◆◆ Scientific
◆◆ Research

A Gibbs Sampling Algorithm to Estimate the Parameters of a Volatility Model: An Application to Ozone Data

Verónica De Jesús Romo[1*], Eliane R. Rodrigues[2], Guadalupe Tzintzun[3]

[1]Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico
[2]Instituto de Matemáticas, Universidad Nacional Autónoma de México, Mexico City, Mexico
[3]Instituto Nacional de Ecología y Cambio Climático, Secretaría de Medio Ambiente y Recursos Naturales, Mexico City, Mexico
Email: *veronica.dejesus@ciencias.unam.mx, eliane@math.unam.mx, tzintzun@ine.gob.mx

ABSTRACT

In this work we consider a stochastic volatility model, commonly used in financial time series studies, to analyse ozone data. The model considered depends on some parameters and in order to estimate them a Markov chain Monte Carlo algorithm is proposed. The algorithm considered here is the so-called Gibbs sampling algorithm which is programmed using the language R. Its code is also given. The model and the algorithm are applied to the weekly ozone averaged measurements obtained from the monitoring network of Mexico City.

1. Introduction

We are going to split this section into two parts. One containing a general literature review and another focussing on the description of what happens in Mexico City.

1.1. General Literature Review

It is a well know fact that when exposed to high levels of pollution, human beings may present serious health issues (see for instance [1-3]). In particular, if ozone concentration levels stay above 0.11 parts per million (0.11 ppm) for a long period of time, then a very sensitive part of the population may present a deterioration in their health (see for instance [4-7]).

Since exposure to high levels of pollution can cause serious damage to people's health as well as the environment in general, there are several works in the literature that deal with the problem of modelling the behaviour of pollutants in general and in particular, ozone. Among them, we may quote the early work [8] considering extreme value theory. We also have [9] using time series analysis; [10] and [11] considering Markov chain models; [12] with an analysis of the behaviour the maximum measurements of ozone with an application to the data from Mexico City; [13,14] using homogeneous Poisson processes, and [15,16] using non-homogeneous Poisson processes (for several Markov and Poisson models applied to air pollution problems see for instance

[17]). [18-20] present studies that analyse the impact on air quality of the driving restriction imposed in Mexico City.

Besides keeping track of the level of the daily maximum concentration of a pollutant and checking for a decreasing trend in the measurements, it might also be of importance for the environmental authorities to keep track of the degree of variability of the pollutant's concentration. That is important because, for instance, it may help to see how the variability behaves after some measures aiming to reduce its level are taken. A decreasing of the daily concentration might occur, but it is also interesting to know if the variability also decreases. Following in that direction we have, for instance, [21,22] and [23] where some univariate and bivariate stochastic volatility models are used to study the weekly averaged ozone concentration in Mexico City. Also, in [23] and [24] we have the use of multivariate stochastic volatility models to study the behaviour of five pollutants present in the city of São Paulo, Brazil.

Many more works may be found in the literature. Some of those works consider spatial models and neural networks, among other methodologies. Some of them consider different pollutants as well. In this work we are going to consider only ozone since this pollutant is the one still causing problems in Mexico City. In the next section we describe the situation in Mexico City, the problem to be studied as well as the methodology considered to study it.

*Corresponding author.

1.2. Description of the Problem in Mexico City

During the past twenty five years, environmental authorities in Mexico have been implementing several measures in order to decrease the level of pollution in several cities throughout the country. In particular, in Mexico City, those measures have caused an improvement in the air quality. For instance, from 1997 to 2005 the maximum hourly ozone measurements decreased by 30%. Moreover, during the period of time ranging from 2008 to 2010, the mean concentration of the daily maximum measurements did not surpassed the value 0.1 ppm. Taking into account that the Mexican environmental standard for ozone is 0.11 ppm ([25]) and individuals should not be exposed, on average for a period of one hour or more once a year, this value for the mean is not so bad. However, that does not mean that high levels do not occur. In order to see that, note that during the period from 01 January 2008 to 31 December 2010, there were 491 days with the daily maximum measurement surpassing the value 0.11 ppm, 22 days when the threshold 0.17 ppm was surpassed, and one day of surpassings of the threshold of 0.2ppm used to declare emergency alerts in Mexico City.

Figure 1 shows the plots of the daily maximum ozone measurements obtained during the period ranging from 01 January 2008 to 31 December 2010. The horizontal line indicates the Mexican ozone standard of 0.11 ppm.

Looking at **Figure 1**, we may see that even though the mean of the measurements in all regions is below 0.11 ppm, we may still have many days in which the daily maximum measurement is well above 0.11.

As seen in **Figure 1**, it would be useful to analyse how the ozone measurements variability have behaved in more recent years. In here we are analysing this variability. Once we are able to model it we will able to estimate the frequency at which emergency situations may occur. In the present paper, in order to analyse the behaviour of this variability, we also consider stochastic volatility models to analyse the ozone data set collected from the monitoring network of the Metropolitan Area of Mexico City (www.sma.df.gov.mx/simat/) corresponding to three years of the daily maximum ozone measurements. The advantage of using stochastic volatility models to analyse time series is that they assume the existence of two processes modelling the series. One modelling the observations and another modelling the latent volatility.

The model considered here is a particular case of the multivariate case considered in [23] and [24]. The difference here is that instead of using five pollutants we are going to concentrate only on ozone measurements obtained in five regions of Mexico City. Another difference is that in previous works, estimation of the parameters of the model was performed using the Gibbs sampling algorithm internally implemented in the software WinBugs ([26]). That poses some difficulties such as slow con-
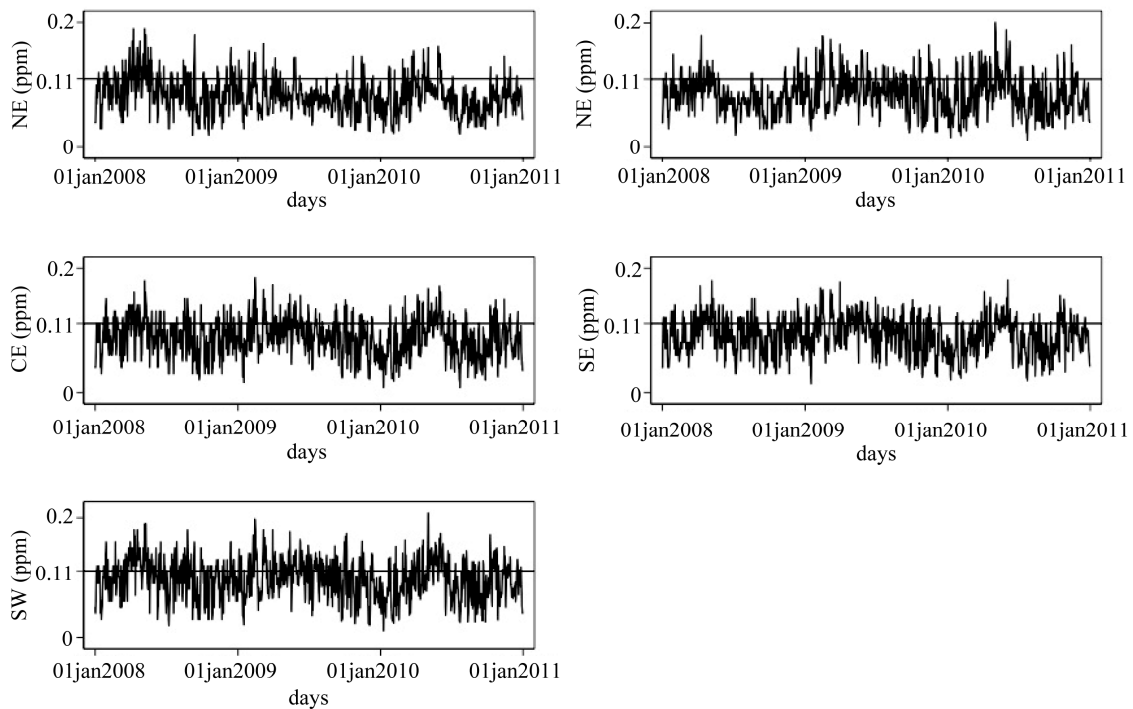

Figure 1. Daily maximum ozone measurements for regions NE, NW, CE, SE and SW for the period ranging from 01 January 2008 to 31 December 2010.

vergence of the algorithm and the use of very specific and very informative prior distributions. That was responsible for the need of running the algorithm several times until the right values for the hyperparameters were obtained. In here, the Gibbs sampling algorithm is programmed in R. The convergence was attained more rapidly and the prior distributions, even though informative, were not so difficult to adjust. Hence, another aim here is to present the Gibbs sampling algorithm, and its code, used to estimate the parameters of the stochastic volatility model and apply it to the case of ozone data from Mexico City.

