II Encontro Galaico-Português de Biometria
Santiago de Compostela, 30 de Junho, 1 e 2 de Julho de 2016

Dynamic fator Models for bivariate Count Data: an application to fire activity

Magda Monteiro[1], Isabel Pereira[2] e Manuel G. Scotto[3]

[1]ESTGA & CIDMA & Universidade de Aveiro
[2]Departamento de Matemática & CIDMA& Universidade de Aveiro
[3]Departamento de Matemática/IST &CEMAT & Universidade de Lisboa

ABSTRACT

The study of forest fire activity, in its several aspects, is essencial to understand the phenomenon and to prevent environmental public catastrophes. In this context the analysis of monthly number of fires along several years is one aspect to have into account in order to better comprehend this tematic. The goal of this work is to analyze the monthly number of forest fires in the neighboring districts of Aveiro and Coimbra, Portugal, through dynamic factor models for bivariate count series. We use a bayesian approach, through MCMC methods, to estimate the model parameters as well as to estimate the common latent factor to both series.

Keywords and key sentences: Dynamic latent variables; MCMC; bivarite Poisson distribution; wildfires.

1. INTRODUCTION

Wildefires are a very common phenomenon in Portugal and each year, a large amount of forest land is destroyed. In the years between 1996 and 2014, the average of fires were around 19 000 and the annual average of burned forest area was around 108 000 ha ([1]). The scale of these values is related with climate conditions, topography and also with a combination of various socio-economic factors including the large scale replacement of native Portuguese forest by fire-prone tree species (such as pine and eucalypt), which are more profitable, and the decline in traditional practices like grazing and forestry which helped to reduced the accumulation of flammable materials ([5],[4]). Is in this context that is extremely important to study forest fire activity in Portugal in its different aspects in which is included the study of time series that result from this activity such as monthly number of fires.
This work analizes the monthly number of forest fires in the neighboring districts of Aveiro and Coimbra, Portugal. Since we have two neighbour regions in the study, we can use bivariate models that allow potentially complex interdependence between series. As these data have a discrete nature we have to use appropriate multivariate models to describe them. The work in [2] presented a dynamic factor multivariate model to analyze Stock-Market Trading Activity. In this model, counts are assumed to be conditionally independently distributed with a discrete distribution whose averages are latent random variables which depend on a set of unknown factors which are autoregressive gaussian processes of order 1. In this work it will be used a bivariate dynamic model with one common latent factor and the temperature as an explanatory variable.

2. BIVARIATE DYNAMIC FACTOR MODEL

Consider a bivariate vector of counts $Y_t = (Y_{t,1}, Y_{t,2})$, recorded at time t (1- Aveiro, 2- Coimbra). In a dynamic multivariate count model the dynamics is introduced at the level of the latent factors. Therefore counts are assumed to be conditionally independently distributed with Poisson distribution

$$p(y_{t,j}|\lambda_{t,j}) = e^{-\lambda_{t,j}} \frac{\lambda_{t,j}^{y_{t,j}}}{y_{t,j}!}, \ t = 1 \cdots, T; \ j = 1, 2 \tag{1}$$

whose means $\lambda_{t,j}$ are latent random variables. Having into account the sample autocorrelation functions in Figure , it is also also assumed that the logarithm of each component of the mean vector, $\boldsymbol{\lambda}_t = (\lambda_{t,1}, \lambda_{t,2})$, is a linear function of the type

$$\ln(\lambda_{t,j}) = \mu_{t,j} + \gamma_j Z_t + \beta_j X_{t,j}, \quad j = 1, 2, \tag{2}$$

where Z_t is an unknown random factor which is common to both components of the vector of counts, $X_{t,j}$ is the monthly average temperature in region j and $\mu_{t,j}$ is the intercept which depends on the month of the year, $\mu_{t,j} = \mu_{s,j} \times I(t = s + 12k)$, $s = 1, 2, \cdots, 12$ and some natural k. It was considered that $\gamma_1 = 1$ for normalization and to eliminate indeterminacies in the factor scale.

