Transfer Functions in Dynamic Generalized Linear Models

Mariane B. Alves, Dani Gamerman and Marco Antonio R. Ferreira

September, 2006

Resumo

As one specifies the effect of a regressor in a time series analysis it is sometimes necessary to assume that fluctuations in that variable do not have only immediate impact on the mean response, but that its effects somehow propagate to future times. We adopt, in the present work, transfer functions to model such impacts, represented by structural blocks E_t present in dynamic generalized linear models' predictors.

All the inference is carried under the Bayesian paradigm and in the context above two sources of difficulties emerge for the analytical derivation of posterior distributions: non-Gaussian nature of the response, associated to non-conjugate priors and, if there are autoregressive parameters in the block E_t, also non-linearity of the predictor on these parameters.

The purpose of this work is the to produce complete Bayesian inference on generalized dynamic linear models with transfer functions, using Monte Carlo Markov Chain methods to build samples of the posterior joint distribution of the parameters involved in such models.

Several transfer structures are specified, associated to Poisson, Binomial, Gamma and Inverse Gaussian responses. Simulated data are analyzed under the resulting models in order to access their performance. Finally, the Gamma models are applied to real data concerning the cumulative effect of daily rain volumes over pollutant levels.

Keywords: Transfer functions, Dynamic Generalized Linear Models, Bayesian inference, Monte Carlo Markov chain.

1 Introduction

There are several situations in which one aims to model the cumulative impact of a regressor on a response variable through time. For instance, one could be interested in the effect of a marketing campaign on the sales of some product or in the way rain volumes influence the flow of a river or yet in the effect of a vaccine on the counts of some disease. In all these cases, knowledge of the mechanisms of propagation of such impacts should be useful not only for predictive purposes but also to improve control over the output process, planning optimal times to intervene on the inputs.

Clearly the simplest way to specify, at each time t, the overall impact of a regressor X over a response Y, both observed temporally, is through distributed lag models, in which such effect is represented by:

$$E_t = \sum_{j=0}^{s} \beta_j X_{t-j}. \tag{1}$$

The form above presents two major difficulties. One is that it is presumed in (1) that s - the regressor's horizon of influence on Y - is known, and that the regressor's effect is null for $j > s$, what may result in sub estimation of the overall effect of X. Actually, in practical applications, the pure specification of s demands careful analysis. Another problem emerges when the effect of X on the output exhibits significant persistence, therefore implying moderate to large values of s, in which case the estimation of the parameters $\beta_0, \cdots, \beta_s$ is compromised by the autocorrelation in $X_t, \cdots, X_{t-s}$.

Thus it is usual to impose restrictions on the coefficients of a distributed lag model (Klein, 1958; Solow, 1960; Almon, 1965). A fundamental work on restricted distributed lags is the one by Koyck (1954), in which it is assumed that in an infinite distributed lag model with coefficients $\beta_0, \beta_1, \cdots, \beta_b, \cdots$, as from lag b ($b \geq 0$) the coefficients in (1) exhibit geometric decay, governed by:

$$\beta_j = \rho^j \beta, \quad (j \geq b, \quad 0 \leq \rho \leq 1). \tag{2}$$

The estimation of (2) resumes to ρ and β, thus resulting in a much more parsimonious formulation than (1). Besides, the structure (2) allows the estimation (instead of arbitration) of the period of significant influence of X over Y.

Other forms of restriction will be examined. Essentially, Koyck's formulation is a particular case of the transfer function models adopted in the present work. We assume, as described by Box et al. (1994), that the relation between the input X and the output Y can be represented by an effect E_t structured as:

$$\begin{aligned} (1 - \rho_1 B - \cdots - \rho_r B^r) E_t &= (\omega_0 - \omega_1 B - \cdots - \omega_s B^s) X_{t-b} \\ &= (\omega_0 B^b - \omega_1 B^{b+1} - \cdots - \omega_s B^{b+s}) X_t \end{aligned}$$

or

$$\rho(B) E_t = \omega(B) B^b X_t = \beta(B) X_t, \tag{3}$$

where B is the lag operator: $BE_t = E_{t-1}$.

Alternatively, the effect of X can be written as a linear filter:

$$\begin{aligned} E_t &= \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \cdots \\ &= \frac{\beta(B)}{\rho(B)} X_t \\ &= \nu(B) X_t. \end{aligned} \tag{4}$$

The polynomial $\nu(B) = \nu_0 + \nu_1 B + \nu_2 B^2 + \cdots$ is called *transfer function* and represents the cumulative effect of the regressor X on the output process. The weights $\nu_0, \nu_1, \nu_2, \cdots$ are called *impulse response function* and express the instantaneous impact of X on the output process at present and further times. In the following sections we refer to $\nu(B)$ in (4) as TF(r, s, b), indicating that $\nu(B)$ is the resulting transfer function from the ratio between two polynomials on B: one of order r, applied on the output, and one of order s applied on the regressor lagged by b instants.

The filter (4) is stable if, for $|B| \leq 1$, $\nu(B)$ is convergent, implying that finite fluctuations in the input X result in finite fluctuations in the output process Y. The stability conditions for transfer functions are equivalent to those of stationarity for ARMA models.

Substituting (4) in (3) results in the following identity:

$$(1 - \rho_1 B - \cdots - \rho_r B^r)(\nu_0 + \nu_1 B + \nu_2 B^2 + \cdots) = (\omega_0 - \omega_1 B - \cdots - \omega_s B^s)B^b, \tag{5}$$

which, equated on B gives:

$$\nu_j = \begin{cases} 0 & j < b \\ \rho_1 \nu_{j-1} + \rho_2 \nu_{j-2} + \cdots + \rho_r \nu_{j-r} + \omega_0 & j = b \\ \rho_1 \nu_{j-1} + \rho_2 \nu_{j-2} + \cdots + \rho_r \nu_{j-r} - \omega_{j-b} & j = b+1, b+2, \cdots, b+s \\ \rho_1 \nu_{j-1} + \rho_2 \nu_{j-2} + \cdots + \rho_r \nu_{j-r} & j > b+s. \end{cases}$$

Hence, several shapes can be obtained even with low orders r and s, as may be seen in figure 1.

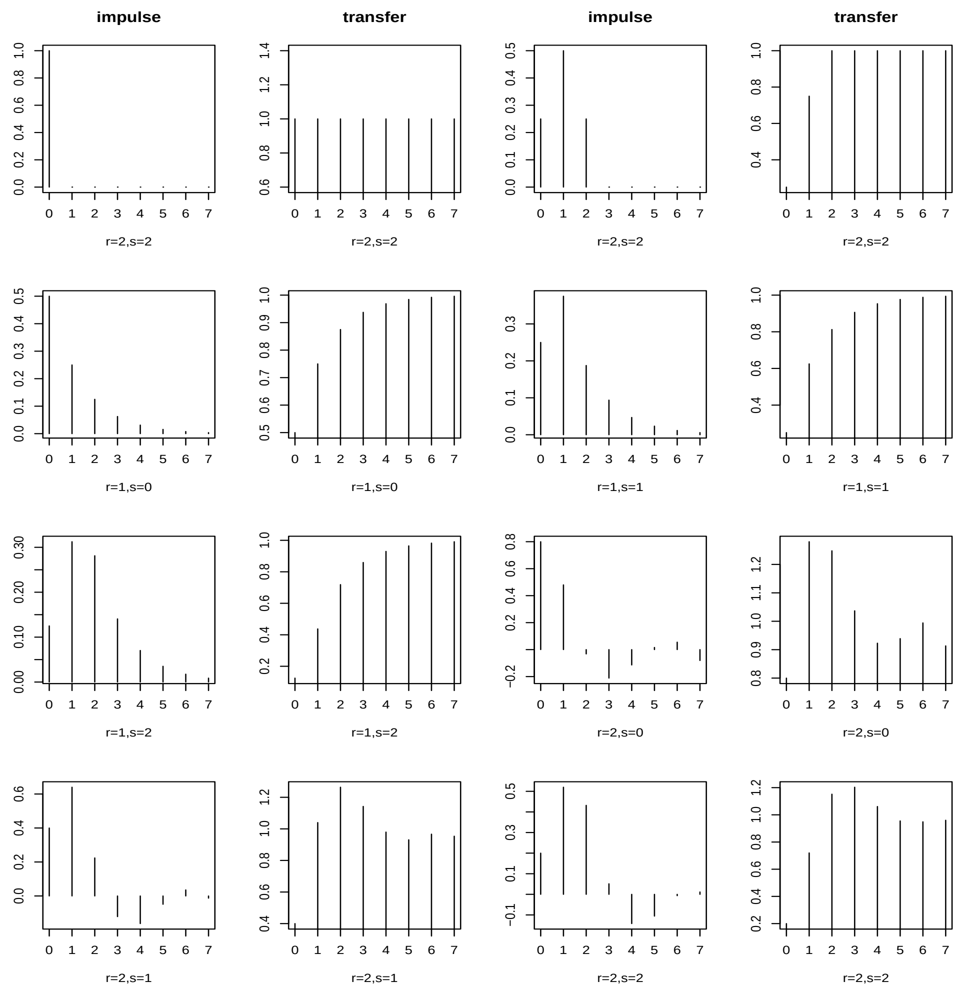

Figura 1: Examples of transfer and impulse response functions with $0 \leq r, s \leq 2$ and $b = 0$.