This paper is organised as follows. Section 2 presents the stochastic volatility model considered here as well as the vector of parameters that need to be estimated. A Bayesian formulation used to estimate the parameters of the model considered in Section 2 is given in Section 3. In Section 4 the model and parameters estimation described in Sections 2 and 3 are applied to the data set obtained from the monitoring network of Mexico City. Section 5 presents some remarks and discussion of the results obtained. Finally, in Section 6 we conclude. An Appendix with the computer code used in the estimation of the parameters of the model is given after the list of reference.

2. A Stochastic Volatility Model

Stochastic volatility models have been used, in general, to analyse the variance of time series of financial returns (see for instance [27-30]). Until then, the existing methodology considered to model the volatility could be characterised by those used to model the variance based on function of past observations of the series (for instance the ARCH—auto regressive conditional heteroscedastic-models proposed by [31]) and of the delayed values of the series (GARCH—generalised auto regressive conditional heteroscedastic—models proposed by [32]). Volatility models are characterised by modelling the volatility of the times series in function of a non observable stochastic process. Hence, the main difference between the two approaches is that models based on GARCH assume that the volatility is observed in a period ahead in time, while in the stochastic volatility models, the volatility is a latent variable that cannot be observed without an error.

In this section the stochastic volatility model used to study the behaviour of weekly average measurements of the ozone in several regions of Mexico City is described. Many types of multivariate stochastic volatility models may be found in the literature (see for example [30]). The one considered here may be described as follows. Let $N \geq 1$ and $T > 0$ be fixed known integer and real numbers, respectively. (In here, N will represent the number of weeks in the observational period $[0,T]$

whose weekly average measurements of ozone we are considering.) Also, let $K > 0$ be an integer number recording the number of different regions where measurements are taken. Hence, for $X_j(t), t \in [0,T]$ the daily maximum ozone concentration measured on region j on the tth day, $j = 1,2,\cdots,K$, define

$$Z_j(t), t = 1,2,\cdots,N \text{ by } Z_j(1) = (1/7)\sum_{t=1}^{7} X_j(t),$$

$Z_j(2) = (1/7)\sum_{t=8}^{14} X_j(t)$, and so on. Therefore, $Z_j(t)$ is the weekly averaged ozone concentration in region j in the tth week, $t = 1,2,\cdots,N, j = 1,2,\cdots,K$.

Denote by $Y_j(t), t = 2,3,\cdots,N$ the log-return series defined by $\log\left[Z_j(t)/Z_j(t-1)\right], j = 1,2,\cdots,K$. Let

$$Y(t) = \left(Y_1(t), Y_2(t), \cdots, Y_K(t)\right)', t = 1,2,\cdots,N$$

be a K-dimensional vector whose coordinates are the K times series $Y_i(t), i = 1,2,\cdots,K$. The vector $Y(t), t = 1,2,\cdots,N$ is known as the log-return vector of the series $Z_j(t), j = 1,2,\cdots,K, t = 1,2,\cdots,N$. The set $D = \left\{Y(1), Y(2), \cdots, Y(N)\right\}$ will denote the set of observed data.

As in [23] and [24], we assume that $Y(t)$ may be written as $Y(t) = H(t)\epsilon(t)$, where $\epsilon(t) = \left(\epsilon_1(t), \epsilon_2(t), \cdots, \epsilon_K(t)\right)$ is an error vector and $H(t)$ is a $K \times K$ diagonal matrix given by $H(t) = \text{diag}\left(e^{h_1(t)/2}, e^{h_2(t)/2}, \cdots, e^{h_K(t)/2}\right)$, with $h(t) = \left(h_1(t), h_2(t), \cdots, h_K(t)\right), t = 1,2,\cdots,N,$ a vector of latent variables to be specified later. Therefore, we may write $Y_j(t), j = 1,2,\cdots,K$, as $Y_j(t) = e^{h_j(t)/2}\epsilon_j(t)$.

Also, we assume that $h(t), t = 1,2,\cdots,N$ is such that its coordinates follow an AR(1) model, *i.e.*, for $j = 1,2,\cdots,K$

$$\begin{cases} h_j(1) = \mu_j + \eta_j(1) \\ h_j(t) = \mu_j + \phi_j\left[h_j(t-1) - \mu_j\right] + \eta_j(t), t = 2,3,\cdots,N \end{cases}$$

where $-1 < \phi_j < 1$, and $\eta(t) = \left(\eta_1(t), \eta_2(t), \cdots, \eta_K(t)\right)$ is assumed to have a multivariate Normal distribution with mean vector $\mathbf{0} = (0,0,\cdots,0)$ and variance-covariance matrix Σ_η the diagonal matrix $\text{diag}\left(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2, \cdots, \sigma_{\eta_K}^2\right)$.

Remark. Note that by definition we have, for $j = 1,2,\cdots,K$, that $h_j(1)$ has a Normal distribution $N\left(\mu_i, \sigma_{\eta_j}^2\right)$ and that given $h_j(t-1)$, the variable $h_j(t)$ has a Normal distribution $N\left(\mu_j + \phi_j\left[h_j(t-1) - \mu_j\right], \sigma_{\eta_j}^2\right), t = 2,3,\cdots,N$.

We also assume that $\epsilon(t)$ has a multivariate Normal distribution with mean vector $\mathbf{0} = (0,0,\cdots,0)$ and vari-

ance-covariance matrix $\Sigma_\epsilon = \left(\rho_{ij}^\epsilon\right)_{i,j=1,2,\cdots,K}$, where

$\rho_{ii}^\epsilon = 1$ is the variance of $\epsilon_i(t)$, and ρ_{ij}^ϵ is the covariance of $\epsilon_i(t)$ and $\epsilon_j(t), i, j = 1, 2, \cdots, K, i \neq j$, $t = 1, 2, \cdots, N$. Therefore, by definition, we have that given $\boldsymbol{h}(t)$, the vector $\boldsymbol{Y}(t)$ has a multivariate Normal distribution with mean vector $\boldsymbol{0}$ and variance-covariance matrix given by $\Sigma_Y = \left(\rho_{ij}^Y\right)_{i,j=1,2,\cdots,K}$, where $\rho_{ii}^Y = e^{h_i(t)}$

and $\rho_{ij}^Y = \rho_{ij}^\epsilon e^{\left(h_i(t)+h_j(t)\right)/2}, i, j = 1, 2, \cdots, K, i \neq j$.

Remark. Note that since the volatility of the series $Y(t), t = 1, 2, \cdots, N$, is by definition $\sigma^2(t) = E\left[Y^2(t) \mid Y(1), Y(2), \cdots, Y(t-1)\right]$, the vector $\boldsymbol{h}(t) = \left(h_1(t), h_2(t), \cdots, h_K(t)\right)$ is the vector of the logarithm of the volatility of the series studied.

In the present work we will be considering the case where $\rho_{ij}^\epsilon = 0$ for $i \neq j, i, j = 1, 2, \cdots, K$ *i.e.*, we assume independence among the coordinates of the vector $\boldsymbol{\epsilon}$. Hence, now $\Sigma_\epsilon = \text{diag}(1, 1, \cdots, 1)$. That is, we are considering only Model I given in [24]. This means that the problem can be viewed as having K times series which can be analysed separately.

