# Inference for the hyperparameters of structural models under classical and Bayesian perspectives: a comparison study

Thiago Rezende dos Santos[a]*and Glaura C. Franco[b]

[a]*Department of Mathematics, UFOP, MG, Brazil*
[b]*Department of Statistics, UFMG, MG, Brazil*

July 13, 2010

**Abstract**

Structural models, or dynamic linear models as they are known in the Bayesian literature, have been widely used to model and predict time series using a decomposition in non-observable components. Due to the direct interpretation of the parameters, structural models are a powerful and simple methodology to analyse time series in several areas, such as economy, climatology, environmental sciences, among others. The parameters of such models can be estimated either using maximum likelihood or Bayesian procedures, generally implemented using conjugate priors, and there are plenty of works in the literature employing both methods. But are there situations where one of these approaches should be preferred? In this work, instead of conjugate priors for the hyperparameters, the Jeffreys prior is used in the Bayesian approach, along with the uniform prior, and the results are compared to the maximum likelihood method, in an extensive Monte Carlo study. Interval estimation is also evaluated and, to this purpose, bootstrap confidence intervals are introduced in the context of structural models and their performance is compared to the asymptotic and credibility intervals. A real time series of a Brazilian electric company is used as illustration.

*Corresponding author: Thiago Rezende dos Santos. Address: Universidade Federal de Ouro Preto, Morro do Cruzeiro, Ouro Preto, MG, 35400-000, Brazil. E-mail: thiagords@ufmg.br

Keywords: state space models, dynamic linear models, bootstrap, confidence and credibility intervals.

# 1 Introduction

One way of modelling a time series is through its decomposition in non-observable components by means of a state space formulation. Under a classical perspective, the modelling is performed using structural models, proposed by Harvey (1989). A very similar approach can be obtained by using a Bayesian procedure, which in this case is known as dynamic linear models (DLM) (West and Harrison; 1997). Parameter estimation can be performed by estimating the system matrices obtained from the state space equations. Both methods, classical and Bayesian, use the Kalman filter (Kalman; 1960) to obtain the estimates of the latent variables, that represent the non-observable components, and their variance, named hyperparameters, but differ in the way the parameter estimation is performed.

In the classical approach, the parameter estimates are obtained maximizing the likelihood function and inferences can be done via the asymptotic distribution of the parameters (Harvey; 1989). More details about the use of structural models can be found in Harvey, Koopman and Shephard (2004), Durbin and Koopman (2001) and Harvey and Proietti (2005).

Under the Bayesian approach, the parameter estimation may be quite cumbersome, as most of the times the posterior distribution of the parameters is extremely complicated or can even not exist. Due to this difficulty, Markov chain Monte Carlo (MCMC) methods are often employed (Gamerman and Lopes; 2006, see) to perform the estimation. Although nowadays there are some alternative procedures, like particle learning filters (Carvalho et al.; 2009), that can be computationally faster, the MCMC method was chosen here because of its robustness and easiness of implementation. Recently, many authors have worked on the Bayesian parameter estimation in dynamic linear models using MCMC (see, for instance, Carter and Kohn (1994), Lopes, Moreira and Schmidt (1999), Reis, Salazar and Gamerman (2006) and Schmidt, Gamerman and Moreira (1999)). **Especially for trend and cyclical models, the Bayesian approach can yield more informative results by using prior information for the periodicity (Harvey, Trimbur and Dijk; 2007).**

Another way of making inferences about the parameters, with a significa-

tive gain for short series, is using the bootstrap. The most common procedure to perform the bootstrap in time series is to resample the residuals of the fitted model, which are generally uncorrelated if there is not order misspecification (Efron and Tibshirani; 1993). This procedure is simple to apply and leads to good theoretical behaviour of the estimates, although it is not robust against violation of the model assumptions. Nevertheless, there are many papers in the literature employing this approach in the state space context, such as Franco et al. (2008), Franco and Souza (2002), Pfeffermann and Tiller (2005), Stoffer and Wall (1991), Wall and Stoffer (2002) and Rodrigues and Ruiz (2009), just to cite a few.

