat{\alpha}_{k,LSE} = \frac{(n-1)\sum_{t=2}^{n} X_{k,t}X_{k,t-1} - \left(\sum_{t=2}^{n} X_{k,t}\right)\left(\sum_{t=2}^{n} X_{k,t-1}\right)}{(n-1)\sum_{t=2}^{n} X_{k,t-1}^2 - \left(\sum_{t=2}^{n} X_{k,t-1}\right)^2}$$

and

$$\hat{\mu}_{k,LSE} = \frac{\sum_{t=2}^{n} X_{k,t} - \hat{\alpha}_{k,LSE}\sum_{t=2}^{n} X_{k,t-1}}{(n-1)}.$$

3.1.3 Moment Estimators

Considering that the one step ahead prediction error is

$$e_{k,t} = X_{k,t} - \mathrm{E}(X_{k,t}|X_{k,t-1}), \qquad k = 1, 2, \ldots, r,$$

we have that $\mathrm{E}(e_{k,t}|X_{k,t-1}) = 0$, $\mathrm{E}(X_{k,t-1}e_{k,t}|X_{k,t-1}) = 0$ and the corresponding sample moments are the following

$$\begin{cases} \frac{1}{n-1}\sum_{t=2}^{n}(X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta) = 0 \\ \\ \frac{1}{n-1}\sum_{t=2}^{n} X_{k,t-1}(X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta) = 0 \end{cases}, \tag{7}$$

for k=1,2,...,r . This system has $2r$ equations and $2r+1$ unknown parameters so it will be necessary to add another equation in order to estimate all the parameters. Through the analysis of covariance matrix given in (3), we observe that

$$\mathrm{Cov}(X_{i,t}, X_{j,t}) - \frac{\delta}{1 - \alpha_i\alpha_j} = 0\,, \qquad i, j = 1, 2, \ldots, r, i \neq j,$$

being the corresponding sample moment given by

$$\frac{1}{\binom{r}{2}}\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\left[\frac{1}{n-1}\sum_{t=2}^{n}(X_{i,t} - \bar{X}_{i\cdot})(X_{j,t} - \bar{X}_{j\cdot}) - \frac{\delta}{1-\alpha_i\alpha_j}\right], \ \ i, j = 1, 2, \ldots, r, \tag{8}$$

with $\bar{X}_{k\cdot} = \sum_{t=2}^{n} X_{k,t}/(n-1),\ k = 1, 2, \ldots, r.$
The system (7) together with (8), allow us to obtain the estimators of the parameters δ, α_k and $\lambda_k,\ k = 1, \ldots, r$, which are given by

$$\hat{\alpha}_{k,MM} = \frac{(n-1)\sum_{t=2}^{n} X_{k,t}X_{k,t-1} - \left(\sum_{t=2}^{n} X_{k,t}\right)\left(\sum_{t=2}^{n} X_{k,t-1}\right)}{(n-1)\sum_{t=2}^{n} X_{k,t-1}^2 - \left(\sum_{t=2}^{n} X_{k,t-1}\right)^2},$$

$$\hat{\delta}_{MM} = \frac{\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{t=2}^{n}\left[(X_{i,t} - \bar{X}_{i\cdot})(X_{j,t} - \bar{X}_{j\cdot})\right]}{(n-1)\sum_{i=1}^{r-1}\sum_{j=1}^{r}[1/(1-\alpha_i\alpha_j)]},$$

$$\hat{\lambda}_{k,MM} = \frac{\sum_{t=2}^{n} X_{k,t} - \hat{\alpha}_{k,m.m.}\sum_{t=2}^{n} X_{k,t-1}}{(n-1)} - \hat{\delta}_{MM}.$$

3.2 Bayesian Approach

It is well known that Bayesian inference is based on the posterior distribution, since this distribution contains all the available information about the unknown parameters $\boldsymbol{\theta}$. After observing the particular sample $\mathbf{x}_n$, the updated information about $\boldsymbol{\theta}$ is expressed by Bayes theorem through posterior distribution which is given by,

$$\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n}) = \frac{L(\mathbf{x}_{r,n};\boldsymbol{\theta}|\mathbf{x}_{.1})\pi(\boldsymbol{\theta})}{\int_\Theta L(\mathbf{x}_{r,n};\boldsymbol{\theta}|\mathbf{x}_{.1})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto L(\mathbf{x}_{r,n};\boldsymbol{\theta}|\mathbf{x}_1)\pi(\boldsymbol{\theta})\,, \quad \boldsymbol{\theta}\in\Theta, \tag{9}$$

where $\pi(\boldsymbol{\theta})$ denotes the prior distribution. In a Bayesian framework it is necessary to assign priors to each parameter. In this work, the prior distributions considered are the beta and gamma distributions since they are conjugated of binomial and Poisson distributions, respectively. Therefore, beta distribution with parameters $a_k, b_k > 0$ is the prior for α_k, $\alpha_k \frown \text{Be}(a_k, b_k)$, and gamma distributions with parameters $c_k, d_k > 0$, $\lambda_k \frown \text{Ga}(c_k, d_k)$ and $e, f > 0$, $\delta \frown \text{Ga}(e, f)$ are the priors for λ_k and δ, respectively.
Moreover, we assume independence between α_k, λ_k and δ, for $k = 1, 2, \dots, r$, as well as the knowledge of hiperparameters a_k, b_k, c_k, d_k, e and f, $k = 1, 2, \dots, r$. Therefore, the prior distribution of the $2r+1$ parameters $(\alpha_1, \alpha_2, \dots, \alpha_r;\ \lambda_1, \lambda_2, \dots, \lambda_r; \delta)$ has the form,

$$\pi(\boldsymbol{\theta}) = \pi(\delta)\prod_{k=1}^r \pi(\alpha_k)\pi(\lambda_k) \propto \delta^{e-1}\exp(-f\delta)\prod_{k=1}^r {\alpha_k}^{a_k-1}(1-\alpha_k)^{b_k-1}\lambda_k^{c_k-1}\exp(-d_k\lambda_k).$$

Thus, by Bayes theorem it follows from the prior and the likelihood (4), that the posterior distribution is given by the following expression

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n}) \propto= \delta^{e-1}\exp(-f\delta)\left(\prod_{k=1}^r {\alpha_k}^{a_k-1}(1-\alpha_k)^{b_k-1}\lambda_k^{c_k-1}\exp(-d_k\lambda_k)\right)\times \\ \times\left(\prod_{k=1}^r\prod_{t=2}^n\sum_{i=0}^{M_{k,t}}\exp[-(\lambda_k+\delta)]\frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!}\binom{x_{k,t-1}}{i}\alpha_k^i(1-\alpha_k)^{x_{k,t-1}-i}\right).\end{aligned} \tag{10}$$