Clearly, the shape of an impulse response function (and consequently the shape of a transfer function) depends on the behavior of the solutions of the difference equation defining it. So, according to Box et al. (1994), considering a transfer function of order (r,s,b), that is, autoregression of order r on E_t and s lags on $B^b X_t$, we have:

- b null values ν_0, ν_1, ν_{b-1}
- $s - r + 1$ values $\nu_b, \nu_{b+1}, \nu_{b+s-r}$ without fixed pattern (if $s < r$ there are no such values)
- values ν_j, com $j \geq b + s - r + 1$ following the pattern dictated by the difference equation of order r, with initial values given by $\nu_{b+s}, \cdots, \nu_{b+s-r+1}$.

Essentially, the pattern followed by an impulse response function is determined by the roots of the autoregressive polynomial $\rho(B)$: distinct real roots provide geometric decay terms, distinct complex roots determine senoidal terms and equal roots provide polynomial terms. Thus the impulse response function is a combination of geometric, polynomial and senoidal terms.

Transfer functions are well established and applied in different contexts, particularly in Econometrics and Engineering. Nevertheless, there are few records of estimation of such models under the Bayesian paradigm. Pole (1988) and Ravines et al. (2006) present Bayesian estimation of transfer function models respectively applying Gaussian quadrature and Markov chain Monte Carlo methods to obtain posterior information. However, in both cases - as it is usual in the traditional formulation for this class of models - it is assumed that the output Y is Gaussian, which in some cases may demand transformations with the cost of lack of interpretability. Another usual restriction is the assumption that the process governing the relation between X and Y is static, what leads us to the class of Bayesian dynamic models, in which parameters are allowed to evolute according to some stochastic law.

West et al. (1985) defines the class of dynamic generalized linear models, in which responses are assumed to belong to the exponential family of distributions, resulting in an extension to dynamic linear models, defined by Harrison and Stevens (1976), in which, despite the dynamic pattern addressed to structural parameters, responses are restricted to follow the Gaussian distribution. Dynamic generalized linear models are also an extension to generalized linear models, presenting the same observational form, but imposing stochastic fluctuation to structural parameters, thus implying more flexible predictors. A good description of dynamic models and recent developments in this area can be found in Migon et al. (2005).

We follow West et al. (1985), inserting a transfer function in dynamic generalized linear models as an structural component E_t, which represents the overall effect accumulated by X over the mean response at time t. In this context two sources of difficulties emerge for the analytical derivation of posterior distributions: non-Gaussian nature of the response, associated to non-conjugate priors and, if there are autoregressive parameters also non-linearity of the predictor on these parameters.

West et al. (1985) approach the posterior via linear Bayes, in terms of first and second moments, an usual practice during the 80's to deal with non-normality / non-linearity in dynamic models. Also, some works have avoided integration by the determination of posterior modes (Singh and Roberts, 1992; Fahrmeir, 1992, 1997). Since the beginning of the 90's several sampling based approaches have been developed in order to derive posterior inference. In such approaches samples of the densities involved in the posterior analysis are propagated and updated, instead of the densities themselves.

There are several examples of approximation of the posterior distribution in non-gaussian/ non-linear dynamic models using sampling importance (Gordon et al., 1993; Durbin and Koopman, 2000). Another possibility, applied in the present work, is to use Markov chain Monte Carlo (MCMC) methods, which have been applied no non-Gaussian/ non-linear dynamic models by Carlin et al. (1992), Carter and Kohn (1994), De Jong and Shephard (1995), Shephard and Pitt (1997), Knorr-Held (1999), Geweke and Tanizaki (2001), Durbin and Koopman (2002) and Ferreira and Gamerman (2000), among others.

Our purpose is to produce complete Bayesian inference on generalized dynamic linear models with transfer functions in their predictors, using Markov Chain Monte Carlo methods to build samples of the posterior joint distribution of the parameters involved in such models.

In the next section we present the proposed models, specified for Poisson, Binomial, Gamma and Inverse Gaussian distributions and describe the simulation procedure to obtain posterior samples. Section 4 exhibits the results of the estimation procedure to some simulated data, created in order to evaluate the performance of the method. In section 5 we apply Gamma models to real data concerning the effect of rain volumes on particulate material.

2 Model

In the general form of the proposed models we suppose that $Y_t \sim \mathcal{F}(\chi_t)$, $\mathcal{F}$ a distribution in the exponential family with natural parameter χ_t, so that:

$$
\begin{aligned}
p(y_t|\chi_t) &\propto \exp\left\{\frac{y_t\chi_t - b(\chi_t)}{\phi_t}\right\}, \quad t = 1, ...T, \\
g(\mu_t) &= \eta_t = \mathbf{F}_t^{'}\boldsymbol{\alpha}_t + \mathbf{Z}'_t\boldsymbol{\delta} + E_t \\
\boldsymbol{\alpha}_t &= \mathbf{G}_t\boldsymbol{\alpha}_{t-1} + \mathbf{u}_t, \qquad u_t \sim MN(\mathbf{0}, \mathbf{W_t}), \\
E_t &= \nu(B)X_t
\end{aligned} \tag{6}
$$

in which:

- g is a monotone differentiable link function;
- $\mu_t = E[Y_t|\chi_t] = \dot{b}(\chi_t) = \frac{db(\chi_t)}{d\chi_t}$.;
- η_t is a predictor whose basic structure is:

  $\boldsymbol{\alpha}_t$, vector of parameters governing the evolution of the regressors contained in $\mathbf{F}_t$, which may include trend as well as other covariates;

  $\mathbf{G}_t$ is an evolution matrix, governing the stochastic evolution of $\boldsymbol{\alpha}_t$, which may contain unknown parameters;

  $\boldsymbol{\delta}$, is a vector of parameter associated to the covariates $\mathbf{Z}_t$, whose impact on the mean response is supposed constant ;

  E_t, is the structural block denoting present and past effects of a regressor on the mean response function μ_t, following (4) and depending on a parameter vector $\boldsymbol{\psi}$.

Specifically, the proposed models are obtained by the combination of different observational distributions $\mathcal{F}$ and forms of transfer functions defining the effects E_t, as described in (4). The model is completed with a prior distribution $\boldsymbol{\alpha}_1 \sim NM(\mathbf{a}_1, \mathbf{R}_1)$. If all the components of the evolution matrix $\mathbf{G}_t$ are known and $\mathbf{W}_t = \mathbf{W}\ \forall t$, the parametric vector to be estimated is $\boldsymbol{\theta} = (\phi_t, \boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_T, \mathbf{R}_1, \mathbf{W}, \boldsymbol{\delta}', \boldsymbol{\psi}')$.

2.1 Observational Structure

The likelihood function based on model (6) is given by:

$$
\prod_{t=1}^{T} \exp\left\{\frac{y_t\chi_t - b(\chi_t)}{\phi_t}\right\}. \tag{7}
$$

We consider the following distributions in the exponential family:

- $Y_t \sim$ Poisson $(\lambda_t) \rightarrow \chi_t = \log(\lambda_t)$, $\quad b(\chi_t) = e^{\chi_t}$ and $\quad \phi_t = 1$, with canonical link:

$$
\log(\lambda_t) = \eta_t; \tag{8}
$$

- $Y_t \sim$ Binomial$(n, p_t) \rightarrow \chi_t = \log\left(\frac{p_t}{1-p_t}\right)$, $\quad b(\chi_t) = \log[1 + e^{\chi_t}]$ and $\quad \phi_t = 1/n$, with canonical link:

$$
\log\left(\frac{p_t}{1-p_t}\right) = \eta_t; \tag{9}
$$

- $Y_t \sim$ Gamma$(\varphi, \lambda_t) \rightarrow \chi_t = -\frac{\lambda_t}{\varphi}$, $\quad b(\chi_t) = -\log(-\chi_t)$ e $\quad \phi_t = 1/\varphi$. In order to guarantee positive mean response μ_t, we adopt the logarithm link function, instead of the canonical link μ_t^{-1}:

$$\log\left(\frac{\varphi}{\lambda_t}\right) = \eta_t; \tag{10}$$

- $Y_t \sim$ Inverse Gaussian$(\mu_t, \sigma^2) \rightarrow \chi_t = -\frac{1}{2\mu_t^2}$, $\quad b(\chi_t) = -\sqrt{-2\chi_t}$, $\quad \phi_t = \sigma^2$ and Once again, in order to guarantee positive mean response, μ_t, we adopt the logarithm link, instead of the canonical $-\frac{1}{2\mu_t^2}$:

$$\log(\mu_t) = \eta_t. \tag{11}$$

2.2 Transfer Function Specification

In this section, we detail the analytical forms adopted to model the cumulative effect of the regressor X_t, E_t. Throughout this section we assume $b = 0$. When $b > 0$ the effects obtained are just delayed by b periods. In subsection 2.2.1 we use the structure (4) with $r = 0$ and $s > 0$, assuming that s is relatively high. We contour the autocorrelation problems involved in such specification imposing restrictions on the regressor's coefficients. generally with $r = 1, 0 \leq b \leq 2$. In subsections 2.2.2 and 2.2.3 we address the cases TF$(1, 0, 0)$ and TF$(1, 0 \leq s \leq 2$,0). In subsection 2.2.5 we assume an extension to the structure TF$(1, 0, 0)$, assigning a dynamic pattern to the β coefficients in (4). Finally, in subsection 2.2.4 we add unstructured error terms to the effect E_t, in order to account for factors not present in the predictor.