According to the previous paragraphs, there is a fairly large literature applying classical and Bayesian procedures to state space models, whereas research comparing the two methodologies is very restricted. Thus, it would be an interesting issue to evaluate the performance of both approaches, addressing features such as bias and mean square errors of the point estimation procedures as well as interval estimation for the parameters.

Thus, the main contribution of this work is to implement the bootstrap and Bayesian methods, comparing these methodologies to the classical maximum likelihood approach to make inferences for the hyperparameters in structural models. In order to make a fair comparison between classical and Bayesian methods and also because there is generally little information about the behavior of the hyperparameters, Bayesian inference will be based on the least possible subjective information. Therefore, instead of the inverse gamma distribution usually employed in the literature as prior for the variance of the non-observable components, uniform and Jeffreys priors are used, implemented using a hybrid version of the Metropolis-Hastings algorithm, once the priors are not conjugated. The Local Level Model (LLM), Local Linear Trend Model (LLT) and Structural Basic Model (SBM) are used to this purpose.

In addition, 95% confidence and credibility intervals for the hyperparameters are built. The confidence intervals include the asymptotic and some bootstrap intervals, namely the percentile, the bias corrected and the bias corrected and accelerated (Efron and Tibshirani; 1986). The procedures are compared based on the percentage of times the intervals contain the real value of the hyperparameter, in a sequence of Monte Carlo simulations. An application to a real time series is also presented, as an illustration of the methodologies.

This paper is organized as follows. Section 2 presents the structural

models considered in this work, along with classical and Bayesian parameter estimation. Section 3 presents the computational procedures used, that is, bootstrap and MCMC. In Section 4 are the simulation results obtained for estimation and confidence intervals. Section 5 applies the methodologies to a real time series and Section 6 concludes the work.

# 2 Structural model and parameter estimation

A univariate time series $y_t$ can be decomposed in a sum of its unobserved components, such as trend ($\mu$), seasonal ($\gamma$) and disturbances ($\epsilon$). A procedure that has been widely used to model a series based on this approach is the structural model (Harvey; 1989; West and Harrison; 1997, see). Following this procedure, a time series $y_t, \quad t = 1, ..., n$, that presents these characteristics can be written as

$$y_t = \mu_t + \gamma_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad i.i.d. \tag{1}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \quad i.i.d. \tag{2}$$

$$\beta_t = \beta_{t-1} + \xi_t, \quad \xi_t \sim N(0, \sigma_\xi^2) \quad i.i.d. \tag{3}$$

$$\gamma_t = \gamma_{t-1} - ... - \gamma_{t-s+1} + \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2) \quad i.i.d. \tag{4}$$

where $s$ is the number of seasonal periods and $\epsilon_t$, $\eta_t$, $\xi_t$ and $\omega_t$ are white noise disturbances mutually uncorrelated. The above model is known as the Structural Basic Model (SBM). **Although there are nowadays several representations for linear seasonal models in the structural approach (see Proietti (2000) for details), this work will use the form given in Equation (4), as stated in Harvey and Todd (1983), as this is a commomly used specification in empirical works.**

If equation (4) and the $\gamma_t$ component in equation (1) are dropped, the resulting model is called the Local Linear Trend Model (LLT). This model arises when the series presents a trend with an increasing (or decreasing) slope along time. The simplest model is the Local Level Model (LLM). It

arises when the series moves along time as a random walk. To write this model, only equations (1) and (2) are used, dropping the terms $\beta_{t-1}$ and $\gamma_t$.

Equations (1)-(4) are usually written in the state space form, to allow the use of linear algorithms, such as the Kalman filter (Kalman; 1960), in the estimation of the unobserved components. In this form, there are only two equations, the observation and the state equations:

$$y_t = \mathbf{z}_t'\boldsymbol{\alpha}_t + d_t + \epsilon_t, \quad \epsilon_t \sim N(0, h_t) \tag{5}$$

$$\boldsymbol{\alpha}_t = \mathbf{T}_t\boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \mathbf{R}_t\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \tag{6}$$

where $\boldsymbol{\alpha_t}$ is the state vector, $\mathbf{z_t}$, $\mathbf{T_t}$ and $\mathbf{R_t}$ are system matrices, $d_t$ and $\mathbf{c_t}$ are covariates, $\epsilon_t$ are uncorrelated disturbances with variance $h_t$, $\boldsymbol{\eta_t}$ is a vector of serially uncorrelated disturbances whose covariance matrix is given by $\mathbf{Q_t}$ and $\boldsymbol{\eta_t}$ and $\epsilon_t$ are independent. Once the model is in the state space form, the Kalman filter can be applied to estimate the hyperparameter vector, $\boldsymbol{\psi} = (\psi_1, \psi_2, ..., \psi_p)$, which in this case are the variances of the disturbances.