The Bayes estimate for $\boldsymbol{\theta}$ is the mean of this distribution which cannot be obatined analitically. Thus we use the Gibbs sampler in order to generate values of $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$. Through Gibbs sampler and based on a irredutible Markov chain with state space Θ whose stationary distribution is $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$, a sequence of correlated realizations is generated. In this context the algorithm is based on the fact that (Besag, 1974 and Gelfand and Smith, 1990), if the joint distribution $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$ is positive over its entire domain, then it is uniquely determined by the m full conditional distributions $\pi(\theta_i|\mathbf{x}_{r,n}, \boldsymbol{\theta}_{-i}), i = 1, 2, \dots, m$, where $\boldsymbol{\theta}_{-i}$ represents the vector $\boldsymbol{\theta}$ after being removed θ_i component. The full conditional posterior densities are

- for α_k

$$\pi(\alpha_k|\boldsymbol{\alpha}_{-k}, \boldsymbol{\lambda}, \delta, \mathbf{x}_{r,n}) = \pi(\alpha_k|\lambda_k, \delta, \mathbf{x}_{k\cdot})$$

$$\propto \alpha_k^{a_k-1}(1-\alpha_k)^{b_k-1}\prod_{t=2}^n\sum_{i=0}^{M_{k,t}}\frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!}\binom{x_{k,t-1}}{i}\alpha_k^i(1-\alpha_k)^{x_{k,t-1}-i},$$

with $\boldsymbol{\alpha}_{-k} = (\alpha_1, \ldots, \alpha_{k-1}, \alpha_{k+1}, \ldots, \alpha_r)$, $\mathbf{x}_{k\cdot} = (x_{k,t} : t = 1, 2, \ldots, n)$;

- for λ_k

$$\begin{aligned}\pi(\lambda_k|\boldsymbol{\lambda}_{-k}, \alpha, \delta, \mathbf{x}_{r,n}) &= \pi(\lambda_k|\alpha_k, \delta, \mathbf{x}_{k\cdot}) \\ &\propto \lambda_k^{c_k-1} \exp[-(\lambda_k d_k)] \textstyle\prod_{t=2}^{n} \sum_{i=0}^{M_{k,t}} \exp[-(\lambda_k + \delta)] \\ &\quad \frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!} \binom{x_{k,t-1}}{i} \alpha_k^i (1-\alpha_k)^{x_{k,t-1}-i},\end{aligned}$$

with $\boldsymbol{\lambda}_{-k} = (\lambda_1, \ldots, \lambda_{k-1}, \lambda_{k+1}, \ldots, \lambda_r)$;

- for δ

$$\begin{aligned}\pi(\delta|\boldsymbol{\alpha}, \boldsymbol{\lambda}, \mathbf{x}_{r,n}) &\propto \delta^{e-1} \exp(-f\delta) \textstyle\prod_{k=1}^{r} \prod_{t=2}^{n} \sum_{i=0}^{M_{k,t}} \exp[-(\lambda_k + \delta)] \\ &\quad \frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!} \binom{x_{k,t-1}}{i} \alpha_k^i (1-\alpha_k)^{x_{k,t-1}-i}.\end{aligned}$$

The generation of pseudo-random numbers through the full conditional posterior densities may be achieved through the Adaptive Rejection Sampling (ARS) if the functions were surely log-concave. However, since this is not generally the case, we use Adaptive Rejection Metropolis Sampling (ARMS), which is an hybrid method introduced by Gilks *et al.*(1995). Thus, in Gibbs sampler each value $\boldsymbol{\theta}_{-i}$ is generated from $\pi(\theta_i|\mathbf{x}_{r,n}, \boldsymbol{\theta}_{-i})$ through ARMS algorithm in the following way:

Algorithm 1 *1. generate a random sample of the model* (2)*;*

2. *calculate the initial estimates of $\alpha_1, \ldots, \alpha_r$ and δ, by the moments method; denote them by $\alpha_{1,0}, \ldots, \alpha_{r,0}$ and δ_0;*

3. *using ARMS method, simulate for each $k = 1, 2, \ldots, r$,*

$$\lambda_{k,1} \text{ from } \pi(\lambda_k|\mathbf{x}_{k\cdot}, \delta_0, \alpha_{k,0}),$$

and

$$\alpha_{k,1} \text{ from } \pi(\alpha_k|\mathbf{x}_{k\cdot}, \delta_0, \lambda_{k,1});$$

4. *simulate, using ARMS method,*

$$\delta_1 \text{ from } \pi(\delta|\mathbf{x}_{r,n}, \alpha_{1,1}, \ldots, \alpha_{r,1}, \lambda_{1,1}, \ldots, \lambda_{r,1});$$

5. *repeat steps 3. and 4. with $i = 2, \ldots, nig$ (number of Gibbs sampler iterations); that is, for $k = 1, 2, \ldots, r$,*

$$\lambda_{k,i} \text{ is simulated from } \pi(\lambda_k|\mathbf{x}_{k\cdot}, \delta_{i-1}, \alpha_{k,i-1})$$

$$\alpha_{k,i} \text{ is simulated from } \pi(\alpha_k|\mathbf{x}_{k\cdot}, \delta_{i-1}, \lambda_{k,i})$$

$$\delta_i \text{ is simulated from } \pi(\delta|\mathbf{x}_{r,n}, \alpha_{1,i}, \ldots, \alpha_{r,i}, \lambda_{1,i}, \ldots, \lambda_{r,i});$$

6. *despising the first b values (corresponding to the burn-in period) and picking up each value, obtain a sample with $m = (nig - b)/l$ elements. Denote the corresponding sample means by: $\alpha_{k,B}^{(1)}$, $\lambda_{k,B}^{(1)}$ e $\delta_B^{(1)}$;*

7. *repeat nrep times the steps 1. to 6..*

Afterwards Bayes estimates can be calculated through the expression

$$\hat{\alpha}_{k,B} = \frac{1}{nrep}\sum_{i=1}^{nrep}\alpha_{k,B}^{(i)}, \quad \hat{\lambda}_{k,B} = \frac{1}{nrep}\sum_{i=1}^{nrep}\lambda_{k,B}^{(i)} \quad \text{and} \quad \hat{\delta}_B = \frac{1}{nrep}\sum_{i=1}^{nrep}\delta^{(i)}.$$

4 Predictive Inference

Let $\mathbf{x}_n = \{X_{k,t} : k = 1,\ldots,r, t = 2,\ldots,n\}$ be a sample generated by the Poisson SUINAR(1) model. We aim at obtaining the h-step-ahead predictor of $X_{k,n+h}$, $\hat{X}_{k,n+h}$. We begin by presenting some results fundamental to the understanding of the work.