Therefore, the joint density function of $\boldsymbol{Y}(t)$ given $\boldsymbol{h}(t)$ is

$$f\left(\boldsymbol{Y}(t) \mid \boldsymbol{h}(t)\right)$$
$$= (2\pi)^{-K/2} \exp\left\{-\frac{1}{2}\sum_{j=1}^K \left[h_j(t) + e^{-h_j(t)} Y_j^2(t)\right]\right\}. \quad (1)$$

This means that given $\boldsymbol{h}(t) = \left(h_1(t), h_2(t), \cdots, h_K(t)\right)$, the vector of the log-returns $\boldsymbol{Y}(t) = \left(Y_1(t), Y_2(t), \cdots, Y_K(t)\right)$ has a Normal distribution with mean vector $\boldsymbol{0} = (0, 0, \cdots, 0)$ and variance-covariance matrix $\boldsymbol{\Sigma}_Y = \text{diag}\left(e^{h_1(t)}, e^{h_2(t)}, \cdots, e^{h_K(t)}\right)$.

The vector of parameters to be estimated is $\boldsymbol{\theta} = \left(\boldsymbol{\mu}; \boldsymbol{\phi}; \boldsymbol{\sigma}_{\boldsymbol{\eta}}^2\right)$, where $\boldsymbol{\mu} = \left(\mu_1, \mu_2, \cdots, \mu_K\right)'$, $\boldsymbol{\phi} = \left(\phi_1, \phi_2, \cdots, \phi_K\right)'$ and $\boldsymbol{\sigma}_{\boldsymbol{\eta}}^2 = \left(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2, \cdots, \sigma_{\eta_K}^2\right)'$. (In here, v' indicates the vector v transposed.)

3. Bayesian Formulation of the Models

In this section the expression for the complete conditional marginal density functions of the parameters, that are used in the Gibbs sampling algorithm, are presented. We also determine the prior distribution of the parameters of the model. Therefore, define $\boldsymbol{\varphi} = (\boldsymbol{\theta}, \boldsymbol{h})$, where $\boldsymbol{h} = \left(\boldsymbol{h}(1), \boldsymbol{h}(2), \cdots, \boldsymbol{h}(N)\right)$. The aim is to obtain an expression for the posterior distribution $P(\boldsymbol{\varphi} \mid \boldsymbol{D})$ and from it obtain the estimated parameters. Note that we have that $P(\boldsymbol{\varphi} \mid \boldsymbol{D}) \propto L(\boldsymbol{D} \mid \boldsymbol{\varphi}) P(\boldsymbol{h} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta})$, where $L(\boldsymbol{D} \mid \boldsymbol{\varphi})$ is the likelihood function of the model,

$P(\boldsymbol{h} \mid \boldsymbol{\theta})$ and $P(\boldsymbol{\theta})$ are the conditional distribution of the latent variables given the vector of parameters and the prior distribution of the vector of parameters, respectively.

Note that the likelihood function $L(\boldsymbol{D} \mid \boldsymbol{\varphi})$ is given by

$$L(\boldsymbol{D} \mid \boldsymbol{\varphi}) \propto \exp\left(-\frac{1}{2}\sum_{t=1}^N \sum_{j=1}^K \left[h_j(t) + e^{-h_j(t)} Y_j^2(t)\right]\right). \quad (2)$$

Additionally,

$$P(\boldsymbol{h} \mid \boldsymbol{\theta}) = f\left(\boldsymbol{h}(1) \mid \boldsymbol{\theta}\right) \prod_{t=2}^N f\left(\boldsymbol{h}(t) \mid \boldsymbol{h}(t-1), \boldsymbol{\theta}\right).$$

By hypothesis, we have that ([23] and [24])

$$f\left(\boldsymbol{h}(1) \mid \boldsymbol{\theta}\right) \propto \prod_{j=1}^K \left(\sigma_{\eta_j}^2\right)^{-1/2} \exp\left(-\frac{1}{2\sigma_{\eta_j}^2}\left[h_j(1) - \mu_j\right]^2\right).$$

and that for $t = 2, \cdots, N$,

$$f\left(\boldsymbol{h}(t) \mid \boldsymbol{h}(t-1), \boldsymbol{\theta}\right) \propto \prod_{j=1}^K \left(\sigma_{\eta_j}^2\right)^{-\frac{1}{2}}$$
$$\cdot \exp\left(-\frac{1}{2\sigma_{\eta_j}^2}\left[h_j(t) - \mu_j - \phi_j\left(h_j(t-1) - \mu_j\right)\right]^2\right).$$

Hence,

$$P(\boldsymbol{h} \mid \boldsymbol{\theta}) \propto \left[\prod_{j=1}^K \left(\sigma_{\eta_j}^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_{\eta_j}^2}\left[h_j(1) - \mu_j\right]^2\right)\right]$$
$$\prod_{t=2}^N \left[\prod_{j=1}^K \left(\sigma_{\eta_j}^2\right)^{-\frac{1}{2}} \quad (3)$$
$$\exp\left(-\frac{1}{2\sigma_{\eta_j}^2}\left[h_j(t) - \mu_j - \phi_j\left(h_j(t-1) - \mu_j\right)\right]^2\right)\right].$$

The choice of prior distributions for the parameters was based on information obtained from previous studies ([22-24]) and also from the behaviour presented by the data.

The parameters ϕ_j, μ_j and $\sigma_{\eta_j}^2$ will have as their prior distributions a Normal distribution Normal $\left(a_j, b_j\right)$, a Normal $\left(e_j, f_j\right)$ and an inverse Gamma IG $\left(c_j, d_j\right) j = 1, \cdots, K$. The hyperparameters $a_j, b_j, c_j, d_j, e_j, f_j, j = 1, \cdots, K$ will be considered known and will be specified later. (In here, IG (a, b) is the inverse Gamma distribution with mean $b/(a-1)$ and variance $b^2/\left[(a-1)^2(a-2)\right], a > 2$.)

The latent variables $\boldsymbol{h}(t)$ will be considered as parameters to be estimated and will have as their prior distribution a Normal $\left(\mu_j, \sigma_{\eta_j}^2\right)$ for $t = 1$, and for $t = 2, 3, \cdots, N$ its prior distribution is a Normal

$$\left(\mu_j + \phi_j\left(h_j\left(t-1\right)-\mu_j\right), \sigma_{\eta_j}^2\right), j = 1, 2, \cdots, K .$$

Therefore, we have from (2), (3) and the forms of the prior distributions of the parameters, that the complete posterior distribution $P\left(\boldsymbol{\varphi}\middle|\boldsymbol{D}\right)$ is given by

$$P\left(\boldsymbol{\varphi}\middle|\boldsymbol{D}\right) \propto L\left(\boldsymbol{D}\middle|\boldsymbol{\varphi}\right)P\left(\boldsymbol{h}\middle|\boldsymbol{\theta}\right)P\left(\boldsymbol{\theta}\right)$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{N}\sum_{j=1}^{K}\left[h_j\left(t\right)+\mathrm{e}^{-h_j\left(t\right)}Y_j^2\left(t\right)\right]\right\}$$

$$\times\left(\prod_{j=1}^{K}\left(\sigma_{\eta_j}^2\right)^{-1/2}\exp\left\{-\frac{1}{2\sigma_{\eta_j}^2}\left(h_j\left(1\right)-\mu_j\right)^2\right\}\right)$$

$$\times\left[\prod_{t=2}^{N}\left(\prod_{j=1}^{K}\left(\sigma_{\eta_j}^2\right)^{-1/2}\right.\right. \tag{4}$$

$$\left.\left.\exp\left\{-\frac{1}{2\sigma_{\eta_j}^2}\left[h_j\left(t\right)-\mu_j-\phi_j\left(h_j\left(t-1\right)-\mu_j\right)\right]^2\right\}\right)\right]$$

$$\times\left(\prod_{j=1}^{K}\mathrm{e}^{-\frac{1}{2f_j}\left(\mu_j-e_j\right)^2}\mathrm{e}^{-\frac{1}{2b_j}\left(\phi_j-a_j\right)^2}\left(\sigma_{\eta_j}^2\right)^{1-c_j}\mathrm{e}^{-d_j/\sigma_{\eta_j}^2}\right).$$

Due to the complexity of (4), estimation of the parameters of the volatility models may not be performed using usual methods. Hence, Markov chain Monte Carlo algorithms will be used. In particular, a Gibbs sampling algorithm (see for instance [33]) programmed using R is going to be used to obtain a sample of the parameters and with that estimates for them.