The assumptions of the model are:

(1) $E(\alpha_0) = a_0$ and $Cov(\alpha_0, \alpha_0') = P_0$, where $\alpha_0$ is the initial state.

(2) $E(\epsilon_t, \boldsymbol{\eta_t}) = 0$ and $E(\epsilon_t, \boldsymbol{\alpha_t}) = E(\boldsymbol{\eta_t}, \boldsymbol{\alpha_t}) = 0$, for all $t = 1, ..., n$.

The estimation of the hyperparameter vector $\boldsymbol{\psi}$ can be done either using classical methods (i.e, maximizing the likelihood function) or Bayesian methods. In Subsections 2.1 and 2.2 the classical and Bayesian estimation methods considered in this work are described in some detail.

## 2.1 Classical inference

The likelihood function can be obtained through the one-step ahead prediction error, $\nu_t = y_t - \widetilde{y}_{t|t-1}$, calculated using the Kalman filter, assuming that $(y_t|Y_{t-1}) \sim N(\widetilde{y}_{t|t-1}, F_t)$, where $Y_{t-1}$ are the observations up to time $t-1$ and $\widetilde{y}_{t|t-1}$ is the one-step ahead forecasting. For a univariate time series of size $n$, the logarithm of the likelihood function is given by

$$lnL(\boldsymbol{\psi}; Y_n) = ln\prod_{t=1}^{n} p(y_t|Y_{t-1}) = -\frac{n}{2}ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n} ln|F_t| - \frac{1}{2}\sum_{t=1}^{n} \nu_t' F_t^{-1}\nu_t. \tag{7}$$

The estimated hyperparameter vector is the one that maximizes the above likelihood function. As this is a nonlinear function of the hyperparameters, the estimation should be done numerically. In this work, the well-known BFGS optimization algorithm is employed (more details can be seen in Franco et al. (2008)). As in this case the parameters are variances, and therefore positive quantities, a reparameterization should be performed to avoid the problem of obtaining negative variances. Thus, a **logarithmic** transformation is used to take into account the nonnegativity constraints.

Harvey (1989) states that, under some regularity conditions, the maximum likelihood estimator (MLE), $\hat{\boldsymbol{\psi}}$, is asymptotically normal with mean $\boldsymbol{\psi}$ and covariance matrix $Var(\boldsymbol{\psi}) = n^{-1}\mathbf{IA}^{-1}(\boldsymbol{\psi})$, in which

$$\mathbf{IA}(\boldsymbol{\psi}) = \lim_{n\to\infty} n^{-1}\mathbf{I}(\boldsymbol{\psi}),$$

and $\mathbf{I}(\boldsymbol{\psi})$ is the Fisher information matrix, given by

$$I_{ij}(\boldsymbol{\psi}) = \frac{1}{2}\sum_t \left\{ \text{tr}\left[ F_t^{-1}\frac{\partial F_t}{\partial \psi_i} F_t^{-1}\frac{\partial F_t}{\partial \psi_j}\right]\right\} + \text{E}\left\{ \sum_t \left(\frac{\partial \nu_t}{\partial \psi_i}\right)' F_t^{-1}\frac{\partial \nu_t}{\partial \psi_j}\right\}, \qquad (8)$$

where $i, j = 1, ..., p$. For all structural models considered in this work, the regularity conditions cited above are satisfied if $\boldsymbol{\psi}$ is an interior point of the parameter space. If one or more of the elements of $\boldsymbol{\psi}$ lies on the boundary of the parameter space, the joint asymptotic distribution of the MLE can be affected (Harvey; 1989).