2.2.1 Polynomial Lag Models

A distributed lag model (1) is a particular case of (4) with $\rho(B) = 1$. Difficulties associated to such models have been discussed in section 1. One way to restrict the coefficients $\beta_0, \beta_1, \cdots \beta_s$ in a distributed lag model (1), overdrawing such difficulties, is to assume that these coefficients' trajectory may be described by a low order d polynomial on the lags (Almon, 1965), as exhibited in figure 2:

$$\beta_j = \sum_{k=0}^{d} \zeta_k j^k, \quad \text{j=0,...,s,} \tag{12}$$

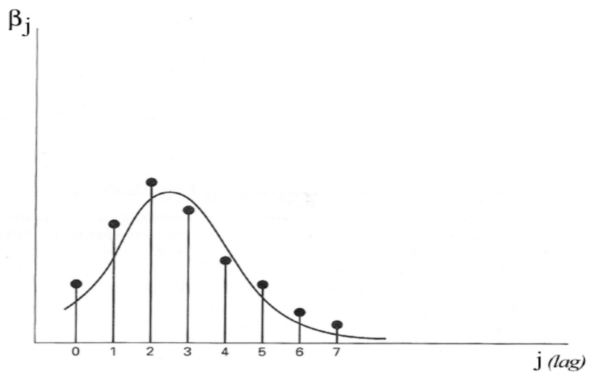

Figura 2: Polynomial approximation to the coefficients β_j in a distributed lag model.

Defining $S_{t_j} = \sum_{i=0}^{s} i^j X_{t-i}$, $\quad j = 0, 1, 2, \cdots, d$ and applying the restrictions (12) to (1), we have:

$$E_t = \zeta_0 S_{t_0} + \zeta_1 S_{t_1} + \zeta_2 S_{t_2} + ... + \zeta_d S_{t_d}. \tag{13}$$

Therefore the parameters $\boldsymbol{\psi} = (\zeta_0, \zeta_1, \cdots, \zeta_d)$ are added to the parameter vector $\boldsymbol{\theta}$. We assume a prior distribution $NM(\mathbf{m}_\zeta, \mathbf{C}_\zeta)$ for $\boldsymbol{\psi}$. Throughout this paper we suppose prior independence among parameters, so that the covariance matrixes in normal multivariate prior are diagonal.

In polynomial distributed models one must choose the degree of the smoothing polynomial, d and also the influence horizon s of the regressor X_t, which can prove to be a non trivial task, leading us to the following specifications, based on autoregressive parameters.

2.2.2 TF$(1, 0, 0)$

Suppose that, in (3), we have $r = 1$, $b = 0$ and $s = 0$, so that:

$$E_t = \rho E_{t-1} + \beta X_t. \tag{14}$$

Recursively solving the difference equation above we have:

$$E_t = \beta X_t + \rho\beta X_{t-1} + \rho^2 \beta X_{t-2} + \cdots. \tag{15}$$

Hence we have the following impulse response function:

$$\nu_j = (\rho B)^j \beta, \quad j = 0, 1, 2, \cdots. \tag{16}$$

If we suppose, without loss of generality, that $\beta > 0$, the impulse response function may exhibit the following patterns:

- $0 < \rho < 1$: geometric decay;
- $-1 < \rho < 0$: geometric decay with alternating signs;
- $\rho > 1$: geometric growth;
- $\rho < -1$: geometric growth with alternating signs.

Such patterns are exhibited in figure 3, with $\beta = 1$ and different values for ρ.

The model is determined by the parameters $\boldsymbol{\psi} = (\rho, \beta, E_0)$, for which we assume the following prior distributions: $\rho \sim U(0, 1)$, $\beta \sim N(m_\beta, C_\beta)$ e $E_0 \sim N(m_E, C_E)$.

The model is stable if $|\rho| < 1$. In this case it is called geometric lag model and was first associated to Koyck (1954). This model postulates the gradual decay of the effect of the regressor X, until such effect eventually diminishes. The autoregressive parameter ρ controls the velocity of the decay. Absolute values of ρ next to 1 imply more persistent effects of X.

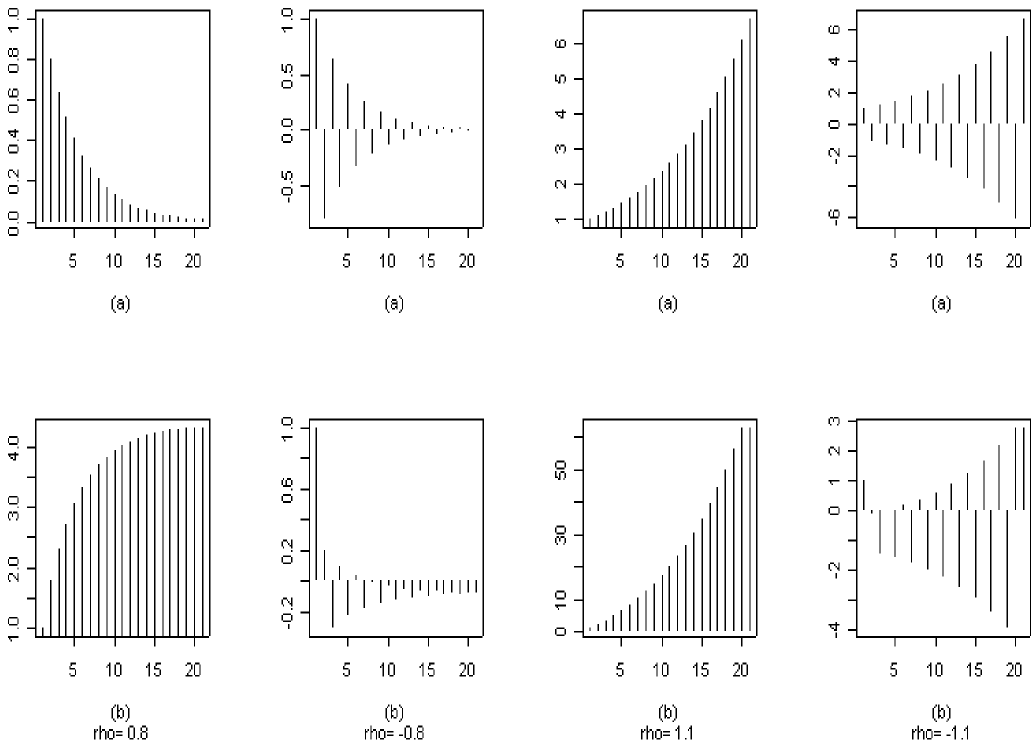

Figura 3: *TF(1,0,0): Behavior of the impulse response function (a) and transfer function (b), for* $\beta > 0$, $|\rho| < 1$ *and* $|\rho| > 1$

.

2.2.3 TF$(1, s > 0, 0)$

Instead of a monotone pattern in the impulse response function as postulated by Koyck's model, if we adopt lags on X it is possible to obtain effects which grow up until a vertex and then present geometric decay. Examples of such behaviors can be seen in figure 1, for $r = 1$ and $1 \leq s \leq 2$. The effect E_t is then expressed by:

$$E_t = \rho E_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \cdots \beta_s X_{t-s}. \tag{17}$$

The vector os parameters to be estimated in this formulation is $\boldsymbol{\psi} = (\rho, \beta_0, \cdots, \beta_s, E_0)$, with prior distributions: $\rho \sim U(0,1)$, $\boldsymbol{\beta} \sim NM(\mathbf{m}_\beta, \mathbf{C}_\beta)$ and $E_0 \sim N(m_E, C_E)$. The intrinsic autocorrelation in $X_t, X_{t-1,\cdots}$ may cause some problem to estimation procedures. Nevertheless, usually good fits are obtained with a low number of lags on X.

2.2.4 TF$(1, s \geq 0$,0) with independent random errors

In models which aim to adjust and predict a stochastic process it is usual to have influential variables omitted, maybe due to difficulties of measurement or even because in most situations it is unrealistic to presume that one can identify all the factors which influence the output.