Therefore, we need the expression for the complete marginal conditional distribution for each parameter. Additionally, in some cases, in order to obtain a sample from those complete conditional marginal distributions, we will need to use a Metropolis-Hastings algorithm (see for instance [34-36]) step inside the Gibbs sampling. Therefore, whenever necessary, the acceptance probability for the Metropolis-Hastings algorithm is also going to be presented.

Hence, let $\boldsymbol{\varphi}-\{a\}$ indicate the vector $\boldsymbol{\varphi}$ without the coordinate containing the parameter a. Therefore, from (4) we have, for $j = 1, 2, \cdots, K$, that

$$P\left(h_j\left(1\right)\middle|\boldsymbol{\varphi}-\left\{h_j\left(1\right)\right\}, Y_j\left(1\right)\right)$$

$$\propto \Psi_{h_j(1)}\left(h_j\left(1\right), Y_j\left(1\right)\right)\exp\left\{-\frac{1}{2\sigma_{\eta_j}^2}\left[h_j\left(1\right)-\mu_j\right]^2\right\},$$

where

$$\Psi_{h_j(1)}\left(h_j\left(1\right), Y_j\left(1\right)\right)$$

$$= \exp\left\{-\frac{1}{2}\left(h_j\left(1\right)+\mathrm{e}^{-h_j(1)}Y_j^2\left(1\right)\right)\right\},$$

and for $t = 2, 3, \cdots, N$,

$$P\left(h_j\left(t\right)\middle|\boldsymbol{\varphi}-\left\{h_j\left(t\right)\right\}, Y_j\left(t\right)\right)$$

$$\propto \Psi_{h_j(t)}\left(h_j\left(t\right), Y_j\left(t\right)\right)$$

$$\times\exp\left\{-\frac{1}{2\sigma_{\eta_j}^2}\left[h_j\left(t\right)-\mu_j-\phi_j\left(h_j\left(t-1\right)-\mu_j\right)\right]^2\right\}$$

where

$$\Psi_{h_j(t)}\left(h_j\left(t\right), Y_j\left(t\right)\right)$$

$$= \exp\left\{-\frac{1}{2}\left[h_j\left(t\right)+\mathrm{e}^{-h_j(t)}Y_j^2\left(t\right)\right]\right\}.$$

Therefore, in the case of $h_j\left(t\right), t = 1, 2, \cdots, N$, if y is the new proposed value for $h_j\left(t\right)$ and x is its actual value, then the acceptance probability in the Metropolis-Hastings step is given by (we use the respective prior distributions to generate the proposed values)

$$\rho\left(x,y\right) = \min\left\{1, \frac{P\left(y\middle|\boldsymbol{\varphi}-\left\{h_j\left(1\right)\right\}, Y_j\left(1\right)\right)\cdot\mathrm{e}^{-\frac{1}{2\sigma_{\eta_j}^2}\left(x-\mu_j\right)^2}}{P\left(x\middle|\boldsymbol{\varphi}-\left\{h_j\left(1\right)\right\}, Y_j\left(1\right)\right)\cdot\mathrm{e}^{-\frac{1}{2\sigma_{\eta_j}^2}\left(y-\mu_j\right)^2}}\right\}$$

$$= \min\left\{1, \frac{\Psi_{h_j(1)}\left(y, Y_j\left(1\right)\right)}{\Psi_{h_j(1)}\left(x, Y_j\left(1\right)\right)}\right\},$$

for $t = 1$ and for $t = 2, 3, \cdots, N$,

$$\rho\left(x,y\right)$$

$$= \min\left\{1, \frac{P\left(y\middle|\boldsymbol{\varphi}-\left\{h_j\left(t\right)\right\}, Y_j\left(t\right)\right)\cdot\mathrm{e}^{-\frac{1}{2\sigma_{\eta_j}^2}\left(x-\mu_j-\phi_j\left(h_j\left(t-1\right)-\mu_j\right)\right)^2}}{P\left(x\middle|\boldsymbol{\varphi}-\left\{h_j\left(t\right)\right\}, Y_j\left(t\right)\right)\cdot\mathrm{e}^{-\frac{1}{2\sigma_{\eta_j}^2}\left(y-\mu_j-\phi_j\left(h_j\left(t-1\right)-\mu_j\right)\right)^2}}\right\}$$

$$= \min\left\{1, \frac{\Psi_{h_j(t)}\left(y, Y_j\left(t\right)\right)}{\Psi_{h_j(t)}\left(x, Y_j\left(t\right)\right)}\right\}.$$

Additionally, we have, for $j = 1, 2, \cdots, K$, that $P\left(\mu_j\middle|\boldsymbol{\varphi}-\left\{\mu_j\right\}, D\right)$ is proportional to a Normal $\left(B_j/A_j, 1/A_j\right)$, where

$$A_j = \frac{1}{\sigma_{\eta_j}^2}\left[1+\left(N-1\right)\left(1-\phi_j\right)^2\right]+\frac{1}{f_j},$$

$$B_j = \frac{1}{\sigma_{\eta_j}^2}\left(h_j\left(1\right)+\left(1-\phi_j\right)\sum_{t=2}^{N}\left[h_j\left(t\right)-\phi_j h_j\left(t-1\right)\right]\right)$$

$$+\frac{e_j}{f_j},$$

that $P\left(\phi_j\middle|\boldsymbol{\varphi}-\left\{\phi_j\right\}, D\right)$ is proportional to a Normal $\left(D_j/C_j, 1/C_j\right)$, with

$$C_j = \left(\frac{1}{\sigma_{\eta_j}^2} \sum_{t=2}^{N} \left(h_j(t-1) - \mu_j \right)^2 + \frac{1}{b_j} \right),$$

$$D_j = \left(\frac{1}{\sigma_{\eta_j}^2} \sum_{t=2}^{N} \left(h_j(t) - \mu_j \right)\left(h_j(t-1) - \mu_j \right) + \frac{a_j}{b_j} \right),$$

and that $P\left(\sigma_{\eta_j}^2 \middle| \boldsymbol{\varphi} - \left\{ \sigma_{\eta_j}^2 \right\}, D \right)$ is proportional to a inverse Gamma IG $\left((N/2) + c_j, (1/2)\left(E_j + F_j + 2d_j \right) \right)$, where

$$E_j = \left(h_j(1) - \mu_j \right)^2,$$

$$F_j = \sum_{t=2}^{N} \left[h_j(t) - \mu_j - \phi_j \left(h_j(t-1) - \mu_j \right) \right]^2.$$

Convergence of the Gibbs sampling algorithm was monitored using the analysis of the trace plots and the package CODA that is part of the software R. Specifically, we used the Geweke ([37]), Raftery-Lewis ([38]), Heidelberger-Welch ([39]) and the Gelman-Rubin ([40]) tests.

4. An Application to Mexico City Ozone Measurement

In this section we apply the stochastic volatility model, described in earlier sections, to ozone measurements from the monitoring network of Mexico City. The Metropolitan Area of Mexico City is divided into five sections, namely, NE (Northeast), NW (Northwest), CE (Centre), SE (Southeast) and SW (Southwest) (see for instance [10], [15] and http://www.sma.df.gob.mx). The data used here correspond to the daily maximum ozone measurements taken from 01 January 2008 until 31 December 2010, corresponding to $T = 1092$ observations. Measurements are taken minute by minute and the 1-hour average results is reported at each monitoring station. The daily maximum measurement for a given region is the maximum over all the maximum averaged values recorded hourly during a 24-hour period by each station placed in the region. During the period of time considered, the mean concentration levels for regions NE, NW, CE, SE and SW are, respectively, 0.0829, 0.0851, 0.0876, 0.0911, and 0.0977 with respective standard deviation given by 0.0286, 0.0303, 0.0306, 0.0294 and 0.0341. We also have a total of $N = 155$ weeks and note that in here $K = 5$. We associate $k = 1, 2, 3, 4$ and 5, with regions NE, NW, CE, SE and SW, respectively.