The analytical form of $\mathbf{I}(\boldsymbol{\psi})$ is not straightforward to compute in this case. In order to simplify the computation of the Fisher information matrix, the expectation operator can be omitted in expression (8), once this new expression is asymptotically equivalent. Franco et al. (2008) show a numerical way of calculating the asymptotic variances of the hyperparameters, based on a suggestion of Harvey (1989) to calculate the derivatives of $\nu_t$ and $F_t$ in the Kalman filter.

A $100(1-\kappa)\%$ asymptotic confidence interval for $\psi_i$, $i = 1, ...p$ is given by

$$\hat{\psi}_i \pm z_{\kappa/2}\sqrt{\text{Var}(\hat{\psi}_i)},$$

in which $z_{\kappa/2}$ is the $\kappa/2$ percentile of the Normal distribution, and $Var(\hat{\psi}_i)$ is obtained from the elements in the diagonal of $Var(\hat{\boldsymbol{\psi}})$.

## 2.2 Bayesian inference

If a prior $\pi(\boldsymbol{\psi})$ for $\boldsymbol{\psi}$ is specified, then the posterior distribution is given by

$$\pi(\boldsymbol{\psi} \mid Y_n) = \frac{L(\boldsymbol{\psi}; Y_n)\pi(\boldsymbol{\psi})}{\int L(\lambda; Y_n)\pi(\lambda)d\lambda}. \tag{9}$$

Bayesian estimation depends on the specification of a prior for all unknown parameters, but most of the times it is not feasible or even impossible. In this case reference priors, obtained through some formal method, can be used. The most common one is the Jeffreys prior (Migon and Gamerman; 1999), which in this case is calculated through a numerical approximation of the Fisher information matrix, $\mathbf{I}(\boldsymbol{\psi})$, and given by

$$\pi(\boldsymbol{\psi}) \simeq det(\mathbf{I}(\boldsymbol{\psi}))^{1/2} \tag{10}$$

where $\mathbf{I}(\boldsymbol{\psi})$ is approximated in the same way as in Section 2.1.

In this work a proper uniform prior is also considered. The prior for the state parameters $(\alpha_1, \ldots, \alpha_n)'$, given $\boldsymbol{\psi}$, can be obtained with the specification in (6) and initial values for $\alpha_0$.

The most used Bayes estimators are the posterior mean, obtained when the quadratic loss function is used, the posterior median, obtained when the absolute loss function is used, and the posterior mode, obtained when the 0-1 loss function is used (Migon and Gamerman; 1999).

Credibility intervals for $\psi_i$, $i = 1, ...p$, can be built as follows. Given $\kappa$, any interval $(t_1, t_2)$ satisfying

$$\int_{t_1}^{t_2} \pi(\psi_i \mid Y_n)d\psi_i = 1 - \kappa \tag{11}$$

is a credibility interval for $\psi_i$ with level $100(1-\kappa)\%$.

# 3 Computational procedures

In this section the computational procedures used to make inferences about the hyperparameter vector are described, namely the bootstrap in the

classical approach and the Metropolis-Hastings algorithm in the Bayesian approach.

## 3.1 Bootstrap

The bootstrap method is a resampling technique that can be used to approximate the theoretical distribution by the empirical distribution of a finite sample of observations. In the context of time series the observations are not independently distributed and therefore the observed series needs to be 'whitened', which means that the serial correlation in the observations must be removed. When the true correlations are identified, the residuals of the model are independent. The bootstrap technique can therefore be applied to the residuals of the model whether the distribution is known (parametric bootstrap) or not (nonparametric bootstrap).

The bootstrap procedure used here is based on the work of Stoffer and Wall (1991). The Kalman filter computes the one-step forecasting errors, which are independently distributed for a correctly specified model. Standard bootstrap methods are applied to the forecast errors and a simple modification is used to obtain the bootstrap for the original observations.