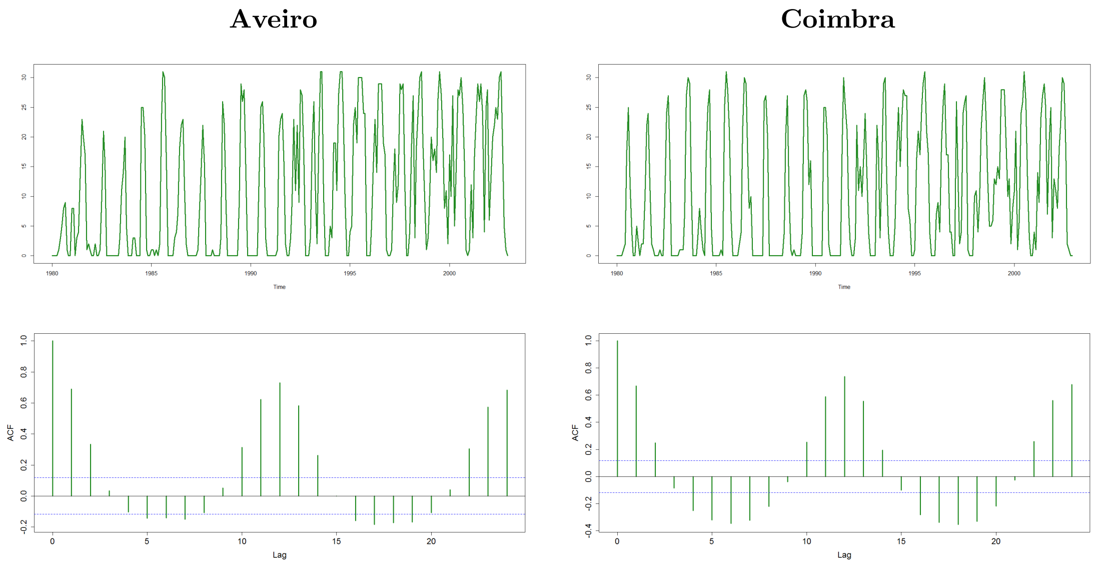

Figure 1: Monthly number of fires in Aveiro and Coimbra (top) and the respective sample autocorrelation functions (bottom).

The possibility of serial and cross-correlation in the counts is introduced through the structure of the latent factor Z_t. It is assumed that the factor follows an AR(1) Gaussian process, $Z_t|Z_{t-1} \sim N(\phi Z_{t-1}, \sigma^2)$.

The estimation of the model can be made either in a classical framework, using maximum likelihood estimation based uppon efficient importance sample, or through a bayesian approach, which will be considered here. Considering $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}, \boldsymbol{\beta})$ and $\boldsymbol{\theta}_2 = (\phi, \sigma)$, the vector of all parameters involved and $h(\boldsymbol{\theta})$ its prior distribution, and having into account that (Z_t) is an unknown Gaussian process, the prior distribution for the unknown components of the model can be written as $h(\boldsymbol{\theta}, \boldsymbol{z}) = h(\boldsymbol{\theta})h(\boldsymbol{z}|\boldsymbol{\theta}_2)$, so the posterior distribution of the unknown components can be obtained

$$h(\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{y}) = \frac{h(\boldsymbol{\theta}) \prod_{j=1}^{2} \prod_{t=1}^{T} f(z_1) f(z_t|z_{t-1}, \boldsymbol{\theta}_2) p(y_{t,j}|\lambda_{t,j})}{\int \cdots \int h(\boldsymbol{\theta}) \prod_{j=1}^{2} \prod_{t=1}^{T} f(z_1) f(z_t|z_{t-1}, \boldsymbol{\theta}) p(y_{t,j}|\lambda_{t,j}) d\boldsymbol{\theta} d\boldsymbol{z}}. \tag{3}$$

Since this distribution does not have a closed form, it will be use Markov Chain Monte Carlo methods (MCMC).

3. APPLICATION TO FIRE ACTIVITY

The bivariate dynamic factor model for counts introduced in Section 2 is used to analyze the dynamics and correlations of the monthly number of fires in Aveiro and Coimbra Counties in the period of january 1980 to December 2002. Although the data on the monthly of fires are known for a longer period of time, there are no available data about the temperature in the remaining period. The descriptive statistics are presented in Table 1 and there exists a positive correlation between series of counts. The estimation of the parameters of the model were made using a bayesian

	Aveiro	**Coimbra**
Mean	10.62	10.15
Median	5	6.5
standard dev.	11.49	10.66
Minimum	0	0
Maximum	31	31
correlation	0.74	

Table 1: Descriptive statistics for the number of fires

approach, through the use of MCMC methods with Metropolis-Hastings algorithm, using *Winbugs* software. Since we do not have prior information about parameters we used non-informative priors for them. We considered that parameters were independent, and for σ^2 we used the inverse Gamma with both parameters 0.001 and for the remaining parameters we used a normal prior distribution with a mean zero and a variance of 10000. A burnin period of 150000 were considered in order to the chain reaches the steady state (we use the package coda in R software), and a lag of 1000 to form a sample, with dimension 1000, from the posterior distribution. The results of the estimation procedure are presented in Table 2 and Figure 2 presents the predictions of the latent factor with 95% HPD intervals. The standardized Pearson residuals (see Table 3) do not present significant correlation but they present values of the means and standard deviations far from the expected values, i.e., the characteristics of mean zero and standard deviation 1. It was also analized the quantile quantile plot of randomized PIT residuals that revealed deviations from normality in the left tail. In order to improve the modeling results one can use other discrete distribution such as negative binomial, since dispersion index is equal to 1 in the Poisson distribution and also one can adjust the structure of the factor Z_t so it can capture all the dynamics in the series.