In order to address the possible effect of omitted variables we add unstructured error terms to the effect E_t:

$$E_t = \rho E_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \cdots \beta_s X_{t-s} + \varepsilon_t, \quad \varepsilon_t \sim N(0, Q_\varepsilon). \tag{18}$$

In this formulation, although we impose more flexibility to the predictor η_t, the impulse response and transfer functions are maintained constant through time, since ρ e β coefficients are constant. The parameter vector $\boldsymbol{\theta}$ is completed with $\boldsymbol{\psi} = (\rho, \beta_0, \cdots, \beta_s, E_0,$
$\varepsilon_1, \cdots, \varepsilon_T, Q_\varepsilon)$, with prior specification given by: $\rho \sim U(0,1)$, $\boldsymbol{\beta} \sim NM(\mathbf{m}_\beta, \mathbf{C}_\beta)$, $E_0 \sim N(m_E, C_E)$, $Q_\varepsilon \sim GI(\frac{n_\varepsilon}{2}, \frac{n_\varepsilon s_\varepsilon}{2})$.

The relative magnitudes of the errors ε_t as compared to the magnitude of the whole predictor η_t may serve as an indicative of the power of the regressors present in η_t to predict the output.

2.2.5 TF$(1,0,0)$ with Dynamic Gain Factors

It is assumed by Koyck's model, described in subsection 2.2.2, that an unitary fluctuation on X at any time t, results in a constant immediate impact β. This formulation can be extended if we suppose that such impact varies through time, so that:

$$E_t = \rho E_{t-1} + \beta_t X_t. \tag{19}$$

with the dynamic in β_t dictated by some stochastic law, for instance, a random walk :

$$\beta_t = \beta_{t-1} + \upsilon_t, \quad \upsilon_t \sim N(0, Q). \tag{20}$$

The resulting impulse response function presents dynamic magnitude and shape. One example of such behavior is exhibited in figure 4, for $\rho = 0.9$ and $Q = 0.005$, at times $t = 100, 300$ e 500. The static model (14) is obtained as a particular case with $Q = 0$.

Figura 4: Dynamic impulse response function determined by a random walk on β_t, obtained with $\rho = 0.9$ and $Q = 0.005$, evaluated on three different instants.

The model is completely specified by the parameter vector $\boldsymbol{\theta}$ containing $\boldsymbol{\psi} = (\rho, E_0, \beta_1, \cdots, \beta_T, Q)$, Once the errors υ_t, $t = 1, \cdots, T$ are known, the structural coefficients β_t, $t = 1, \cdots, T$ are easily obtained. Due to computational issues detailed further in section 3 we then work in terms of $\boldsymbol{\psi}_\upsilon = (\rho, E_0, \upsilon_1, \cdots, \upsilon_T, Q)$, whose components follow prior distributions: $\rho \sim U(0,1)$, $E_0 \sim N(m_E, C_E)$, $\upsilon_1 \sim N(a_\upsilon, R_\upsilon)$, $Q \sim GI(\frac{n_\upsilon}{2}, \frac{n_\upsilon s_\upsilon}{2})$.

3 Posterior Distribution and Computational Issues

We assume that $\mathbf{G}_t = \mathbf{G}$ and $\mathbf{W}_t = \mathbf{W}$, constant $\forall t$, assigning the following prior distributions to the components of the parametric vector $\boldsymbol{\theta}$: $\boldsymbol{\delta} \sim NM(\mathbf{a}_\delta, \mathbf{R}_\delta)$, with $\mathbf{R}_\delta$ supposed known, and usually a diagonal matrix, $\boldsymbol{\alpha}_1 \sim NM(\mathbf{a}_1, \mathbf{R}_1)$, and inverse Wishart for $\mathbf{W}$. Hence the posterior distribution associated to the proposed models is given by:

$$\begin{aligned} \pi(\boldsymbol{\theta}|y_1, \cdots, y_T) \propto \quad & \prod_{t=1}^{T} \exp\left\{\tfrac{y_t \chi_t(\theta) - b[\chi_t(\theta)]}{\phi_t}\right\} \\ & \times \exp\left\{-\tfrac{1}{2}(\boldsymbol{\delta} - \mathbf{a}_\delta)' \mathbf{R}_\delta^{-1} (\boldsymbol{\delta} - \mathbf{a}_\delta)\right\} \\ & \times \exp\left\{-\tfrac{1}{2}(\boldsymbol{\alpha}_1 - \mathbf{a}_1)' \mathbf{R}_1^{-1} (\boldsymbol{\alpha}_1 - \mathbf{a}_1)\right\} \\ & \times \prod_{t=2}^{T} \pi(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \mathbf{W}) \\ & \times \pi(\mathbf{W}) \pi(\boldsymbol{\psi}) \prod_{t=1}^{T} \pi(\phi_t). \end{aligned} \tag{21}$$

This distribution does not present closed form and is approached computationally, through Markov Chain Monte Carlo methods. We use Gibbs sampler and in particular, the only parameters whose full conditionals are available are: $\mathbf{W}$, which follows an inverse Wishart posterior distribution and, if we assume inverse Gamma priors for Q_ε and Q, these also have inverse Gamma posterior distributions.

The remaining parameters in $\boldsymbol{\theta}$ are sampled via Metropolis-Hastings steps in the Gibbs sampler. We follow Gamerman (1998), using proposal densities obtained as the full conditionals for each of the components of $\boldsymbol{\theta}$ in the work model:

$$\tilde{y}_t = \eta_t + \tilde{v}_t, \quad \tilde{v}_t \sim N(0, \tilde{V}_t). \tag{22}$$

where η_t follows (6) and $\tilde{y}_t$ are modified observations given by:

$$\tilde{y}_t = \eta_t + (y_t - \mu_t)\dot{g}(\mu_t) \tag{23}$$

with variance:

$$\tilde{V}_t = \tilde{V}_t(\theta_t) = \phi_t \ddot{b}(\chi_t)[\dot{g}(\mu_t)]^2, \tag{24}$$

where $\dot{g}$ and $\ddot{b}$ respectively denote first and second derivatives of the functions g and b.

In dynamic models, the convergence of the Markov chain is perturbed by the inherent autocorrelation in structural parameters such as $\boldsymbol{\alpha}_t$ (6) and β_t (20) (Früwirth-Schnatter, 1994; De Jong and Shephard, 1995; Shephard and Pitt, 1997; Knorr-Held, 1999). We follow Gamerman (1998), expressing such parameters in terms of their respective evolution errors, u_t and υ_t, for which we assume prior serial independence. Then these evolution errors are updated following a single movement Metropolis-Hastings scheme, with proposal densities given by the work model based on (22). Thus, at each step of the algorithm, $\mathbf{u}_1, \cdots, \mathbf{u}_T$ and/or $\upsilon_1, \cdots \upsilon_T$ are sampled. The computational cost implied is the reconstruction of the structural parameters $\boldsymbol{\alpha}_t$ and/or β_t, at each iteration of the MCMC procedure, but we expect that the gain due to the absence of prior autocorrelation overcomes such costs. We assume that the evolution errors covariance matrix $\mathbf{W}_t = \mathbf{W}$ constant $\forall t$ and assign an inverse Wishart distribution to $\mathbf{W}$, basing the inference procedure on $\boldsymbol{\theta}_\mathbf{u} = (\phi_t, \mathbf{u}_1, \cdots, \mathbf{u}_T, \mathbf{W}, \boldsymbol{\delta}', \boldsymbol{\psi}'_\upsilon)$, with $\boldsymbol{\psi}_\upsilon$ written in terms of the evolution errors υ_t, $t = 1, \cdots, T$, in (20) , and hence the resulting posterior distribution is:

$$\begin{aligned} \pi(\boldsymbol{\theta}_\mathbf{u}|D_T) \propto \quad & \prod_{t=1}^{T} \exp\left\{\tfrac{y_t \chi_t(\theta_\mathbf{u}) - b[\chi_t(\theta_\mathbf{u})]}{\phi_t}\right\} \\ & \times \exp\left\{-\tfrac{1}{2}(\boldsymbol{\delta} - \mathbf{a}_\delta)' \mathbf{R}_\delta^{-1} (\boldsymbol{\delta} - \mathbf{a}_\delta)\right\} \\ & \times \prod_{t=1}^{T} \exp\left\{-\tfrac{1}{2}(\mathbf{u}_t)' \mathbf{W}^{-1} (\mathbf{u}_t)\right\} \\ & \times \pi(\mathbf{W}) \pi(\boldsymbol{\psi}_\upsilon) \prod_{t=1}^{T} \pi(\phi_t). \end{aligned} \tag{25}$$

In most time series applications, one aims to predict a vector a vector of future values $\mathbf{y}_f = (y_{T+1}, y_{T+2}, \cdots, y_{T+h})$ given the observed data $\mathbf{y} = (y_1, y_2, \cdots, y_T)$. The predictive distribution is given by:

$$\begin{aligned} p(\mathbf{y}_f|\mathbf{y}) &= \int p(\mathbf{y}_f|\mathbf{y},\theta)\pi(\theta|\mathbf{y})d\theta \\ &= \int p(\mathbf{y}_f|\theta)\pi(\theta|\mathbf{y})d\theta \\ &= E_{(\theta|\mathbf{y})}[p(\mathbf{y}_f|\theta)] \end{aligned} \tag{26}$$

and thus once the posterior distribution is obtained the predictive distribution follows directly. Hence our focus is on the approximation of the posterior distribution (25).