The full procedure works as follows. Using the Kalman filter, the innovations, $\nu_t$, and their variances, $F_t$, are obtained recursively. Next, the innovations should be centered and re-scaled to account for problems as non-zero means or heteroscedastic residuals. The bootstrap innovations, $e_t^*$, are obtained by resampling the centered and re-scaled innovations, with replacement. Let the vector $\boldsymbol{S}_t$ be defined as

$$\boldsymbol{S}_t = \begin{bmatrix} \boldsymbol{a}_{t+1|t} \\ y_t \end{bmatrix},$$

in which $\boldsymbol{a}_{t+1|t}$ is a linear estimator for the state vector $\boldsymbol{\alpha}_{t+1}$, based on $Y_t = (y_1, \ldots, y_t)'$, with variance $\boldsymbol{P}_{t+1|t}$. Then vector $\boldsymbol{S}_t$ is updated as

$$\boldsymbol{S}_t = \begin{bmatrix} \boldsymbol{T}_t & \boldsymbol{0} \\ \boldsymbol{z}_t & \boldsymbol{0} \end{bmatrix} \boldsymbol{S}_{t-1} + \begin{bmatrix} \boldsymbol{T}_t \boldsymbol{K}_t \sqrt{F_t} \\ \sqrt{F_t} \end{bmatrix} \nu_t, \tag{12}$$

where $\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1} \boldsymbol{z}_t' F_t^{-1}$ is the Kalman gain's matrix. The bootstrap series $y_t^*$ can be calculated solving Eq. (12) by substituting $\nu_t$ by $e_t^*$ and using the estimated values $F_t$ and $\boldsymbol{K}_t$ obtained from the Kalman filter. The initial conditions of the Kalman Filter are fixed at their given values and $\boldsymbol{\psi}$ is held fixed at $\hat{\boldsymbol{\psi}}$. The MLE for the bootstrap series, $y_t^*$, will be called here $\hat{\boldsymbol{\psi}}^*$.

Three methods of constructing bootstrap confidence intervals, proposed in Efron and Tibshirani (1986), are considered in this study. The first procedure, the percentile interval, although being reliable, does not have satisfactory coverage probabilities. Thus, Efron and Tibshirani (1993) proposed two improved versions of the percentile method, the bias-corrected and the bias-corrected and accelerated method. These later two procedures are shown to perform better than the former one, in the sense that they can approach the exact confidence intervals, when such intervals can be calculated, and have accurate coverage probabilities in most situations, apart from some small sample cases.

For each one of the methods described below, $B$ bootstrap series $y^{*1}, y^{*2}, ..., y^{*B}$ are generated and the estimated hyperparameter vector, $\hat{\boldsymbol{\psi}}^* = (\hat{\psi}_1^*, ..., \hat{\psi}_p^*)$, is calculated for each series.

**i) Percentile (PERC)**

Taking the $\kappa/2$ and $1-\kappa/2$ percentiles of the bootstrap distribution of $\hat{\psi}_i^*$, $i = 1, ..., p$, the $100(1-\kappa)\%$ percentile interval is defined by

$$[\hat{\psi}_i^{*(\kappa/2)}, \hat{\psi}_i^{*(1-\kappa/2)}].$$

Thus, in practice, after estimating the values of $\psi_i$ for each of the $B$ bootstrap series, they are ordered and the $100(\kappa/2)^{th}$ and $100(1-\kappa/2)^{th}$ values are taken as the lower and upper points of the intervals, respectively.

**ii) Bias-corrected (BC)**

This method also uses the percentiles of the bootstrap distribution, but not exactly the $(\kappa/2)^{th}$ and $(1-\kappa/2)^{th}$. Instead, it corrects these values for possible bias in the estimation of $\psi_i$, through a quantity $m_0$ which measures the median bias of $\hat{\psi}_i$.

The $100(1-\kappa)\%$ bias-corrected interval is defined by

$$[\hat{\psi}_i^{*(\kappa_1)}, \hat{\psi}_i^{*(\kappa_2)}]$$

where $\kappa_1 = \Phi\left(2m_0 + z^{(\kappa/2)}\right)$, $\kappa_2 = \Phi\left(2m_0 + z^{(1-\kappa/2)}\right)$, $\Phi$ is the cumulative distribution function of a $N(0,1)$ and $z^{(\kappa)}$ its $100\kappa^{th}$ percentile point.