According to the definition of SUINAR(1) process, we have that

$$X_{k,n+h} = \alpha_k \circ X_{k,n+h-1} + \epsilon_{k,n+h}. \tag{11}$$

Iterating backwards h times, equation (11) can be written as

$$X_{k,n+h} = \alpha_k^h \circ X_{k,n} + \sum_{j=1}^{h}\alpha_k^{h-j}\circ\epsilon_{k,n+j}, \quad h = 1,2,\ldots.$$

Since $X_{k,n}$ is independent of $\epsilon_{k,n+j}$, $j = 1,\ldots,h$, the conditional distribution of $X_{k,n+h}$ on $X_{k,n}$ is

$$\begin{aligned} &P\left(X_{k,n+h} = x \,\middle|\, X_{k,n}\right) = P\left(\alpha_k^h \circ X_{k,n} + \textstyle\sum_{j=1}^{h}\alpha_k^{h-j}\circ\epsilon_{k,n+j} = x \,\middle|\, X_{k,n}\right) \\ &= \textstyle\sum_{y=0}^{\min X_{k,n},x} P\left(\alpha_k^h\circ X_{k,n} = y \,\middle|\, X_{k,n}\right) P\left(\textstyle\sum_{j=1}^{h}\alpha_k^{h-j}\circ\epsilon_{k,n+j} = x-y\right). \end{aligned}$$

Noting that $\alpha_k \circ X_{k,n} \,\Big|\, X_{k,n} \sim Bi(X_{k,n},\alpha_k)$ and $\epsilon_{k,t}\sim P(\lambda_k)$, it follows easily that the distribution of $X_{k,n+h}|X_{k,n}$ is the convolution of the distribution of the innovation process, Poisson distribution with parameter $(\lambda_k+\delta)(1-\alpha_k^h)/(1-\alpha_k)$, and that resulting from the binomial thinning operation, binomial distribution with parameters $X_{k,n}$ and α_k^h. This result, proved in Silva (2005a), is established in the following theorem:

Theorem 1 *For the Poisson SUINAR(1) model, the distribution of $X_{k,n+h}$ given $X_{k,n}$ is the convolution of a binomial distribution with parameters $X_{k,n}$ and α_k^h and a Poisson distribution with parameter $(\lambda_k+\delta)(1-\alpha_k^h)/(1-\alpha_k)$. That is to say, $X_{k,n+h}|X_{k,n}$ has the moment generating function*

$$\varphi_{X_{k,n+h}|X_{k,n}}(s) = \left[\alpha_k^h e^s + (1-\alpha_k^h)\right]^{x_{k,n}} \exp\left\{(\lambda_k+\delta)\frac{1-\alpha_k^h}{1-\alpha_k}(e^s-1)\right\}. \tag{12}$$

Thus, the probabilitity function of $X_{k,n+h}|X_{k,n}$, $k = 1,2\ldots,r$, is given by

$$\begin{aligned} p(x_{k,n+h}|x_{k,n}) &= P(X_{k,n+h} = x|X_{k,n} = x_{k,n}) = \textstyle\sum_{i=0}^{\min(x,x_{k,n})}\binom{x_{k,n}}{i}(\alpha_k^h)^i(1-\alpha_k^h)^{x_{k,n}-i} \\ &\exp\left[-(\lambda_k+\delta)\tfrac{1-\alpha_k^h}{1-\alpha_k}\right]\tfrac{1}{(x-i)!}\left[(\lambda_k+\delta)\tfrac{1-\alpha_k^h}{1-\alpha_k}\right]^{x-i}, k = 1,2\ldots,r. \end{aligned} \tag{13}$$

Since $\lim_{h\to+\infty}\varphi_{X_{k,n+h}|X_{k,n}}(s) = \exp\left[\frac{\lambda_k+\delta}{1-\alpha_k}(e^s-1)\right]$, the corollary follows.

Corollary 1 *$X_{k,n+h}|X_{k,n}$ has the Poisson limit distribution with parameter $(\lambda_k+\delta)/(1-\alpha_k)$.*

4.1 Classical Prediction

4.1.1 Forecasts of future observations

Analogously to the study made by Silva *et al.*(2006) concerning prediction in PoINAR(1) processes, we will calculate two predictors of $X_{k,n+h}$. One of them is based on the minimization of mean square error and the other minimizes the mean absolute error.

Due to the fact that the best predictor which minimizes the mean square error is $\hat{X}_{k,n+h}$ =E$[X_{k,n+h}|X_{k,n}]$ and according to expression (12), it comes straightforwardly that $\mathrm{E}[X_{k,n+h}|X_{k,n}] = \varphi'_{X_{k,n+h}|X_{k,n}}(s)$ calculated on $s = 0$. Therefore

$$\hat{X}_{k,n+h} = \mathrm{E}[X_{k,n+h}|X_{k,n}] = \alpha_k^h X_{k,n} + \frac{1-\alpha_k^h}{1-\alpha_k}(\lambda_k + \delta), \quad k = 1, 2, \ldots, r. \tag{14}$$

This method hardly produces coherent predictions in the sense that forecasts of integer values must be integer values as well (see Chatfield, 2001). In order to obtain coherent predictions for X_{n+h}, Freeland and McCabe (2003) suggest using the value which minimizes the expected absolute error given the sample, i.e., the value that minimizes $E[|X_{n+h} - \hat{X}_{n+h}| \,\big|\, X_n]$. Let $m_{k,h}$ be the median of the conditional distribution $X_{k,n+h}|X_{k,n}$. It can be proved that $\mathrm{E}[|X_{k,n+h} - \hat{m}_{k,n+h}| \,\Big|\, X_{k,n}]$ has a global minimum in $\hat{m}_{k,n+h} = m_{k,h}$; in this sense, this means that median of the predictive distribution is the best predictor of $X_{k,n+h}$.

4.1.2 Prediction Intervals

A prediction interval is always more informative than a point forecast. The method of obtaining confidence intervals for the predicted value is based on the probability function of the h-steps-ahead forecast error, which is given by

$$e_{k,n+h}|\mathbf{x}_{r,n} = X_{k,n+h} - \hat{X}_{k,n+h} = X_{k,n+h} - \alpha_k^h x_{k,n} - \frac{1-\alpha_k^h}{1-\alpha_k}(\lambda_k + \delta).$$

It is worth to mention that $e_{k,n+h}$ is a discrete variable taking values on $\{j - \alpha_k^h x_{k,n} - [(\lambda_k + \delta)(1 - \alpha_k^h)/(1-\alpha_k)];\ j = 0, 1, 2 \ldots\}$; hence has the probability function,

$$\begin{aligned} P\left(e_{k,n+h}|\mathbf{x}_{r,n} = j - \alpha_k^h x_{k,n} - (\lambda_k+\delta)\frac{1-\alpha_k^h}{1-\alpha_k}\right) &= \mathrm{P}(X_{k,n+h} = j|X_{k,n} = x_{k,n}) = \\ = \exp\left[-(\lambda_k+\delta)\frac{1-\alpha_k^h}{1-\alpha_k}\right] \textstyle\sum_{i=0}^{\min(j,x_{k,n})} \frac{[(\lambda_k+\delta)(1-\alpha_k^h)/(1-\alpha_k)]^{j-i}}{(j-i)!} &\times \binom{x_{k,n}}{i}(\alpha_k^h)^i(1-\alpha_k^h)^{x_{k,n}-i}. \end{aligned}$$

Once known the probability function of the forecast error, the $100\gamma\%$ confidence interval for $X_{k,n+h}$ is given by

$$(\hat{X}_{k,n+h} + e_{t_1}, \hat{X}_{k,n+h} + e_{t_2}), \tag{15}$$

where $\hat{X}_{k,n+h}$ is defined by (14), e_{t_1} is the greatest value $e_{k,n+h}|\mathbf{x}_{r,n}$ such as $\mathrm{P}(e_{k,n+h}|\mathbf{x}_{r,n} \leq e_{t_1}) \leq (1-\gamma)/2$ and e_{t_2} is the lowest value of $e_{k,n+h}|\mathbf{x}_{r,n}$, such as $\mathrm{P}(e_{k,n+h}|\mathbf{x}_{r,n} \leq e_{t_2}) \geq (1+\gamma)/2$.