4 Simulated Data

In order to evaluate the identifiability of the proposed models as well as the performance of the MCMC scheme adopted, we have simulated artificial data accordingly to the proposed models. Specifically in the following subsections we present the results of the estimation procedure for two simulated datasets: a binomial response model including in its predictor a dynamic level and a TF(1,0,0) structure and a Poisson response model including a dynamic TF(1,0,0) structure due to dynamic gain factors.

4.1 Binomial(n,p_t) with TF(1,0,0) and Dynamic Level

Assuming the following structure:

$$\begin{aligned} y_t &\sim Binomial(n, p_t) \\ \eta_t &= \alpha_t + E_t + \delta_1 Z_{1t} + \delta_2 Z_{2t} \\ \alpha_t &= \alpha_{t-1} + u_t, \qquad u_t \sim N(0, W), \\ E_t &= \rho E_{t-1} + \beta X_t, \end{aligned} \tag{27}$$

we have simulated 1000 observations $(y_1, \cdots, y_{1000})$ with $n = 5$ and the following values for the components of the parameter vector: $\beta = 0.05$, $\rho = 0.9$, $E_0 = 0.5$, $\delta_1 = -0.1$, $\delta_2 = 0.1$, $W = 0.005$, conditioning on the regressors X_t, Z_{1t} and Z_{2t}, exhibited in figure 5.

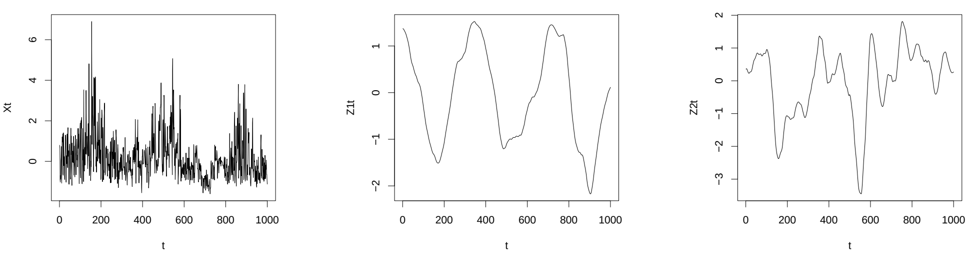

Figura 5: *Binomial(*$5, p_t$*): Covariates used to simulate the binomial data.*

The full conditional distribution for W is Inverse Gamma. For the remaining parameters, whose full conditional is not availible for sampling, we have adopted Metropolis-Hastings steps in the Gibbs sampler.

Following Gamerman (1998), The evolution errors $u_1, \cdots, u_{1000}$ have been sampled individually, with proposal distributions based on (23) and(24). The same proposal has been used to update β, E_0 and the vector (δ_1, δ_2). We have adopted a random walk proposal for logit(ρ).

Histograms obtained after convergence of the MCMC procedure for the samples of the posterior distributions of the parameters involved in model (27) are exhibited in figure 6. Vertical lines indicate the "real" values of each parameter.

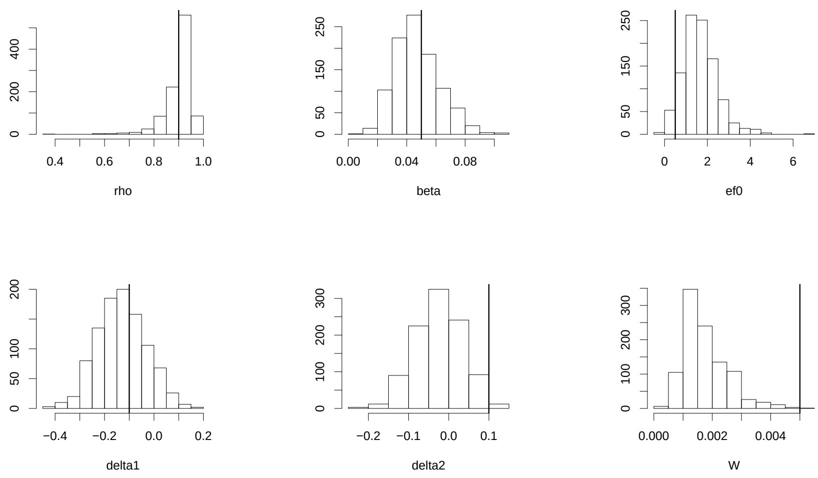

Figura 6: *Binomial(5, p_t) simulated data: Histograms of the posterior distribution samples for the parameters involved in model (27).*

Both transfer function parameters, β and ρ - which account for the immediate impact and the memory of impact of X_t on the mean response - are very well estimated. There is significant uncertainty on the value of the impact of X_t already accumulated in the beginning of the analysis, E_0, which is just as expected. The estimation procedure has been able to recover the value of δ_1, but δ_2 is slightly underestimated, maybe due to the more volatile trajectory of Z_2. Also underestimeted is the variance of the evolution errors, W, which is quite difficult to estimate, once it is not present in the likelihood. Although the trajectory of the posterior mean of dynamic level α_t is quite smooth, its theoretical values have been captured by 95% credibility limits, with only a few isolated points outside these bounds, as can be seen in figure 7.

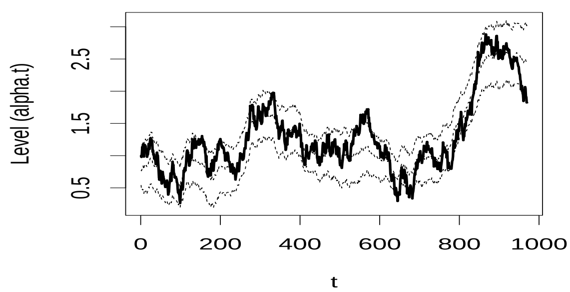

Figura 7: *Binomial(5, p_t) simulated data. Dynamic level α_t: traces of theoretical α_t (bold line), posterior mean and 95 % credibility limits estimated through MCMC.*

The estimation of the evolution of the impact of X_t and its accumulated effect through time, as measured respectively by the impulse response function and transfer function is exhibited in figure 8, with accurate point estimates given by the posterior means of each function. Except for a few points, the mean response function is also captured by 95% credibility bounds, as can be seen in figure 9.

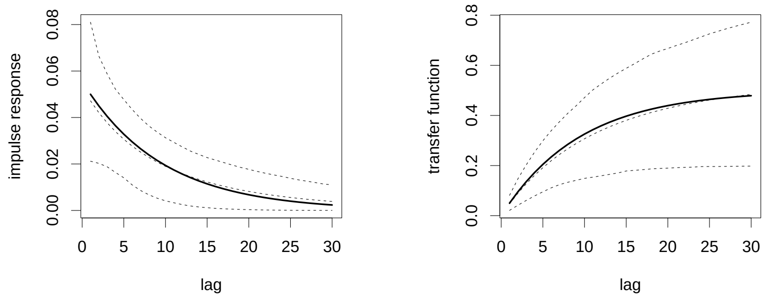

Figura 8: *Binomial($5, p_t$) simulated data. Impulse response and transfer functions: theoretical values (solid line), posterior mean and 95 % credibility limits estimated through MCMC.*

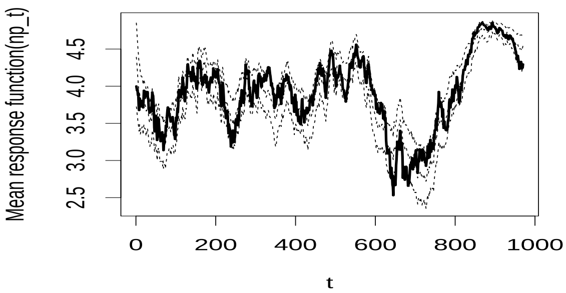

Figura 9: *Binomial($5, p_t$) simulated data. Mean response function np_t: traces of theoretical values (bold line), posterior mean and 95 % credibility limits estimated through MCMC.*

4.2 Poisson(λ_t) with TF(1,0,0) and Dynamic Gain Factors

Using the same regressor variables exhibited in figure 5 and assuming that the immediate impact X_t has on the mean response λ_t may vary through time, we have simulated 1000 points (y_1, y_2, y_{1000}), according to the formulation

$$\begin{aligned} y_t &\sim Poisson(\lambda_t) \\ \eta_t &= \alpha + E_t + \delta_1 Z_{1t} + \delta_2 Z_{2t} \\ E_t &= \rho E_{t-1} + \beta_t X_t, \\ \beta_t &= \beta_{t-1} + \upsilon_t, \qquad \upsilon_t \sim N(0, Q), \end{aligned} \tag{28}$$

adopting the following theoretical values for the parameters: $\alpha = 1.0$, $\rho = 0.7$, $E_0 = 0.1$, $\delta_1 = -0.05$, $\delta_2 = 0.05$ and $Q = 0.0005$.