The value of $m_0$ is calculated using the proportion of $\hat{\psi}_i^*$ in the bootstrap samples that are smaller than the $\hat{\psi}_i$ estimated in the original series,

$$m_0 = \Phi^{-1}\left(\frac{\#\hat{\psi}_i^*(b) < \hat{\psi}_i}{B}\right).$$

**iii)Bias-corrected and accelerated (BCa)**

The BCa interval, besides calculating the quantity $m_0$, also calculates the *acceleration*, given by

$$a = \frac{\sum_{j=1}^{n}(\hat{\psi}_i(.) - \hat{\psi}_i(j))^3}{6\left[\sum_{j=1}^{n}(\hat{\psi}_i(.) - \hat{\psi}_i(j))^2\right]^{3/2}}$$

where $\hat{\psi}_i(j)$ is the hyperparameter estimate calculated for the sample with the $j^{th}$ observation deleted and $\hat{\psi}_i(\cdot) = \sum_{j=1}^{n}\hat{\psi}_i(j)/n$.

Thus, the $100(1-\kappa)\%$ BCa interval is given by

$$[\hat{\psi}_i^{*(\kappa_1)}, \hat{\psi}_i^{*(\kappa_2)}]$$

with $\kappa_1 = \Phi\left(m_0 + \frac{m_0 + z^{(\kappa/2)}}{1 - a(m_0 + z^{(\kappa/2)})}\right)$ and $\kappa_2 = \Phi\left(m_0 + \frac{m_0 + z^{(1-\kappa/2)}}{1 - a(m_0 + z^{(1-\kappa/2)})}\right)$.

The BCa interval is second order accurate and it has also the advantage of being transformation-respecting (Efron and Tibshirani; 1993).

## 3.2 Metropolis-Hastings Algorithm

Markov chain Monte Carlo (MCMC) methods have been widely used in the Bayesian inference, as they enable the generation of a sample of the posterior distribution of a parameter or a random object of interest using algorithms as the Gibbs sampling and Metropolis-Hastings. The main idea of the MCMC methods is to construct a Markov chain from which it is easy to generate a trajectory and that leads to an equilibrium distribution, after a sufficiently large number of iterations, equal to the distribution of interest (Gamerman and Lopes; 2006, see).

In this study, a hybrid version of the Metropolis-Hastings (M-H) algorithm (Hastings; 1970; Metropolis et al.; 1953) is adopted to obtain the mean and the quantiles of the distribution of interest. Given a target distribution $\pi(\cdot)$, the Metropolis-Hastings algorithm and its variants enable the creation of ergodic Markov chains that have $\pi(\cdot)$ as the equilibrium distribution. This

algorithm is used to simulate a posterior distribution through a candidate-generating or transition density, $q(y|x)$, from a state $x$ to another state $y$, with the following properties

- $\int q(y|x)dy = 1$;
- $q(y|x)$ can be evaluated for all $x$ and $y$;
- for each $x$ it is possible to generate realizations with distribution $q(.|x)$.

In the hybrid version, the $p$ hyperparameters, $(\psi_1, \ldots, \psi_p)$, are updated separately, with different candidate-generating densities.

The M-H algorithm is initiated from an arbitrary point, $\boldsymbol{\psi}^{(0)}$, and evolves to the next point of the chain through the candidate-generating density in the following way

1. Generate $\psi_i^{(j)} \sim q_i(.|\psi_i^{(j-1)})$, $i = 1, ...p$;
2. Calculate the Hastings ratio
   $R_i = \left[\pi_i(\psi_i^{(j)})q_i(\psi_i^{(j-1)}|\psi_i^{(j)})\right] / \left[\pi_i(\psi_i^{(j-1)})q_i(\psi_i^{(j)}|\psi_i^{(j-1)})\right]$,
   where $\pi_i(\psi_i)$ is the full conditional of $\psi_i$;
3. Fix the acceptance probability equal to $\min\{1, R_i\}$;
4. Obtain the next value of the chain as $\psi_i^{(j)}$ with probability $\min\{1, R_i\}$, or $\psi_i^{(j-1)}$ with probability $1 - \min\{1, R_i\}$;
5. repeat the above steps until the convergence of the chain.

The acceptance probabilities $R_i$ are defined to ensure a reversible chain and its convergence to the equilibrium distribution $\pi(\cdot)$.