4.2 Bayesian Prediction

To obtain the Bayesian predictive function we use the randomness of both, the future observation $X_{k,n+h}$ we want to predict and the vector of unknown parameters $\boldsymbol{\theta}$. Moreover, information about

$\boldsymbol{\theta}$ is contained in the observed sample $\mathbf{x}_{r,n}$ and is quantified on the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$. Thus the following definition.

Definition 1 *Let $\boldsymbol{\theta} \in \Theta$ be the vector of unknown parameters. The h steps-ahead bayesian posterior predictive distribution is defined by*

$$\pi(x_{k,n+h}|\mathbf{x}_{r,n}) = \int_\Theta \pi(x_{n+h}; \boldsymbol{\theta}|\mathbf{x}_{r,n}) d\boldsymbol{\theta} = \int_\Theta p(x_{k,n+h}|\mathbf{x}_{r,n}; \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n}) d\boldsymbol{\theta}, \tag{16}$$

where $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$ is the posterior probability density function of $\boldsymbol{\theta}$ and $p(x_{k,n+h}|\mathbf{x}_{r,n}; \boldsymbol{\theta})$ is the classic predictive function.

The predictive distribution $X_{n+h}|\mathbf{x}_{r,n}$ given by (16) is looked upon as containing all the accumulated information on the future values. Therefore, the Bayesian predictor of $X_{k,n+h}$ can be calculated through the mean value, the median or the mode of the predictive function $\pi(x_{k,n+h}|\mathbf{x}_{r,n})$.

4.2.1 Forecasts of future observations

According to the Definition 1, the h-steps-ahead Bayesian predictive function for the kth individual of the SUINAR(1) model is given by,

$$\begin{aligned}
\pi(x_{k,n+h}|\mathbf{x}_{r,n}) &= \int_{\Theta_k} \pi(x_{k,n+h}, \boldsymbol{\theta}_k|\mathbf{x}_{r,n}) d\boldsymbol{\theta}_k \\
&= \int_{\Theta_k} p(x_{k,n+h}|\mathbf{x}_{r,n}, \boldsymbol{\theta}_k)\pi(\boldsymbol{\theta}_k|\mathbf{x}_{r,n}) d\boldsymbol{\theta}_k \\
&= \int_{\Theta_k} p(x_{k,n+h}|x_{k,n}, \boldsymbol{\theta}_k)\pi(\boldsymbol{\theta}_k|\mathbf{x}_{r,n}) d\boldsymbol{\theta}_k,
\end{aligned} \tag{17}$$

where,$\theta_k = (\delta, \alpha_k, \lambda_k)$, $p(x_{k,n+h}|x_{k,n}, \theta_k)$, $k = 1, 2, \ldots, r$, is given by (13) and $\pi(\theta_k|\mathbf{x}_{r,n})$ is the posterior probability density function of θ_k defined by,

$$\begin{aligned}
&\pi(\theta_k|\mathbf{x}_{r,n}) \propto \pi(\theta_k)L(\mathbf{x}_{r,n}, \delta, \lambda_k, \alpha_k|\mathbf{x}_{.1}) \propto \\
&\propto \delta^{e-1}\exp(-f\delta){\alpha_k}^{a_k-1}(1-\alpha_k)^{b_k-1}\lambda_k^{c_k-1}\exp(-d_k\lambda_k)\times \\
&\left(\prod_{t=2}^n \sum_{i=0}^{M_{k,t}} \exp[-(\lambda_k+\delta)]\frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!}\binom{x_{k,t-1}}{i}\alpha_k^i(1-\alpha_k)^{x_{k,t-1}\,-\,i}\right).
\end{aligned}$$

Usually, $X_{k,n+h}$ is predicted by $E(X_{k,n+h}|\mathbf{x}_{r,n})$ which does not seem feasible here due to the complexity of equation (17). Thus we propose two methodologies to deal with the problem. In the first approach, using the expected value properties $E(X_{k,n+h}|\mathbf{x}_{r,n})$ is rewriten as follows.

$$\begin{aligned}
E[X_{k,n+h}|\mathbf{x}_{r,n}] &= E[E(X_{k,n+h}|\mathbf{x}_{r,n}, \theta_k)|\mathbf{x}_{r,n}] \\
&= E[\alpha_k^h X_{k,n} + (1-\alpha_k^h)(\lambda_k+\delta)/(1-\alpha_k)\,\Big|\,\mathbf{x}_{r,n}] \qquad \text{por (14)} \\
&= X_{k,n}E(\alpha_k^h|\mathbf{x}_{r,n}) + E[(1-\alpha_k^h)(\lambda_k+\delta)/(1-\alpha_k)\,\Big|\,\mathbf{x}_{r,n}].
\end{aligned}$$

Now, the mean values $E(\alpha_k^h|\mathbf{x}_{r,n})$ and $E[(1-\alpha_k^h)(\lambda_k+\delta)/(1-\alpha_k)\,\Big|\,\mathbf{x}_{r,n}]$, can be estimated using Gibbs methodology jointly with ARMS algorithm to generate values of the full conditional distributions: $(\delta^1, \delta^2, \ldots, \delta^m)$, $(\alpha_k^1, \alpha_k^2, \ldots, \alpha_k^m)$ and $(\lambda_k^1, \lambda_k^2, \ldots, \lambda_k^m)$ for $k = 1, 2, \ldots, r$, necessary to the evaluation of correspondent ergodic means (see section 3.2). Thus, $X_{k,n+h}$ can be estimated by,

$$\hat{X}_{k,n+h} = x_{k,n}\frac{1}{m}\sum_{i=1}^m(\alpha_k^i)^h + \left[\frac{1}{m}\sum_{i=1}^m \frac{1-(\alpha_k^i)^h}{1-\alpha_k^i}\left(\lambda_k^i+\delta^i\right)\right]. \tag{18}$$

The second approach applies Tanner composition method, Tanner (1996), to the SUINAR(1) model. A sample $(X_{k,n+h,1}, X_{k,n+h,2}, \ldots, X_{k,n+h,m})$ is generated from the predictive distribution (17) using Algorithm 2 described bellow. Then, the forecast for the future observation $X_{k,n+h}$ can be calculated through the sample mean, median or mode.