We have adopted the same MCMC scheme detailed in subsection 27, with single movements based on proposals built from (23) for the parameters, except for Q, whose full conditional is available for sampling and logit(ρ), for which we have adopted a random walk proposal. Figure 10 exhibits histograms of the posterior distribution samples, showing very good results, maybe except for δ_1 and δ_2, both slightly overestimated.

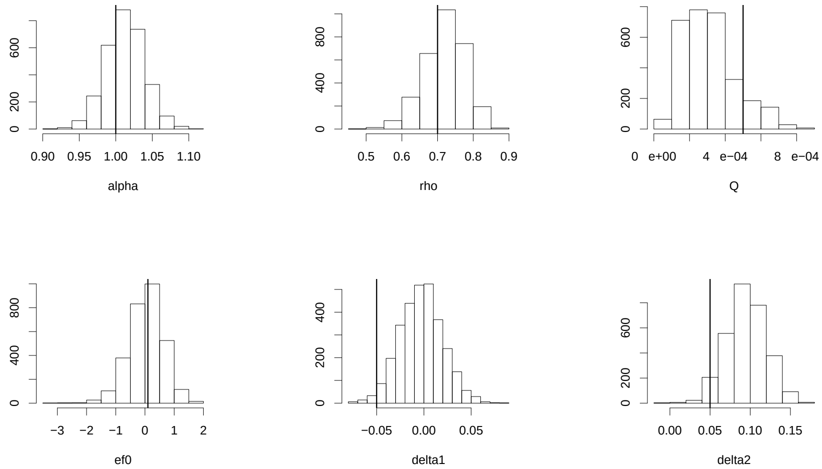

Figura 10: *Poisson(λ_t) simulated data: Histograms of the posterior distribution samples for the parameters involved in model (28).*

As can be seen in figure 10, the posterior distribution of Q exhibits positive skewness, with higher density to values of Q under the theoretical postulated value. Consequently the resulting estimated trajectory of β_t is smoother than its theoretical trajectory, but still very well estimated if we consider the estimated credibility limits exhibited in figure 11.

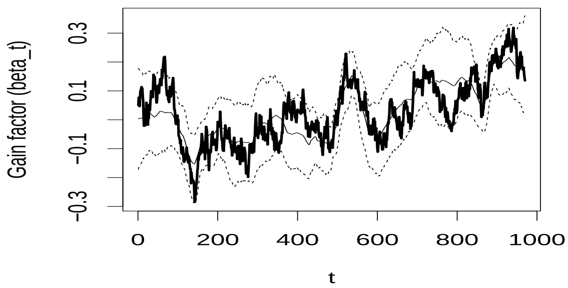

Figura 11: *Poisson(λ_t) simulated data. Dynamic gain factor β_t: traces of theoretical β_t (bold line), posterior mean and 95 % credibility limits estimated through MCMC.*

Figure 12 exhibits the excellent results in the estimation of impulse response functions at three instants and following, in figure 13 it can be seen that the mean response function has also been very well captured by the MCMC scheme.

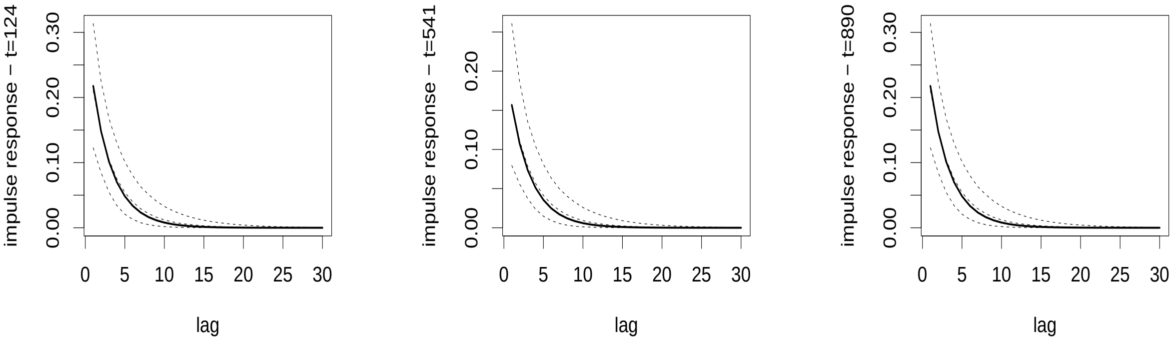

Figura 12: *Poisson(λ_t) simulated data. Transfer response function at times* $t = 124$, $t = 541$ *and* $t = 890$*: theoretical values(solid lines), posterior mean and 95 % credibility limits estimated through MCMC.*

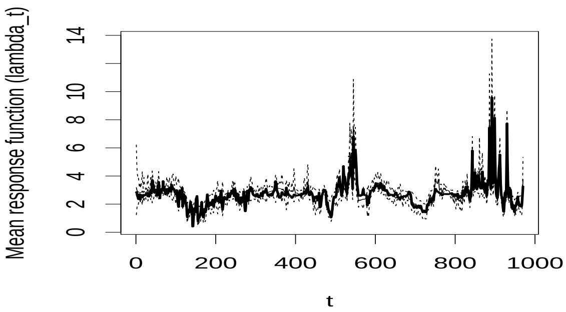

Figura 13: *Poisson(λ_t) simulated data. Mean response function* λ_t*: traces of theoretical values (bold line), posterior mean and 95 % credibility limits estimated through MCMC.*

5 Application: Effect of Rain Volumes on Particulate Material

Particulate matter is a pollutant constituted by liquid and solid material in suspense in atmosphere due to its diminished size (we consider particles with diameter inferior to $10\mu m$). Particulate matter serves as a vehicle to other substances like hydrocarbon and metals, which aggregate to the particles. Larger particles are retained in the superior part of the respiratory system, but the thinner ones may affect pulmonary alveolus. Among the symptoms associated with the inhalation of particulate matter are allergies, asthma e chronic bronchitis. Thus it is usual to consider particulate matter levels as one of the regressors in epidemiological models for such outputs. Nevertheless, it is usual to have periods without information on the pollutant, due to problems in the monitor system. Hence, models which aim to predict pollutant levels are very relevant in this context.

We propose to model particulate matter levels (PM_{10}) in Rio de Janeiro - Brazil, using as regressors climatic variables, such as daily minimum temperature and relative humidity, with particular attention focused on the effect of rain volumes over the levels of such pollutant. We assume that the effect of a rainy day may affect the pollutant levels in the following days, thus adopting transfer functions to model such inertia. We also include calendar dummies in the predictor.

The data refer to 670 observations daily collected in 5 neighborhoods, from September/2000 to August/2002. We do not consider spatial effects and so the the level of the pollutant used for each day in the analysis period is the average level over the monitor stations. Meteorological conditions are registered in four different spots and once again we consider the average level of temperature and humidity over the four measurement stations. Figure 14 exhibits PM_{10} levels ($\mu g/m^3$) recorded in the analysis period and meteorological variables are exhibited in figure 15.

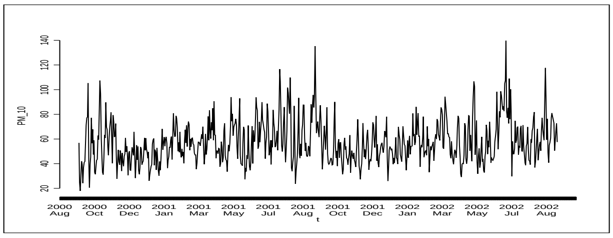

Figura 14: *Daily levels of PM_{10} ($\mu g/m^3$) in Rio de Janeiro, Brazil - September/2000 to August/2002.*

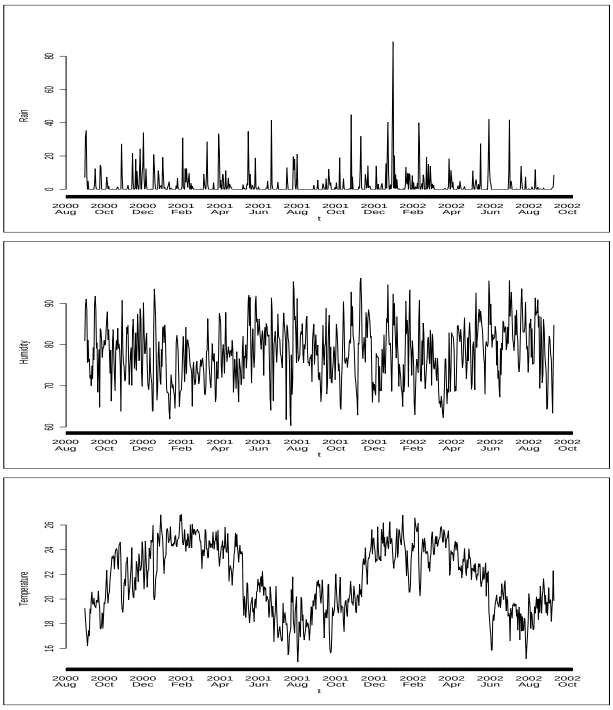

Figura 15: *Meteorological variables in Rio de Janeiro, Brazil - September/2000 to August/2002.*

We denote the PM_{10} level at day t by y_t and the fitted models follow the structure:

$$
\begin{aligned}
\mathrm{y}_t &\sim Gama(\varphi, \lambda_t) \\
\log\left(\frac{\varphi}{\lambda_t}\right) = \eta_t &= \alpha + E_t + \boldsymbol{\delta}'_C \mathbf{C}_t + \boldsymbol{\delta}'_D \mathbf{D}_t + \boldsymbol{\delta}'_S \mathbf{S}_t,
\end{aligned}
$$

in which:

- E_t is a structural block representing the cumulative effect of rain volumes up to time t, expressed by different forms adopted to transfer functions;

- $\mathbf{C}'_t = (Humidity_t, Temperature_t)$;
- $\mathbf{D}'_t = (Mon_t, Tue_t, Q = Wed_t, Thu_t, Fri_t, Sat_t, Hol_t)$ is a vector of dummy variables, accounting for the week days and holidays;
- $\mathbf{S}'_t = \left(Cos\left(\frac{2\pi t}{365}\right), Sin\left(\frac{2\pi t}{365}\right)\right)$. is a vector of seasonal effects.