Figure 2 presents the plots of the weekly averaged measurements for each of the regions during the period of time considered here.

Note that the weekly averaged measurements also present low values in a large number of weeks. However, it seems that the variation from one week to another may, in some cases, be very high. Therefore, it is of interest to know how this variability is behaving. Using stochastic volatility models we will be able to obtain information on how the variation behaves and with the estimated values
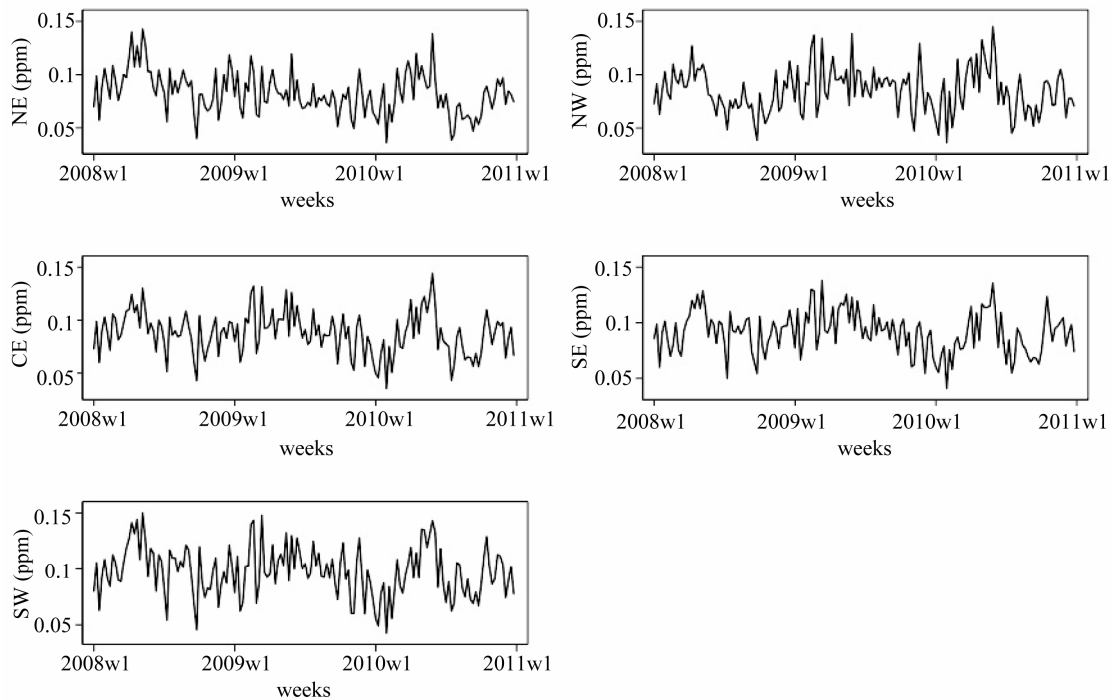

Figure 2. Weekly averaged ozone measurements for regions NE, NW, CE, SE and SW for the period ranging from 01 January 2008 to 31 December 2010.

of the parameters of the model, we may perform predictions about future behaviour of this variability. As said before, estimation of the parameters is going to be made using a sample from the respective complete conditional marginal posterior distributions which will be drawn using the Gibbs sampling algorithm run for three separate chains. Each chain run 21000 steps and after a burn-in period of 2000 steps, a sample of size 3800, taken every 5th generated value, was obtained from each one. Hence, estimation of the parameters was performed using a sample of size 11400. The hyperparameters of the prior distributions are $a_j = 0$, $b_j = 1$, $c_j = 3$, $d_j = 3$, $e_j = 0$, $f_j = 10$, $j = 1, 2, 3, 4, 5$. **Table 1**

presents the mean, the standard deviation (indicated by SD) and the 95% credible intervals for the parameters of the model when data from each of the regions are used.

In **Figures 3-5**, we have the plots of the estimated (dashed lines) and observed (solid lines) log-returns of the weekly averaged measurements for the entire observational period and including the predicted and the observed values for the month of January 2011 (not included in the data when estimating the parameters).

Observing **Figures 3-5**, we may see that, most of the time, the estimated log-returns of the weekly averaged measurements provide a good fit to the observed ones.

Table 1. Posterior mean, standard deviation (indicated by SD) and 95% credible intervals of the parameters of interest for all regions.

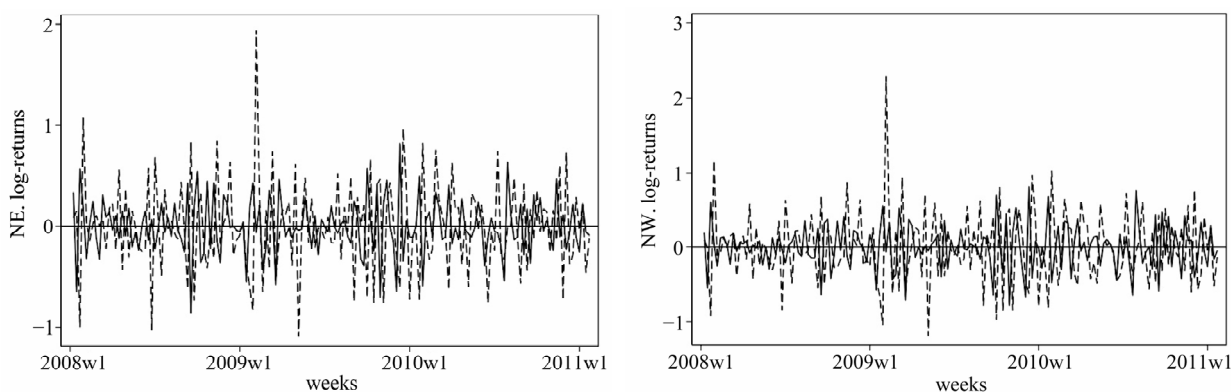| Region	Parameter	Mean	SD	95% Credible interval
NE	μ	−2.864	0.174	(−3.205, −2.524)
	ϕ	0.154	0.150	(−0.14, 0.448)
	σ_η^2	0.837	0.281	(0.287, 1.387)
NW	μ	−2.689	0.182	(−3.046, −2.332)
	ϕ	0.163	0.152	(−0.135, 0.46)
	σ_η^2	0.912	0.328	(0.269, 1.555)
CE	μ	−2.689	0.167	(−3.016, −2.362)
	ϕ	0.141	0.151	(−0.155, 0.437)
	σ_η^2	0.757	0.255	(0.257, 1.256)
SE	μ	−2.921	0.173	(−3.26, −2.582)
	ϕ	0.135	0.143	(−0.145, 0.416)
	σ_η^2	0.888	0.292	(0.315, 1.46)
SW	μ	−2.668	0.171	(−3.003, −2.333)
	ϕ	0.131	0.150	(−0.163, 0.425)
	σ_η^2	0.839	0.287	(0.277, 1.401)