Algorithm 2 *1. Calculate an initial estimate α_0 and δ_0 for α_k and δ, respectively, using a classic estimation method from a sample $\{X_{k,t} : k = 1, \ldots, r, t = 2, \ldots, n\}$ of the Poisson SUINAR(1) defined by (2),;*

2. *using Gibbs methodology jointly with adaptive rejection Metropolis sampling (ARMS), sample values of the triplets $(\alpha_{k,1}, \lambda_{k,1}, \delta_1)$, $(\alpha_{k,2}, \lambda_{k,2}, \delta_2), \ldots, (\alpha_{k,m}, \lambda_{k,m}, \delta_m)$ from the full conditional distributions of α_k, λ_k and δ;*

3. *for each i $(i = 1, \ldots, m)$ draw $X_{k,n+h,i}$ from $\pi(x_{k,n+h}|x_{r,n}, \alpha_{k,i}, \lambda_{k,i}, \delta_i)$, using the inverse transformation method adapted to discrete variables. That means,*

   (a) *sample u from $U(0, 1)$,*

   (b) *evaluate the lowest integer value s: $\sum_{i=0}^{s} \pi(x_{k,n+h}|x_{r,n}, \alpha_i, \lambda_i, \delta_i) \geq u$,*

   (c) *consider $X_{k,n+h,i} = s$.*

Thus, we have sampled $X_{k,n+h,1}, X_{k,n+h,2}, \ldots, X_{k,n+h,m}$ from the posterior predictive distribution.

4.2.2 HPD predictive intervals

In this section Highest Probability Density (HPD) predictive intervals are obtained from the posterior predictive distribution (Paulino *et al.*, 2003).

Definition 2 *$R(\gamma) = (X_L, X_R)$ is a prediction interval HPD (degree γ) for $X_{k,n+h}$ if*

$$P(X_L \leq X_{k,n+h} \leq X_R) = \sum_{x_{k,n+h}=X_L}^{X_R} \pi(x_{k,n+h}|\mathbf{x}_{r,n}) \geq K_\gamma,$$

where K_γ is the largest constant such that $P[X_{n+h} \in R(\gamma)] \geq \gamma$.

The computation of the HPD interval for $X_{k,n+h}$ is hindered by the lack of an explicit expession for the posterior predictive probability function, equation (17). However an estimate of $R(\gamma)$ may be obtained using Chen and Shao (1999) algorithm which is outlined next.

Algorithm 3 *1. draw a sample from $\pi(x_{k,n+h}|\mathbf{x}_{r,n})$ (Algorithm 2);*

2. *order the sample values $X_{(k,n+h,1)}, X_{(k,n+h,2)}, \ldots, X_{(k,n+h,m)}$, obtained in 1.;*

3. *for fixed γ, calculate the intervals*

$$\hat{R}_i(\gamma) = \left(X_{(k,n+h,i)}, X_{(k,n+h,i+[m\gamma])}\right), \quad 1 \leq i \leq m - [m\gamma],$$

where $[m\gamma]$ is the integer part of $m\gamma$. Choose for $100\gamma\%$ HPD interval for $X_{k,n+h}$, the $\hat{R}(\gamma)$ with smallest amplitude.

$\hat{R}(\gamma)$ is an estimator of $R(\gamma)$, whose asymptotic properties are valid under certain regularity conditions (Theorem 7.3.1., Chen *et al.*, 2000).
Noting that we are considering point processes, the Algorithm 3 can produce more than one interval. When this is the case we choose for $\hat{R}(\gamma)$ the interval with highest absolute frequency, between those with smaller amplitude; in the case of equality the absolute frequencies, the interval considered is the one with smaller inferior limit as suggested by Chen *et al.* (2000).

5 Simulation Study

In this section the small sample properties of the estimation and forecasting methods proposed are accessed by means of a simulation study. The data are generated according to the model (2) with $r = 5$, values for α_k and λ_k as described in Table 1 and setting $\delta = 2$, in a total of nine models. For each model, 200 time series of dimension $n = 25, 50, 100$ are generated.

Table 1: Values of the vector parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ used to simulate the samples

k		1	2	3	4	5	k		1	2	3	4	5
	$\boldsymbol{\alpha}_s$	0.2	0.2	0.1	0.1	0.2		$\boldsymbol{\lambda}_s$	1.5	1.0	1.0	1.5	1.0
α_k	$\boldsymbol{\alpha}_l$	0.8	0.8	0.8	0.9	0.9	λ_k	$\boldsymbol{\lambda}_l$	3.0	3.0	2.5	2.5	3.0
	$\boldsymbol{\alpha}_{sl}$	0.2	0.8	0.9	0.1	0.2		$\boldsymbol{\lambda}_{sl}$	3.0	0.5	1.0	3.0	0.1

5.1 Parameter Estimation

To calculate the Bayesian estimates we use vague prior distributions, considering all the hyperparameters approximately null. This choice is due to the fact that, for one hand we are dealing with simulated samples hence there is no available prior information, and for the other hand the main purpose is to compare the performance between classical and Bayesian methodologies. In Algorithm 1, we set nig =3100, with $b = 1100$ as burn-in period and $l = 20$, to reduce autocorrelation between MCMC samples. We consider 200 independent replicates.
A problem that occurs frequently when estimating INAR models by classic methodology is that the estimates for the parameters α_k are inadmissible, that is to say that $\alpha_k \notin (0, 1)$. In this study these samples are eliminated.
The performance of the estimation methods is illustrated in Tables 2 and 3 for two particular situations of the Poisson SUINAR(1) model. In Table 2 we consider the model $(\boldsymbol{\alpha}_{sl}, \boldsymbol{\lambda}_{sl})$ with parameters $\boldsymbol{\alpha}_{sl} : \alpha_1 = 0.2, \alpha_2 = 0.8, \alpha_3 = 0.9, \alpha_4 = 0.1, \alpha_5 = 0.2$, $\boldsymbol{\lambda}_{sl} : \lambda_1 = 3.0, \lambda_2 = 0.5, \lambda_3 = 1.0, \lambda_4 = 3.0, \lambda_5 = 0.1$ and $\delta = 2$ which is caracterized by both α_k and λ_k ranging from low to high values, meaning that the mean of the innovations varies among the individuals. Table 3 presents the estimation results for the model$(\boldsymbol{\alpha}_s, \boldsymbol{\lambda}_l)$ with parameters, $\boldsymbol{\alpha}_s : \alpha_1 = 0.2, \alpha_2 = 0.2, \alpha_3 = 0.1, \alpha_4 = 0.1, \alpha_5 = 0.2$, $\boldsymbol{\lambda}_l : \lambda_1 = 3.0, \lambda_2 = 3.0, \lambda_3 = 2.5, \lambda_4 = 2.5, \lambda_5 = 3.0$ and $\delta = 2$ which is caracterized by low values for the parameters α and high values for the innnovations for all the individuals, with small variation between individuals.