	Parameters						
City	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7
Aveiro	0.07134	0.9322	1.7120	1.6400	1.997	3.344	4.096
	(0.01878)	(0.02167)	(0.0166)	(0.0225)	(0.01933)	(0.0199)	(0.02052)
Coimbra	0.8402	2.0760	2.1910	2.1510	3.3700	4.786	5.4460
	(0.02167)	(0.02131)	(0.0249)	(0.01763)	(0.02338)	(0.02668)	(0.02951)
	μ_8	μ_9	μ_{10}	μ_{11}	μ_{12}	γ	β
Aveiro	4.239	3.842	2.269	-0.4684	-1.124	1	-0.0499
	(0.02066)	(0.01973)	(.01912)	(0.01898)	(0.01956)	–	(0.0005)
Coimbra	5.337	4.403	2.947	0.3264	-0.0.9631	0.9455	-0.1019
	(0.0293)	(0.0273)	(0.02455)	(0.02105)	(0.02349)	(0.00228)	(0.0012)
	factor parameters						
	ϕ	σ			**DIC**		
	0.8247	0.8585			2652.570		
	(0.0012)	(0.0020)					

Table 2: Parameter estimates (MC errors in brackets).

ACKNOWLEDGMENT

This work was supported in part by the Portuguese Foundation for Science and Technology (FCT-Fundação para a Ciência e a Tecnologia), through CIDMA - Center for Research and Development in Mathematics and Applications, within project UID/MAT/04106/2013.

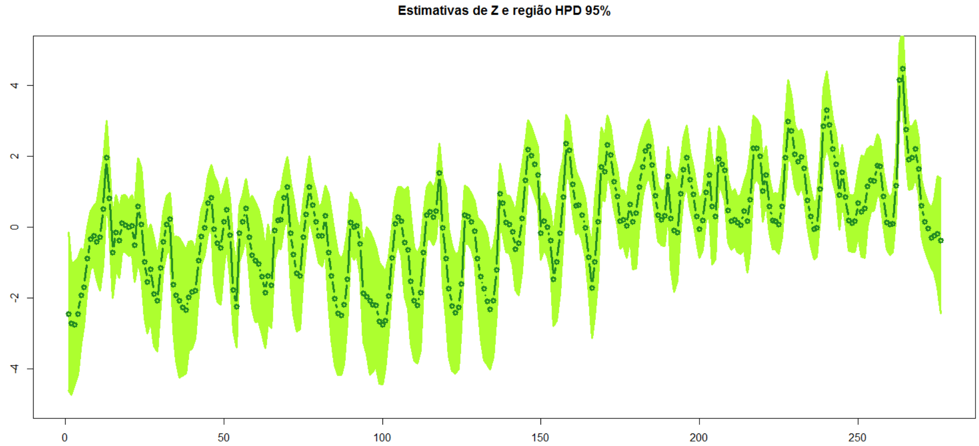

Figure 2: Predictions of factor Z and the respective 95% HPD interval.

Distribution	**Region**	**Mean**	**Standard deviation**	**Portmanteu test p-value**
Poisson	**Aveiro**	-0.105	0.8351	0.082
	Coimbra	-0.104	0.9276	0.239

Table 3: statistics of standardized Pearson residuals and Portmanteu test

References

[1] Instituto da Conservação da Natureza e das Florestas (2016). Estatística - dados sobre incêndios florestais.
http://www.icnf.pt/portal/florestas/dfci/inc/estat-sgif#tip (08/04/2016)

[2] Jung, R.C., Liesenfeld, R., Richard, J.F. (2011). Dynamic factor models for multivariate count data: an application to stock-market trading activity. *Journal of Business & Economic Statistics*, 29, 73–85.

[3] Paulino, C.D., Amaral Turkman, A., Murteira, B. (2003). *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa.

[4] Pereira, J.M.C., Carreiras, J.M.B., Silva, J.M.N., Vasconcelos, M.J., 2006. Alguns conceitos básicos sobre os fogos rurais em Portugal. In: Pereira, J.S., Pereira, J.M.C., Rego, F.C., Silva, J.M.N., Silva, T.P. (Eds.), Incêndios Florestais em Portugal: Caracterização, Impactes e Prevenção. ISAPress, Lisboa, pp. 133–161.

[5] Fernandes, Paulo (2007). Entender porque arde tanto a floresta em Portugal. In Árvores e Florestas de Portugal, Vol. 8 Proteger a Floresta - Incêndios, Pragas e Doenças. Chapter: 3, Publisher: Público, Comunicação Social, SA, & Fundação Luso-Americana para o Desenvolvimento, Editors: Silva J.S., pp.69–91.