In table 1 we present the values obtained to model comparisons criteria DIC (Spiegelhalter et al., 2002) and EPD (Gelfand and Ghosh, 1998), for different specifications of the predictor η_t.

Tabela 1: *Comparison of models fitted to daily PM_{10}, considering cumulative impact of rain volumes.*

Predictor Specification	Transfer Function Specification	DIC	EPD
1: $(r=1, s=0)$, $\boldsymbol{\delta}_S = \mathbf{0}$	$E_t = \rho E_{t-1} + \beta Chuva_t$	(p_D=13.1) 5419.7	74.3
2: $(r=1, s=1)$, $\boldsymbol{\delta}_S = \mathbf{0}$	$E_t = \rho E_{t-1} + \beta_0 Rain_t + \beta_1 Rain_{t-1}$	(p_D=13.6) 5416.9	74.0
3: $(r=1, s=0)$, $\boldsymbol{\delta}_S \neq \mathbf{0}$	$E_t = \rho E_{t-1} + \beta Rain_t$	(p_D=13.9) 5301.5	62.4
4: $(r=1, s=1)$, $\boldsymbol{\delta}_S \neq \mathbf{0}$	$E_t = \rho E_{t-1} + \beta_0 Rain_t + \beta_1 Rain_{t-1}$	(p_D=14.2) 5303.8	62.7
5: $(r=0, s=30, d=2)$, $\boldsymbol{\delta}_S = \mathbf{0}$	$E_t = \sum_{j=0}^{30} \beta_j Rain_{t-j}$ $\beta_j = \sum_{k=0}^{2} \zeta_k j^k$	(p_D=13.1) 5431.2	75.7
6: $(r=0, s=30, d=3)$, $\boldsymbol{\delta}_S = \mathbf{0}$	$E_t = \sum_{j=0}^{30} \beta_j Rain_{t-j}$ $\beta_j = \sum_{k=0}^{3} \zeta_k j^k$	(p_D=13.4) 5415.4	74.0
7: $(r=1, s=0)$, $\boldsymbol{\delta}_S = \mathbf{0}$ erros iid	$E_t = \rho E_{t-1} + \beta Rain_t + \varepsilon_t$ $\varepsilon_t \sim N(0, Q_\varepsilon)$	(p_D=236.6) 5399.9	51.3

r=autoregression order, s=number of lags in rain volume, d=degree of polynomial approximation to β_j, $j = 1, \cdots, s$.

According to the DIC criteria, the best models are the seasonal terms included in the predictor, with a slight advantage to model 3, which presents a TF(1,0,0) structure. On the other hand, EPD criteria favours model 7, in which random erros have been introduced in order to account for factors not explicitly included in the model. We have then turned to the predictive likelihood for $\mathbf{y}_f = (y_{671}, \cdots, y_{700})$ resulting from both models. For each model M the predictive likelihood based on a horizon of $h = 30$ days is given by:

$$\begin{aligned} p(\mathbf{y}_f|M, D_T) &= \int\int p(\mathbf{y}_f, \varphi, \boldsymbol{\lambda}_f|M, D_T)d\varphi d\boldsymbol{\lambda}_f \\ &= \int\int p(\mathbf{y}_f|\varphi, \boldsymbol{\lambda}_f, M, D_T)\pi(\varphi, \boldsymbol{\lambda}_f|M, D_T)d\varphi d\boldsymbol{\lambda}_f \\ &= E_{\varphi,\lambda_f|M,D_T}[p(\mathbf{y}_f|\varphi, \boldsymbol{\lambda}_f, M, D_T)], \end{aligned}$$

Once a sample $\boldsymbol{\varphi} = (\varphi^{(1)}, \cdots, \varphi^{(N)})$ and $(\boldsymbol{\lambda}_f^{(1)}, \cdots, \boldsymbol{\lambda}_f^{(N)})$ of the posterior distribution $\pi(\varphi, \boldsymbol{\lambda}_f|M, D_T)$ is avalilable, the Monte Carlo estimate of the predictive likelihood is given by:

$$\begin{aligned} \hat{E}_{\lambda_f|M,D_T}[p(\mathbf{y}_f|\varphi, \boldsymbol{\lambda}_f, M, D_T)] &= \sum_{n=1}^{N} \frac{p(\mathbf{y}_f|\varphi^{(n)}, \boldsymbol{\lambda}_f^{(n)}, M, D_T)}{N} \\ &= \sum_{n=1}^{N} \frac{\prod_{i=1}^{h} p(y_{T+i}|\varphi^{(n)}, \lambda_{T+i}^{(n)}, M, D_T)}{N}. \end{aligned}$$

The predictive likelihood favors model 7, in accordance to EPD criteria. Following we describe the results obtained from this model.

Descriptive statistics associated to the samples of the marginal posterior distributions of each parameter involved in model 7 are presented in table 2. There were 2 days without PM_{10} information in the analysis period. We have formally treated the uncertainty due to these missing points estimating them in the MCMC procedure together with the other unknown quantities.

Tabela 2: *Effect of rain volumes on PM_{10} levels in Rio de Janeiro, Brazil - model 7: Statistics associated to the posterior distribution samples.*

	φ	α	β	ρ	E_0	Q_ε	$\delta.Humid$	$\delta.Temp$	$\delta.Mon$
Min	16.10	4.11	-0.0881	0.572	-0.07	0.005	-0.05	-0.027	0.00
Q1	21.43	4.27	-0.0693	0.696	0.35	0.015	-0.04	-0.004	0.10
Median	24.11	4.32	-0.0651	0.722	0.51	0.019	-0.03	0.002	0.12
Mean	24.81	4.33	-0.0652	0.719	0.53	0.019	-0.03	0.001	0.12
Q3	27.43	4.38	-0.0608	0.746	0.68	0.023	-0.03	0.008	0.14
Max.	48.85	4.66	-0.0447	0.834	1.47	0.041	-0.01	0.032	0.21
s.e.	1.1659	0.0065	0.0005	0.0020	0.0191	0.0016	0.0004	0.0008	0.0020
	$\delta.Tue$	$\delta.Wed$	$\delta.Thu$	$\delta.Fri$	$\delta.Sat$	$\delta.Hol$	y_{478}	y_{479}	
Min	0.05	0.08	0.10	0.15	0.024	-0.24	25.540	29.260	
Q1	0.15	0.18	0.19	0.21	0.091	-0.13	45.030	50.220	
Median	0.17	0.20	0.21	0.23	0.117	-0.10	52.810	59.200	
Mean	0.17	0.20	0.21	0.24	0.116	-0.10	53.830	60.480	
Q3	0.19	0.22	0.23	0.26	0.141	-0.08	60.460	69.160	
Max.	0.27	0.31	0.32	0.34	0.210	0.02	116.100	119.200	
s.e.	0.0021	0.0021	0.0021	0.0020	0.0021	0.0025	0.7420	0.8586	

We notice a growth trend of the pollutant levels towards Fridays, maybe due to the great volume of traffic usual in Rio de Janeiro on Fridays. Such levels decay on Saturdays. The pollutant levels are also significantly reduced at Hllidays. Temperature does not present a significant impact on PM_{10} levels according to this model and humidity presents negative effect. We are currently working on models with transfer function structure assigned to the climatic variables. The negative estimate of β indicates lower levels of PM_{10} on rainy days. Actually, the rain "washes"the atmosphere, depositing the particles on the ground. Besides, such reductive effect has a significant persistence,as can be seen in figure 16 which exhibits impulse response and transfer function estimatesassociated to one standard error raise on rain levels. Figure 17, presents the reduction on the pollutant level caused by the maximum rain volume observed, compared to a non-rainy day. According to the estimated model, such rain volume would cause a 50% decay on PM_{10} and the pollutant would return to its original level in 20 days.