Figure 3. Observed (solid line) and estimated/predicted (dashed) log-returns of the weekly averaged measurements for regions NE and NW for the period ranging from 01 January 2008 to 31 December 2010/31 January 2011.
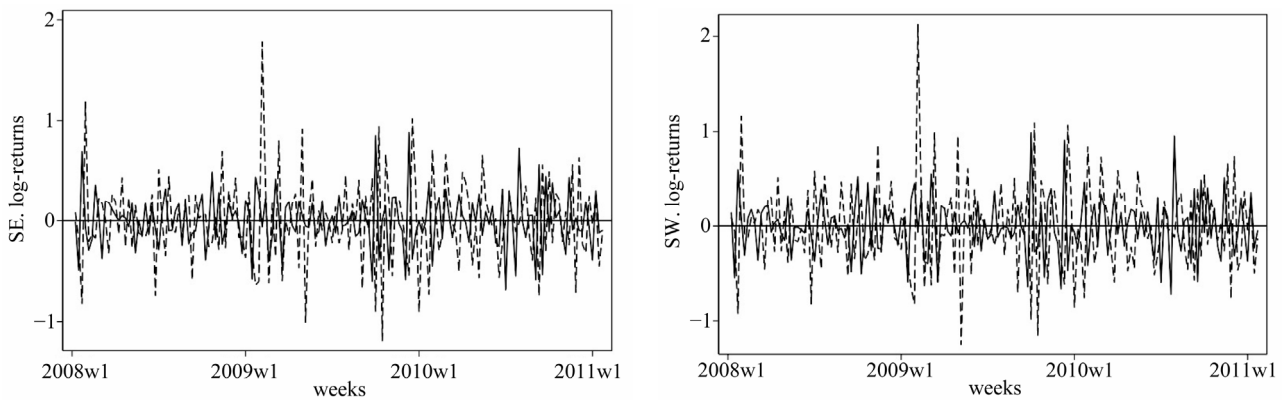
Figure 4. Observed (solid line) and estimated/predicted (dashed) log-returns of the weekly averaged measurements for regions SE and SW for the period ranging from 01 January 2008 to 31 December 2010/31 January 2011.
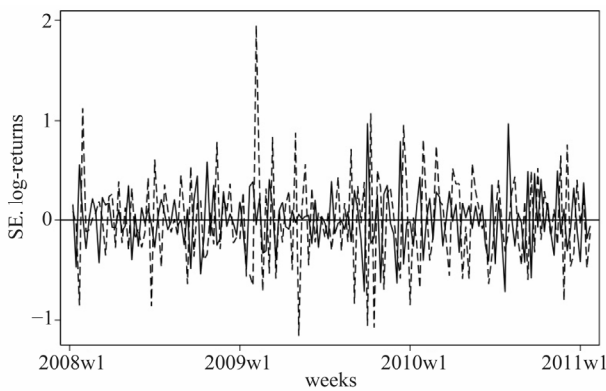

Figure 5. Observed (solid line) and estimated/predicted (dashed) log-returns of the weekly averaged measurements for region CE for the period ranging from 01 January 2008 to 31 December 2010/31 January 2011.

5. Discussion

In this paper we have considered a particular version of a multivariate stochastic volatility model to study the behaviour of the weekly ozone averaged measurements. The interest was in the variability of the series of measurements. The parameters present in the model were estimated using a Gibbs sampling algorithm. This algorithm was implemented using the package R and the code is given in the Appendix of this work. Even though the model considered here is a particular case of a more general version, the general approach may also be implemented by making some adjustments in the code.

Looking at **Table 1**, we may see that the values of the parameters are very similar in all regions. That is justified by the fact that the weekly averaged measurements behave in a very similar way in all regions during the observational period (see **Figure 2**). Hence, the log-returns series will reflect that behaviour. Also from **Table 1**, we may see that the effect of the parameter μ on the latent variables is a negative one in all regions. That effect is also observed in similar studies ([22,23]). Note

that in all regions the value of ϕ is very small. That means that the effect of past values in the present ones is very small. That same behaviour is also seen in [22,23]. However, for the present dataset the effect is even smaller.

Using the estimated parameters given in **Table 1**, we may obtain estimates for the averaged weekly measurements in future weeks. In order to illustrate that we consider the month of January 2011. Note that, by hypothesis we have, for each $j = 1, 2, \cdots, K$, that $Y_j(t) = \log\left(Z_j(t)/Z_j(t-1)\right)$, and therefore, we may write $Z_j(t) = Z_j(t-1)\exp\left[Y_j(t)\right]$, where $Y_j(t) = \exp\left[h_j(t)/2\right]\epsilon_j(t)$. Hence, using the value of the last week of December 2010 (which would be $Z_j(t-1)$), and using the simulated values for $h_j(t)$ and $\epsilon_j(t)$ we may obtain an estimate for the averaged measurement of the first week of January 2011. This values is used as $Z_j(t-1)$ in the next step. Again, we simulate values for $h_j(t)$ and $\varepsilon_j(t)$ and with that we obtain the new $Z_j(t)$, which will correspond to the second week of January 2011, and so on. Those estimated values and the observed ones are given in **Table 2**.

Looking at **Table 2** we may observe that for the first two weeks of January 2011 the estimated weekly averaged values overestimate the observed ones. However, for the last two weeks, the observed weekly averaged measurements are underestimated. The error in the estimation may be justified by the accumulated errors that might occur during the estimation of $h_j(t)$ and $\varepsilon_j(t)$. Nevertheless, the approximation is overall very good. That can also be corroborated by looking at the plots of the log-returns given in **Figures 3-5**.

6. Conclusions

Even though the version of the model considered here assume independence among regions, it is possible to see, by looking at **Figures 3-5** and at **Table 2**, that the

Table 2. Observed and predicted weekly averages ozone measurements for the month of January 2011.

Weeks		NE	NW	CE	SE	SW
1	observed	0.064	0.063	0.057	0.063	0.069
	estimated	0.076	0.07	0.079	0.083	0.084
2	observed	0.08	0.084	0.084	0.085	0.098
	estimated	0.095	0.091	0.099	0.103	0.108
3	observed	0.076	0.071	0.07	0.076	0.078
	estimated	0.06	0.053	0.061	0.066	0.065
4	observed	0.069	0.068	0.065	0.069	0.074
	estimated	0.053	0.047	0.054	0.058	0.057

estimated quantities provide a good approximation to the observed one as well as providing a good prediction to future observations.

Since it provides an instrument to be used to estimate the parameters in volatility models, the programme code given in this work may help those interested in studying volatility of times series in general. Even though the model is applied to air pollution data, the code may be used in any dataset.

We would like to call attention to the fact that a model where correlation amongst the regions is considered may provide a better insight on the behaviour of the volatility of the weekly averaged measurements. That would incorporate the possible effect that the wind direction may have on the measurements in different regions. In the present work we have not dealt with that, but as mentioned before a modification on the complete conditional marginal posterior distributions and on the code of the programme can be easily made in order to incorporate those cases. However, that is the subject of future studies.

7. Acknowledgements

REFERENCES

[1] W. J. Gauderman, E. Avol, F. Gililand, H. Vora, D. Thomas, K. Berhane, R. McConnel, N. Kuenzli, F. Lurmman, E. Rappaport, H. Margolis, D. Bates and J. Peter, "The Effects of Air Pollution on Lung Development from 10 to 18 Years of Age," *New England Journal of Medicine*, Vol. 351, 2004, pp. 1057-1067. doi:10.1056/NEJMoa040610

[2] W. Davis, L. Zaikowski and S. C. Nodvin, "Acid Rain," In: J. Cutler, Ed., *Encyclopedia of Earth*, Cleveland, Washington DC, 2011. http://www.eoearth.org/article/Acid_rain

[3] WHO (World Health Organization), "Air Quality Guidelines—2005. Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide," World Health Organization Regional Office for Europe, 2006.

[4] M. L. Bell, R. Peng and F. Dominici, "The Exposure-Response Curve for Ozone and Risk of Mortality and the Adequacy of Current Ozone Regulations," *Environmental Health Perspectives*, Vol. 114, No. 4, 2005, pp. 532-536. doi:10.1289/ehp.8816

[5] D. Loomis, V. H. Borja-Arbuto, S. I. Bangdiwala and C. M. Shy, "Ozone Exposure and Daily Mortality in Mexico City: A Time Series Analysis," *Health Effects Institute Research Report*, Vol. 75, 1996, pp. 1-46.

[6] M. R. O'Neill, D. Loomis and V. H. Borja-Aburto, "Ozone, Area Social Conditions and Mortality in Mexico City," *Environmental Research*, Vol. 94, No. 3, 2004, pp. 234-242. doi:10.1016/j.envres.2003.07.002

[7] N. Gouveia and T. Fletcher, "Time Series Analysis of Air Pollution and Mortality: Effects by Cause, Age and Socio-Economics Status," *Journal of Epidemiology and Community Health*, Vol. 54, 2000, pp. 750-755. doi:10.1136/jech.54.10.750

[8] R. L. Smith, "Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone," *Statistical Sciences*, Vol. 4, No. 4, 1989, pp. 367-393. doi:10.1214/ss/1177012400

[9] J.-N. Pan and S.-T. Chen, "Monitoring Long-Memory Air Quality Data Using ARFIMA Model," *Environmetrics*, Vol. 19, No. 2, 2008, pp. 209-219. doi:10.1002/env.882

[10] L. J. Álvarez, A. A. Fernández-Bremauntz, E. R. Rodrigues and G. Tzintzun, "Maximum a Posteriori Estimation of the Daily Ozone Peaks in Mexico City," *Journal of Agricultural*, *Biological Statistics*, Vol. 10, No. 3, 2005, pp. 276-290. doi:10.1198/108571105X59017

[11] L. C. Larsen, R. A. Bradley and G. L. Honcoop, "A New Method of Characterizing the Variability of Air Quality-Related Indicators," *Air and Waste Management Association's International Specialty Conference of Tropospheric Ozone and the Environment*, Los Angeles, EUA, 19-22 March 1990.