These results indicate that the method of moments (mm) provides better estimates for small values of α_k ($\alpha_k \leq 0.2$) whereas the maximum likelihood (ml) and Bayesian methodology (B) are more appropriate when the α_k parameter has large values ($\alpha_k \geq 0.8$); however, the Bayesian approach has the advantage of estimating δ, α_k and λ_k separately, which is not possible with the maximum likelihood. Regarding the estimation of λ_k the simulation results indicate that the Bayesian methodology has a better performance when the mean value of entrances is very different from individual to individual, but on contrary, if the differences between the mean values are small, the behavior is not so good. It can be noticed that the method of moments provides always poor estimates for λ_k. Moreover, the parameter δ is underestimated by both methods and the bias increases in the samples where the mean number of entrances differ between the individuals. Regarding the estimation of $\mu_k = \delta + \lambda_k$ the method of moments provides the estimates with smallest bias and whereas the maximum likelihood estimates are the most biased. It is important to note once again that μ_k is estimated as a parameter by ml while $\hat{\mu}_{k,mm} = \hat{\delta}_{mm} + \hat{\lambda}_{k,mm}$.

5.2 Prediction

In this section h-steps-ahead ($h = 1, 2, \ldots, 10$) point forecasts and prediction intervals are obtained using classic methodology, equations 14 and 15 and Bayesian methodology, equation 18 and Algorithm 3 to obtain HPD predictive intervals.
The performance of the forecasting methods is illustrated in Tables 4 and 5 for two particular Poisson SUINAR(1) models.
Table 4 displays forecasts for $x_{k,n+h}$, the jump between $x_{k,n}$ and $x_{k,n+h}$, and the squared errors between $\hat{x}_{k,n+h}$ and $x_{k,n+h}$, considering samples of sizes $n = 25$ and $n = 100$ simulated from the model with parameters ($\boldsymbol{\alpha}_s : \alpha_1 = 0.2, \alpha_2 = 0.2, \alpha_3 = 0.1, \alpha_4 = 0.1, \alpha_5 = 0.2$), ($\boldsymbol{\lambda}_s : \lambda_1 = 1.5, \lambda_2 = 1.0, \lambda_3 = 1.0, \lambda_4 = 1.5, \lambda_5 = 1.0$) and $\delta = 2$.
Table 5 presents similar results for samples generated from the model with parameters ($\boldsymbol{\alpha}_l : \alpha_1 = 0.8, \alpha_2 = 0.8, \alpha_3 = 0.8, \alpha_4 = 0.9, \alpha_5 = 0.9$), ($\boldsymbol{\lambda}_l : \lambda_1 = 3.0, \lambda_2 = 3.0, \lambda_3 = 2.5, \lambda_4 = 2.5, \lambda_5 = 3.0$) and $\delta = 2$.
Additionally Figure 1 presents absolute errors between predicted values and corresponding simulated values, regarding several samples of size 25 of SUINAR(1) model.
According to the present simulation study we can conclude that the results are independent of the prediction method and the methodology. Moreover, the observed prediction error depends on two factors: the jump between $x_{k,n}$ and $x_{k,n+h}$ for $h \leq 4$ and the proximity between $x_{k,n+h}$ and $(\hat{\lambda_k} + \hat{\delta})/(1 - \hat{\alpha_k})$ for large values of h ($h \leq 5$) (remark that $\lim_{h\to\infty} E(X_{k,n+h}|X_{k,n}) = (\lambda_k + \delta)/(1 - \alpha_k)$) (see Figure 1).
Several simulated examples indicate that the variability of the predictive function increases with the magnitude of α_k and λ_k, justifying that the predictions shown in Table 5 are worst than those in Table 4. Moreover it is worthwhile to mention that the values of $\hat{x}_{k,n+h}$ are constant for $h \geq 8$ (Table **??**) when α_k and λ_k are small. In contrast, these values are not constant when α_k and λ_k are large.
There is evidence that the confidence interval gets wider as h increases, as expected and converges to the asymptotic interval. However, the rate of convergence is higher for smaller values of α_k and λ_k.

Table 2: Estimates of $(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \delta)$ model with parameters $\boldsymbol{\alpha}_{sl} = (0.2, 0.8, 0.9, 0.1, 0.2)$, $\boldsymbol{\lambda}_{sl} = (3.0, 0.5, 1.0, 3.0, 0.1)$ and $\delta = 2$ (absolute value of bias in brackets)

		n=25			n=100		
k	α_k	$\hat{\alpha}_{k,mm}$	$\hat{\alpha}_{k,ml}$	$\hat{\alpha}_{k,B}$	$\hat{\alpha}_{k,mm}$	$\hat{\alpha}_{k,ml}$	$\hat{\alpha}_{k,B}$
1	0.2	0.230	0.334	0.256	0.183	0.250	0.197
		(0.03)	(0.13)	(0.06)	(0.02)	(0.05)	(0.00)
2	0.8	0.673	0.847	0.842	0.766	0.865	0.865
		(0.13)	(0.05)	(0.04)	(0.03)	(0.07)	(0.07)
3	0.9	0.794	0.919	0.918	0.873	0.924	0.924
		(0.11)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
4	0.1	0.177	0.275	0.224	0.118	0.174	0.125
		(0.08)	(0.18)	(0.12)	(0.02)	(0.07)	(0.03)
5	0.2	0.143	0.673	0.623	0.155	0.761	0.758
		(0.06)	(0.47)	(0.42)	(0.05)	(0.56)	(0.56)
k	λ_k	$\hat{\lambda}_{k,mm}$		$\hat{\lambda}_{k,B}$	$\hat{\lambda}_{k,mm}$		$\hat{\lambda}_{k,B}$
1	3.0	3.783		3.333	4.004		3.812
		(0.78)		(0.33)	(1.00)		(0.81)
2	0.5	2.445		0.685	1.659		0.782
		(1.95)		(0.19)	(1.16)		(0.28)
3	1.0	4.322		1.088	2.494		1.303
		(3.32)		(0.09)	(1.49)		(0.30)
4	3.0	3.548		3.016	3.845		3.695
		(0.55)		(0.02)	(0.85)		(0.70)
5	0.1	1.082		0.155	1.114		0.154
		(0.98)		(0.06)	(1.01)		(0.05)
k	μ_k	$\hat{\mu}_{k,mm}$	$\hat{\mu}_{k,ml}$	$\hat{\mu}_{k,B}$	$\hat{\mu}_{k,mm}$	$\hat{\mu}_{k,ml}$	$\hat{\mu}_{k,B}$
1	2.2	3.917	3.381	3.779	4.076	3.739	4.005
		(1.08)	(1.62)	(1.22)	(0.92)	(1.26)	(0.99)
2	2.8	2.581	1.097	1.130	1.730	0.977	0.975
		(0.08)	(1.40)	(1.37)	(0.77)	(1.52)	(1.52)
3	2.9	4.457	1.524	1.535	2.565	1.488	1.496
		(1.45)	(1.48)	(1.47)	(0.43)	(1.51)	(1.50)
4	2.1	3.683	3.241	3.462	3.916	3.667	3.889
		(1.32)	(1.76)	(1.54)	(1.08)	(1.33)	(1.11)
5	2.2	1.217	0.436	0.601	1.185	0.331	0.347
		(0.88)	(1.66)	(1.50)	(0.92)	(1.77)	(1.75)
		$\hat{\delta}_{mm}$		$\hat{\delta}_B$	$\hat{\delta}_{mm}$		$\hat{\delta}_B$
$\delta = 2$		0.135		0.446	0.071		0.193