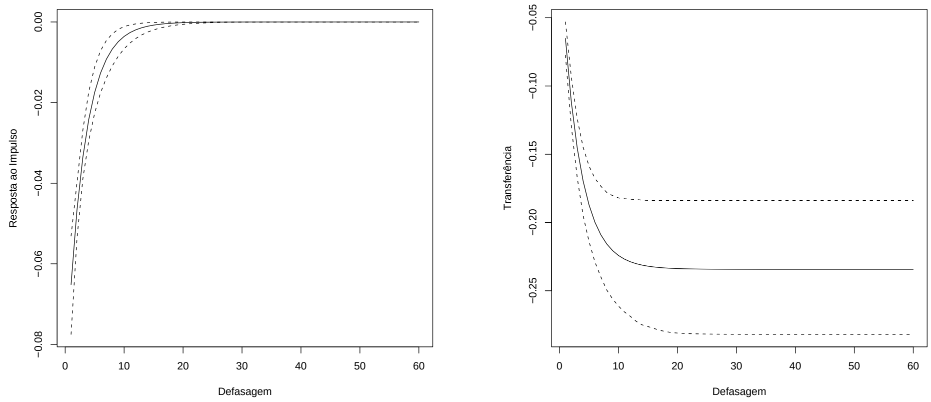

Figura 16: *Effect of rain volumes on PM_{10} in Rio de Janeiro, Brazil- model 7: impulse response and transfer functions associated to one standard error raise on rain volumes.*

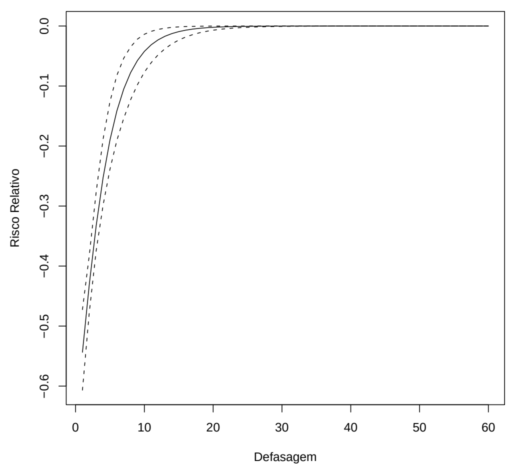

Figura 17: *Effect of rain volumes on PM_{10} in Rio de Janeiro, Brazil- model 7: relative risk associated to the maximum rain volume observed in the analysis period, as compared to non-rainy days.*

Figue 18 exhibits predictive distribution samples' histograms for PM_{10} levels across 30 daysThe vertical lines represent the true observed value of the pollutant on each day. Except for the 12^{th} day in the predictive horizon, in which there has been an unusual high level of pollutant, not captured by the model, the predictive distributions follow very well the observed levels. Predictions based on the estimated predictive mean present a relative error of 17.0%, similar to the median which gives a relative error of 16.8%.

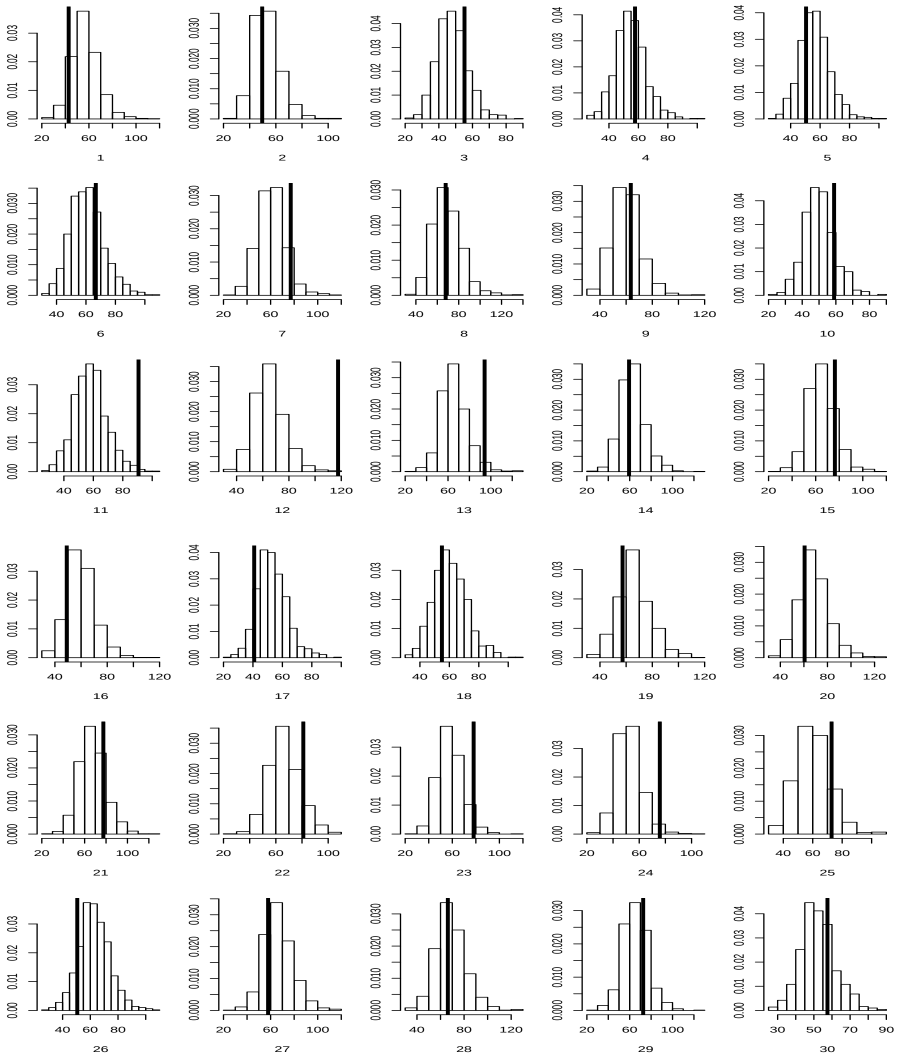

Figura 18: *Efeito de CO sobre óbitos de crianças em SP - modelo 7: Histogramas das amostras das distribuições preditivas para o número de óbitos, com horizontes variando de 1 a 30.*

Referências

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, 33: 178–96.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control.* Prentice-Hall, Englewood Clifs, New Jersey, 3^{rd} edition.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87(418):493–500.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81(3):541–53.

De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, 82(2): 339–50.

Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (com discussão). *Journal of the Royal Statistical Society, série B*, 62(1):3–56.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89:603–15.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association*, 87(418):501–9.

Fahrmeir, L. (1997). Penalized likelihood estimation and iterative Kalman smoothing for non-Gaussian dynamic regression models. *Computational Statistics and Data Analysis*, 24:295–320.

Ferreira, M. A. R. and Gamerman, D. (2000). Dynamic generalized linear models. Em Dey, D. K., Ghosh, S. K., and Mallick, B. K., editores, *Generalized Linear Models*, pages 57–72. Marcel Dekker Inc.

Früwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Jounal of Time Series Analysis*, 15:183–202.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika*, 85: 215–27.

Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85:1–11.

Geweke, J. and Tanizaki, H. (2001). Bayesian estimation of state-apace models using the Metropolis-Hastings algoriyhm within Gibbs sampling. *Computational Statistics and Data Analysis*, 37:151–70.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian bayesian state estimation. *IEE Proceedings-F*, 140(2):107–13.

Harrison, P. J. and Stevens, C. (1976). Bayesian forecasting (com discussão). *Jounal of the Royal Statistical Society, série B*, 38:205–47.

Klein, L. R. (1958). The estimation of distributed lags. *Econometrica*, 26(4):553–65.

Knorr-Held, L. (1999). Conditional prior proposals in dynamic models. *Scandinavian Journal of Statistics*, 26:129–44.

Koyck, L. (1954). *Distributed Lags and Investment Analysis.* North Holland, Amsterdam.

Migon, H. S., Gamerman, D., Lopes, H. F., and Ferreira, M. A. R. (2005). Bayesian dynamic models. Em Dey, D. and Rao, C.R., editores, *Handbook of Statistics*, volume 25, pages 553–88.

Pole, A. (1988). Transfer response models: a numerical approach. Em Bernardo, J., DeGroot, M., Lindley, D., and Smith, A., editores, *Bayesian Statistics*, volume 3, pages 733–46. Oxford University Press.

Ravines, R., Migon, H., and Schmidt, A. (2006). Revisiting distributed lag models through a bayesian perspective (a ser publicado). *Applied Stochastic Models in Business and Industry*.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84:653–67.

Singh, A. C. and Roberts, G. R. (1992). State space modelling of cross-classified time series of counts. *International Statistics Review*, 60:321–36.

Solow, R. M. (1960). On a family of lag distributions. *Econometrica*, 28(2):393–406.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, série B*, 64:583–639.

West, M., Harrison, P. J., and Migon, H. (1985). Dynamic generalized linear model and bayesian forecasting (with discussion). *Journal of the American Statistical Association*, 80(389):73–97.