[12] G. Huerta and B. Sansó, "Time-Varying Models for Ex-

AM

treme Values," *Environmental and Ecological Statistics*, Vol. 14, No. 3, 2007, pp. 285-299. doi:10.1007/s10651-007-0014-3

[13] J. S. Javits, "Statistical Interdependencies in the Ozone National Ambient Air Quality Standard," *Journal of Air Pollution Control Association*, Vol. 30, No. 1, 1980, pp. 58-59. doi:10.1080/00022470.1980.10465918

[14] A. E. Raftery, "Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone," *Statistical Sciences*, Vol. 4, No. 4, 1989, pp. 378-381. doi:10.1214/ss/1177012401

[15] J. A. Achcar, A. A. Fernández-Bremauntz, E. R. Rodrigues and G. Tzintzun, "Estimating the Number of Ozone Peaks in Mexico City Using a Non-Homogeneous Poisson Model," *Environmetrics*, Vol. 19, No. 5, 2008, pp. 469-485. doi:10.1002/env.890

[16] J. A. Achcar, E. R. Rodrigues and G. Tzintzun, "Using Non-Homogeneous Poisson Models with Multiple Change-Points to Estimate the Number of Ozone Exceedances in Mexico City," *Environmetrics*, Vol. 22, No. 1, 2011, pp. 1-12. doi:10.1002/env.1029

[17] E. R. Rodrigues and J. A. Achcar, "Applications of Discrete-Time Markov Chains and Poisson Processes to Air Pollution Modeling and Studies," Springer, New York, 2012.

[18] L. W. Davis, "The Effect of Driving Restrictions on Air Quality in Mexico City," *Journal of Political Economy*, Vol. 116, No. 1, 2008, pp. 39-81. doi:10.1086/529398

[19] G. McKinley, M. Zuk, M. Höjer, M. Avalos, I. Gonzaléz, R. Iniestra, L. Laguna, M. A. Martnez, P. Osnaya, L. M. Reynales, R. Valdés and J. Martnez, "Quantification of Local and Global Benefits from Air Pollution Control in Mexico City," *Environmental Science & Technology*, Vol. 39, No. 7, 2005, pp. 1954-1961. doi:10.1021/es035183e

[20] M. Zavala, S. C. Herndon, E. C. Wood, T. B. Onasch, W. B. Knighton, L. C. Marr, C. E. Kolb and L. T. Molina, "Evaluation of Mobile Emissions Contributions to Mexico City's Emissions Inventory Using on Road and Cross-Road Emission Measurements and Ambient Data," *Atmospheric Chemistry and Physics*, Vol. 9, 2009, pp. 6305-6317. doi:10.5194/acp-9-6305-2009

[21] J. A. Achcar, H. C. Zozolotto and E. R. Rodrigues, "Bivariate Stochastic Volatility Models Applied to Mexico City Ozone Pollution Data," In: G. C. Romano and A. G. Conti, Eds., *Air Quality in the* 21*st Century*, Nova Publishers, New York, 2010, pp. 241-267.

[22] J. A. Achcar, E. R. Rodrigues and G. Tzintzun, "Using Stochastic Volatility Models to Analyse Weekly Ozone Averages in Mexico City," *Environmental and Ecological Statistics*, Vol. 18, No. 2, 2011, pp. 271-290. doi:10.1007/s10651-010-0132-1

[23] H. C. Zozolotto, "Aplicação de Modelos de Volatilidade Estocástica em Dados de Poluição do ar de Duas Grandes Cidades: Cidade do México e São Paulo," Master's Dissertation, Universidade de São Paulo, Ribeirão Preto, 2010.

[24] J. A. Achcar, H. C. Zozolotto, E. R. Rodrigues and P. H. N. Saldiva, "Two Multivariate Stochastic Volatility Models Applied to Air Pollution Data from São Paulo, Brazil," *Advances and Applications in Statistics*, Vol. 20, 2011, pp. 1-23.

[25] NOM, "Modificación a la Norma Oficial Mexicana NOM-020-SSA1-1993," Diario Oficial de la Federación, 2002.

[26] D. J. Spiegelhalter, A. Thomas, N. G. Best and W. R. Gilks, "Winbugs: Bayesian Inference Using Gibbs Sampling," MRC Biostatistics Unit, Cambridge, 1999.

[27] E. Ghysels, A. C. Harvey and E. Renault, "A Stochastic Volatility," In: C. R. Rao and G. S. Maddala, Eds., *Statistical Models in Finance*, North-Holland, Amsterdam, 1996. doi:10.1016/S0169-7161(96)14007-4

[28] S. Kim, N. Shepard and S. Chib, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, Vol. 65, No. 3, 1998, pp. 361-393. doi:10.1111/1467-937X.00050

[29] R. Meyer and J. Yu, "BUGS for a Bayesian Analysis of Stochastic Volatility Models," *Econometrics Journal*, Vol. 3, No. 2, 2000, pp. 198-215. doi:10.1111/1368-423X.00046

[30] J. Yu and R. Meyer, "Multivariate Stochastic Volatility models: Bayesian Estimation and Model Comparison," *Econometric Reviews*, Vol. 25, No. 2-3, 2006, pp. 361-384. doi:10.1080/07474930600713465

[31] R. F. Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, Vol. 50, No. 4, 1982, pp. 987-1007. doi:10.2307/1912773

[32] T. Bollerslev, "Generalized Autoregressive Conditional Heterocedasticity," *Journal of Econometrics*, Vol. 31, No. 3, 1986, pp. 307-327. doi:10.1016/0304-4076(86)90063-1

[33] A. F. M. Smith and G. O. Roberts, "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society Series B*, Vol. 55, 1993, pp. 3-23.

[34] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, "Equation of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics*, Vol. 21, No. 6, 1953, pp. 1087-1092. doi:10.1063/1.1699114

[35] W. K. Hastings, "Monte Carlo Sampling-Based Methods Using Markov Chains and Their Applications," *Biometrika*, Vol. 57, No. 1, 1970, pp. 97-109. doi:10.1093/biomet/57.1.97

[36] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, Vol. 49, No. 4, 1995, pp. 327-335. doi:10.1080/00031305.1995.10476177

[37] J. Geweke, "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," In: J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds., *Bayesian Statistics*, Vol. 4, Clarendon Press, Oxford, 1992, pp. 169-193.

[38] A. Raftery and S. Lewis, "How Many Iterations in the Gibbs Sampler?" In: J. Bemardo, J. Berger, A. Dawid and A. Smith, Eds., *Bayesian Statistics*, Vol. 4, Claredon Press, Oxford, 1992, pp. 763-774.