Table 3: Estimates of $(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \delta)$ of SUINAR(1) model with parameters $\boldsymbol{\alpha}_s = (0.2, 0.2, 0.1, 0.1, 0.2)$, $\boldsymbol{\lambda}_l = (3.0, 3.0, 2.5, 2.5, 3.0)$ and $\delta = 2$ (absolute value of bias in brackets)

		n=25			n=100		
k	α_k	$\hat{\alpha}_{k,mm}$	$\hat{\alpha}_{k,ml}$	$\hat{\alpha}_{k,B}$	$\hat{\alpha}_{k,mm}$	$\hat{\alpha}_{k,ml}$	$\hat{\alpha}_{k,B}$
1	0.2	0.212	0.323	0.238	0.181	0.243	0.201
		(0.01)	(0.12)	(0.04)	(0.02)	(0.04)	(0.00)
2	0.2	0.217	0.320	0.243	0.196	0.267	0.215
		(0.02)	(0.12)	(0.04)	(0.00)	(0.07)	(0.02)
3	0.1	0.180	0.306	0.162	0.125	0.204	0.094
		(0.08)	(0.21)	(0.06)	(0.03)	(0.10)	(0.01)
4	0.1	0.183	0.310	0.167	0.119	0.189	0.088
		(0.08)	(0.21)	(0.07)	(0.02)	(0.09)	(0.01)
5	0.2	0.215	0.325	0.237	0.187	0.256	0.211
		(0.02)	(0.13)	(0.04)	(0.01)	(0.06)	(0.01)
k	λ_k	$\hat{\lambda}_{k,mm}$		$\hat{\lambda}_{k,B}$	$\hat{\lambda}_{k,mm}$		$\hat{\lambda}_{k,B}$
1	3.0	3.926		1.018	4.079		0.673
		(0.93)		(1.98)	(1.08)		(2.33)
2	3.0	3.948		1.050	3.999		0.603
		(0.93)		(1.95)	(1.00)		(2.40)
3	2.5	3.174		0.553	3.411		0.255
		(0.67)		(1.95)	(0.91)		(2.25)
4	2.5	3.150		0.526	3.432		0.272
		(0.65)		(1.97)	(0.93)		(2.23)
5	3.0	3.909		1.026	4.096		0.669
		(0.91)		(1.97)	(1.10)		(2.33)
k	μ_k	$\hat{\mu}_{k,mm}$	$\hat{\mu}_{k,ml}$	$\hat{\mu}_{k,B}$	$\hat{\mu}_{k,mm}$	$\hat{\mu}_{k,ml}$	$\hat{\mu}_{k,B}$
1	2.2	3.947	3.388	3.787	4.079	3.767	3.966
		(1.05)	(1.61)	(1.21)	(0.92)	(1.23)	(1.03)
2	2.2	3.968	3.447	3.819	3.999	3.646	3.897
		(1.03)	(1.55)	(1.18)	(1.00)	(1.35)	(1.10)
3	2.1	3.195	2.702	3.322	3.412	3.102	3.549
		(1.30)	(1.80)	(1.18)	(1.09)	(1.40)	(0.96)
4	2.1	3.171	2.675	3.294	3.433	3.159	3.566
		(1.33)	(1.83)	(1.21)	(1.07)	(1.34)	(0.93)
5	2.2	3.930	3.376	3.795	4.096	3.747	3.963
		(1.07)	(1.62)	(1.21)	(0.90)	(1.25)	(1.03)
		$\hat{\delta}_{mm}$		$\hat{\delta}_B$	$\hat{\delta}_{mm}$		$\hat{\delta}_B$
$\delta = 2$		0.0211		1.3832	0.0005		1.1759

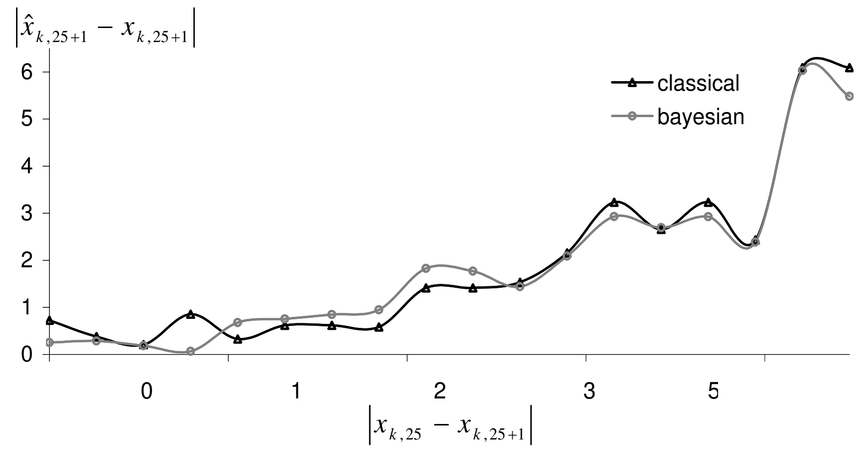

Figure 1: Values of $|\hat{x}_{k,25+1} - x_{k,25+1}|$ with different samples of SUINAR(1) model

.

6 Final Comments

In this work classical and Bayesian approaches to time series analysis and forecasting are applied to the SUINAR(1) models. Regarding the estimation of the model, the Bayesian approach has the advantage of allowing the estimation of all the parameters of the model. However, the two methodologies perform similarly regarding the forecasting of future values.

References

Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, Vol. 8, pp. 261-275.

Berglund, E. and Brännäs, K. (2001). Plant's entry and exit in Swedish municipalities. *The Annals of Regional Science*, Springer, Vol.35, p. 431-448.

Blundell, R., Griffith, R. and Windmeijer, F. (1999). Individual effects and dynamics in count data models. *Working Paper*, W99/3, Institute of Fiscal Studies, London.

Böckenholt, U. (1999) Mixed INAR(1) Poisson regression models: analyzing heterogeneity and serial dependencies in longitudinal count data. *Jounal of Econometrics*, Vol.89, p. 317-338

Chatfield, C. (2001). *Time Series Forecasting.* London: Chapman and Hall/CRC.

Chen, M-H. and Shao, Q-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *J. Comp. and Graphical Stat.*, Vol.8, p. 69-92.

Chen, M-H., Shao, Q-M and Ibrahim, J.G. (2000). *Monte Carlo methods in Bayesian computation.* Springer Series in Statistics.

Freeland, R.K. and McCabe, B.P.M. (2003). Forecasting discrete valued low count series. *International Journal of Forecasting*, Vol.20, p. 427-434.

Table 4: Forecasts for $x_{k,n+h}$ and values of square deviances ($DA^2 = (\hat{x}_{k,n+h} - x_{k,n+h})^2$) of SUINAR(1) model with initial values $\boldsymbol{\alpha}_s = (0.2, 0.2, 0.1, 0.1, 0.2)$, $\boldsymbol{\lambda}_s = (1.5, 1.0, 1.0, 1.5, 1.0)$ and $\delta = 2$.

		n=25					n=100				
			classical		bayesian			classical		bayesian	
h	k	jump	$\hat{x}_{k,n+h}$	DA^2	$\hat{x}_{k,n+h}$	DA^2	jump	$\hat{x}_{k,n+h}$	DA^2	$\hat{x}_{k,n+h}$	DA^2
	1	1	2.672	0.107	2.323	0.458	2	3.191	1.418	3.254	1.571
	2	0	2.272	0.530	2.744	0.066	3	2.864	3.474	3.180	4.752
1	3	0	2.618	0.146	2.712	0.083	1	2.328	0.452	2.233	0.588
	4	0	2.789	0.045	2.841	0.025	1	3.095	3.629	3.178	3.320
	5	3	1.340	7.076	1.307	7.252	0	2.721	0.078	2.812	0.035
	1	1	2.857	3.448	2.626	2.644	2	3.135	1.288	3.216	1.479
	2	2	2.213	1.471	2.679	2.819	2	2.514	0.264	2.876	0.767
2	3	1	2.528	0.279	2.641	0.411	1	2.359	1.847	2.324	1.753
	4	0	2.697	0.092	2.811	0.036	1	2.956	0.002	3.097	0.009
	5	1	1.523	0.228	1.539	0.213	1	2.617	0.381	2.743	0.552
	1	2	2.922	1.162	2.933	1.138	1	3.131	0.017	3.200	0.040
	2	1	2.208	3.211	2.552	2.097	1	2.373	0.393	2.575	0.181
4	3	1	2.502	0.252	2.697	0.486	2	2.362	2.683	2.460	2.372
	4	3	2.639	11.296	2.811	10.170	1	2.932	0.004	2.991	0.000
	5	2	1.675	1.756	1.812	1.411	1	2.564	2.062	2.721	1.636
	1	5	2.927	16.589	3.247	14.085	3	3.131	4.541	3.173	4.722
	2	0	2.208	0.627	2.472	0.279	3	2.358	1.844	2.509	2.277
8	3	1	2.500	0.250	2.781	0.610	0	2.363	0.132	2.443	0.196
	4	2	2.626	5.636	2.799	4.844	0	2.931	1.143	3.021	0.958
	5	0	1.731	0.534	2.114	1.241	1	2.556	0.309	2.705	0.497
	1	0	2.927	0.859	3.259	1.585	3	3.131	4.541	3.200	4.840
	2	2	2.208	7.795	2.626	5.636	0	2.358	0.696	2.429	2.468
10	3	0	2.500	0.250	2.755	0.060	1	2.363	0.406	2.452	0.300
	4	1	2.625	1.891	3.108	0.796	2	2.931	0.867	3.047	1.096
	5	0	1.735	0.540	2.157	1.339	1	2.556	2.085	2.731	1.610
	1		2.927					3.131			
	2		2.208					2.358			
∞	3		2.500					2.363			
	4		2.623					2.931			
	5		1.736					2.556			

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

Paulino, C. D., Amaral Turkman, M.A. and Murteira, B. (2003). *Estatística bayesiana*. Fundação Calouste Gulbenkian. Lisboa.

Silva, N. (2005a). *Análise Bayesiana de Séries Temporais de Valores Inteiros*. Phd thesis, Department of Mathematics, University of Aveiro.

Silva, I., Silva M.E., Pereira, I. and Silva, N. (2005b). Replicated INAR(1) Process. *Methodology and Computing in Applied Probability*, Vol.7, p. 517-542.

Silva, N., Pereira, I., Silva, M.E. (2006). Forecasting in INAR(1) models. *Cadernos de Matemática*, CM06; I-46, Department of Mathematics, University of Aveiro.

Tanner, M.A. (1996). *Tools for statistical inference*. 3rd ed. Springer Verlag, New York.

Table 5: Forecasts for $x_{k,n+h}$ and values of square deviances ($DA^2 = (\hat{x}_{k,n+h} - x_{k,n+h})^2$) of SUINAR(1) model with initial values $\boldsymbol{\alpha}_l = (0.8, 0.8, 0.8, 0.9, 0.9)$, $\boldsymbol{\lambda}_l = (3.0, 3.0, 2.5, 2.5, 3.0)$ e $\delta = 2$.

		n=25					n=100				
			classical		bayesian			classical		bayesian	
h	k	jump	$\hat{x}_{k,n+h}$	DA^2	$\hat{x}_{k,n+h}$	DA^2	jump	$\hat{x}_{k,n+h}$	DA^2	$\hat{x}_{k,n+h}$	DA^2
	1	2	20.585	2.002	20.173	3.338	1	23.667	5.443	24.375	2.641
	2	5	24.906	37.137	24.969	36.373	0	18.680	0.102	18.755	0.060
1	3	1	14.379	0.386	14.244	0.572	1	19.572	2.039	19.782	1.848
	4	0	38.856	0.733	37.932	0.005	1	24.611	0.151	24.677	0.104
	5	3	35.765	10.465	36.073	8.567	5	37.463	29.844	37.173	26.760
	1	1	20.992	3.968	20.254	1.573	1	22.811	1.414	23.876	0.015
	2	1	24.025	8.851	24.442	6.543	3	18.442	5.963	18.576	6.636
2	3	2	14.876	1.263	14.449	2.253	1	19.273	2.983	19.591	1.985
	4	7	39.266	32.879	37.863	50.937	3	25.156	3.400	25.227	3.144
	5	6	35.765	38.875	36.307	32.410	7	57.844	61.528	37.341	53.890
	1	1	21.476	0.227	20.686	0.099	0	21.908	9.560	23.224	3.154
	2	3	22.745	0.065	23.153	0.023	1	18.131	0.017	18.406	0.165
4	3	1	15.035	4.141	14.895	3.591	2	18.918	9.499	19.326	7.150
	4	5	39.557	11.854	37.774	27.311	6	26.076	15.398	26.395	12.996
	5	9	35.109	97.832	36.589	70.745	3	38.418	19.519	37.582	12.831
	1	1	21.826	0.682	21.049	0.002	2	21.381	31.573	22.387	21.280
	2	8	21.375	11.391	22.538	20.593	2	17.862	9.847	17.928	9.437
8	3	2	15.359	11.283	15.537	12.510	2	18.660	11.156	18.837	10.005
	4	4	39.640	5.570	37.370	21.437	5	27.391	2.589	27.647	1.758
	5	5	34.336	44.409	37.207	14.387	4	39.072	3.717	37.995	9.030
	1	2	21.882	0.014	20.714	1.654	6	21.317	5.368	22.037	9.223
	2	7	21.026	4.105	21.638	6.959	7	17.808	67.109	18.054	63.139
10	3	2	15.425	11.731	15.334	11.116	0	18.619	1.907	18.840	1.346
	4	3	39.643	1.841	37.658	11.169	3	27.855	0.731	28.230	1.513
	5	6	33.989	64.176	37.444	20.757	14	39.252	138.016	38.340	161.188
	1		21.937					21.273			
	2		20.384					17.738			
∞	3		15.510					18.577			
	4		39.645					29.664			
	5		29.221					39.640			