[39] P. Heidelberger and P. Welch, "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations," *Communications of the Association for Computing Machinery*, Vol. 24, No. 4, 1983, pp. 233-245. doi:10.1145/358598.358630

[40] A. Gelman and D. B. Rubin, "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Sciences*, Vol. 7, No. 4, 1992, pp. 457-511. doi:10.1214/ss/1177011136

Appendix

```
# GIBBS SAMPLING ALGORITHM
# Auxiliary functions
psi_hj1=function(hj1,yj1){
b_2=exp(-1/2*(hj1+exp(-hj1)*yj1^2))
b_2
}
######################################
psi_hjt=function(hjt,yjt){
b_1=exp(-1/2*(hjt+exp(-hjt)*yjt^2))
b_1
}
######################################
# Initial data
n=21000
burn=2000
n.cadenas=3
salto=5
region=1
seed=987
if(seed>0){set.seed(seed)}
a1=0 ;b1=1
c1=3 ;d1=3
e1=0 ;f1=10
# Generating the initial values
in_mu=rnorm(n.cadenas,0,1)
in_phi=rnorm(n.cadenas,0,0.35)
in_sigma2=1/rgamma(n.cadenas,3,2)
######################################
mide.tiempo=proc.time()
if(region==1){y1=scan("c:/r/rsne3.txt")}
if(region==2){y1=scan("c:/r/rsnw3.txt")}
if(region==3){y1=scan("c:/r/rsce3.txt")}
if(region==4){y1=scan("c:/r/rsse3.txt")}
if(region==5){y1=scan("c:/r/rssw3.txt")}
N=155 # number of weeks (three years)
espmu=e1;espphi=a1;espsigma2=d1/(c1-1)
######################################
# Storing the values of the chains
SIMmu=matrix(0,n,n.cadenas)
SIMphi=matrix(0,n,n.cadenas)
SIMsigma2=matrix(0,n,n.cadenas)
Hjtn=matrix(0,n,N*n.cadenas) # store simulated value of
# h
Contador=matrix(0,N,n.cadenas) # store accepted Metro-
# polis-Hasting values
######################################
for(m in 1:n.cadenas){
######################################
# Gibbs simulation of the latent variables (n=1)
# h_j, t=1,2,...,155; N=155
# First step for t=1
# Metropolis-Hastings step in the Gibbs sampling
anterior=in_mu[m] # initial value for h_j(1)
```

```
u=runif(1)
y=rnorm(1,in_mu[m],sqrt(in_sigma2[m])) # proposal
# density
num=psi_hj1(y,y1[1])
den=psi_hj1(anterior,y1[1])
rho=num/den
Hjtn[1,(N*(m-1)+1)]=anterior+(y-anterior)*(u<=rho) #
# first simulated Gibbs value
Contador[1,m]=(u<=rho)
# simulation for t=2,3,...,N
for(t in 2:N){
hj_ant=Hjtn[1,(N*(m-1)+t-1)] # Gibbs actualization
anterior=rnorm(1,hj_ant,sqrt(in_sigma2[m])) # initial
# value
u=runif(1)
y=rnorm(1,in_mu[m]+in_phi[m]*(hj_ant-in_mu[m]),sqrt
(in_sigma2[m])) # proposed value
num=psi_hjt(y,y1[t])
den=psi_hjt(anterior,y1[t])
rho=num/den
Hjtn[1,(N*(m-1)+t)]=anterior+(y-anterior)*(u<=rho)
Contador[t,m]=(u<=rho)
}
######################################
# mu_1--->n=1
# Gibbs actualisation
# h_1=Hjtn[1,(N*(m-1)+1):(N*(m-1)+N)]
# Gibbs actualisation#
# Parameters of the complete conditional marginal densi-
# ties
##########
A=1/f1+(1+(1-in_phi[m])^2*(N-1))/in_sigma2[m]
diferencias=numeric(N)
for(t in 2:N){
diferencias[t]=h_1[t]-in_phi[m]*h_1[t-1]}
B=e1/f1+(h_1[1]+(1-in_phi[m])*sum(diferencias))/in_si
gma2[m]
C=B/A; D=1/A
##################################################
##
SIMmu[1,m]=rnorm(1,C,sqrt(D))
######################################
# phi_11--->n=1
# Gibbs actualisation#
mu_1=SIMmu[1,m]
#Gibbs actualisation#
#Parameters of the complete conditional marginal densi-
# ties
##########
diferencias1=numeric(N)
for(t in 2:N){
diferencias1[t]=(h_1[t-1]-mu_1)^2}
A=1/b1+sum(diferencias1)/in_sigma2[m]
diferencias2=numeric(N)
```

```
for(t in 2:N){
diferencias2[t]=(h_1[t]-mu_1)*(h_1[t-1]-mu_1)}
B=a1/b1+sum(diferencias2)/in_sigma2[m]
C=B/A; D=1/A
###############################################
##
SIMphi[1,m]=rnorm(1,C,sqrt(D))
##################################
# sigma2_1--->n=1
# Gibbs actualisation#
phi_1=SIMphi[1,m]
# Gibbs actualisation#
# Parameters of the complete conditional marginal densi-
# ties
##########
A=c1+N/2
diferencias=numeric(N)
for(t in 2:N){
diferencias[t]=(h_1[t]-mu_1-phi_1*(h_1[t-1]-mu_1))^2}
B=1/2*((h_1[1]-mu_1)^2+sum(diferencias))+d1
###############################################
##
SIMsigma2[1,m]=1/rgamma(1,A,B)
###########################################
# Gibbs simulation for n=2,3,...,
###########################################
for(k in 2:n){
# Gibbs actualisation#
mu_1=SIMmu[k-1,m]
phi_1=SIMphi[k-1,m]
sigma2_1=SIMsigma2[k-1,m]
# Gibbs actualisation#
# h_j, t=1
anterior=Hjtn[k-1,(N*(m-1)+1)]
u=runif(1)
y=rnorm(1,mu_1,sqrt(sigma2_1))
num=psi_hj1(y,y1[1])
den=psi_hj1(anterior,y1[1])
rho=num/den
Hjtn[k,(N*(m-1)+1)]=anterior+(y-anterior)*(u<=rho)
Contador[1,m]=Contador[1,m]+(u<=rho)
# simulating h_j, t=2,3,...,N
for(t in 2:N){
hj_ant=Hjtn[k,(N*(m-1)+t-1)]
anterior=Hjtn[k-1,(N*(m-1)+t)]
u=runif(1)
y=rnorm(1,mu_1+phi_1*(hj_ant-mu_1),sqrt(sigma2_1))
num=psi_hjt(y,y1[t])
den=psi_hjt(anterior,y1[t])
rho=num/den
Hjtn[k,(N*(m-1)+t)]=anterior+(y-anterior)*(u<=rho)
Contador[t,m]=Contador[t,m]+(u<=rho)
}
##################################
```

```
# mu_1-->>n=k
# Gibbs actualisation#
h_1=Hjtn[k,(N*(m-1)+1):(N*(m-1)+N)]
# Gibbs actualisation#
# Parameters of the complete conditional marginal densi-
# ties
##########
A=1/f1+(1+(1-phi_1)^2*(N-1))/sigma2_1
diferencias=numeric(N)
for(t in 2:N){
diferencias[t]=h_1[t]-phi_1*h_1[t-1]}
B=e1/f1+(h_1[1]+(1-phi_1)*sum(diferencias))/sigma2_1
C=B/A; D=1/A
###############################################
##
SIMmu[k,m]=rnorm(1,C,sqrt(D))
##################################
# phi_11-->>n=k
# Gibbs actualisation#
mu_1=SIMmu[k,m]
# Gibbs actualisation#
#Parameters of the complete conditional marginal densi-
ties
##########
diferencias1=numeric(N)
for(t in 2:N){
diferencias1[t]=(h_1[t-1]-mu_1)^2}
A=1/b1+sum(diferencias1)/sigma2_1
diferencias2=numeric(N)
for(t in 2:N){
diferencias2[t]=(h_1[t]-mu_1)*(h_1[t-1]-mu_1)}
B=a1/b1+sum(diferencias2)/sigma2_1
C=B/A; D=1/A
###############################################
##
SIMphi[k,m]=rnorm(1,C,sqrt(D))
##################################
# sigma2_1-->>n=k
# Gibbs actualisation#
phi_1=SIMphi[k,m]
# Gibbs actualisation#
# Parameters of the complete conditional marginal densi-
# ties
##########
A=c1+N/2
diferencias=numeric(N)
for(t in 2:N){
diferencias[t]=(h_1[t]-mu_1-phi_1*(h_1[t-1]-mu_1))^2}
B=1/2*((h_1[1]-mu_1)^2+sum(diferencias))+d1
###############################################
##
SIMsigma2[k,m]=1/rgamma(1,A,B)
}}
```