There are several possibilities for candidate-generating densities in the literature. In this work, a random walk centered in the last accepted estimate of the hyperparameter, $q_i(\psi_i^{(j)}|\psi_i^{(j-1)}) = N(\psi_i^{(j-1)}, \sigma_i^2)$, $i = 1, ...p$, is used. The values of $\sigma_i^2$ are defined such that the acceptance rates are between 20% and 50% (Chib and Greenberg; 1995).

Like the classical approach, sometimes constraints in the parametric space are necessary, as for example in the case of variances. In these situations, the proposed density should be restricted to this interval and the proposition

distributions are defined as truncated normals to guarantee that the properties defined above are preserved. Due to this fact, a correction is performed in Hastings ratio (step 2). This approach considers the marginal posterior distribution of the hyperparameters once the state parameters are integrated out. This is equivalent to sampling all parameters jointly since the full conditional for the state parameters given the hyperparameter is analytically available. Reis, Salazar and Gamerman (2006) provide substantial empirical evidence in favor of this joint sampling scheme against other blocking procedures.

There are different ways of obtaining full conditional posterior samples of $(\alpha_0, \ldots, \alpha_n)'$. One way is to explore statistical properties of the model in the following way,

$$\pi(\alpha_0, \ldots, \alpha_n | \boldsymbol{\psi}, \boldsymbol{Y}_n) = \pi(\alpha_n | \boldsymbol{\psi}, \boldsymbol{Y}_n) \prod_{t=0}^{n} \pi(\alpha_t | \alpha_{t+1}, \boldsymbol{\psi}, \boldsymbol{Y}_t) \tag{13}$$

where $\pi(\alpha_t | \alpha_{t+1}, \boldsymbol{\psi}, \boldsymbol{Y}_t)$ is given by

$$\begin{array}{l}(\alpha_t | \alpha_{t+1}, \boldsymbol{\psi}, \boldsymbol{Y}_t) \sim N[(\boldsymbol{T}_t' \boldsymbol{Q}^{-1} \boldsymbol{T}_t + \boldsymbol{P}_t^{-1})^{-1} (\boldsymbol{T}_t' \boldsymbol{Q}^{-1} \alpha_{t+1} + \boldsymbol{P}_t^{-1} \boldsymbol{a}_t), \\ (\boldsymbol{T}_t' \boldsymbol{Q}^{-1} \boldsymbol{T}_t + \boldsymbol{P}_t^{-1})^{-1}]\end{array} \tag{14}$$

with $t = 0, \ldots, n-1$.

The results are used in the sampling scheme proposed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994), called Forward Filtering Backward Smoothing (FFBS). The idea is to sample from $(\alpha_0, \ldots, \alpha_n | \boldsymbol{\psi}, \boldsymbol{Y}_n)$ according to the following algorithm:

1. Sample $\alpha_n$ from its updated distribution and set $t = n - 1$;
2. Sample $\alpha_t$ from the distribution in (14);
3. Set $t = t - 1$ and return to step 2 until $t = 0$.

The step 1 is obtained updating the Kalman Filter from $t = 0$ to $t = n$.The updated mean $\boldsymbol{a}_t$ and covariance matrix $\boldsymbol{P}_t$ are saved and used in step 2 of the algorithm.

Another way of sampling $(\alpha_0, \ldots, \alpha_n)'$ can be done through its full conditional. In particular, fast inversion algorithms can be used ensuring that

samples from $(\alpha_0, \ldots, \alpha_n)'$ are quickly drawn. More details about the sampling schemes of $(\alpha_0, \ldots, \alpha_n)'$ can be seen in Migon et al. (2005) and Reis, Salazar and Gamerman (2006).

Another important aspect is the evaluation of the convergence of the algorithm. One of the most popular ways to check the convergence is using the method of Geweke (1992), available in R in the package Coda (Plummer et al.; 2005). The trajectory (trace) of the generated chain can also be observed, as suggested by Gamerman and Lopes (2006), to search for qualitatively similar movements, to verify if the convergence to the stationary distribution was attained. Besides, the autocorrelation function of the generated chains must show a fast decrease, characteristic of stationary distributions.

# 4 Simulation results

In this section the results obtained from Monte Carlo (MC) experiments, implemented in the Ox language (Doornik; 1999), are presented. The performance of the maximum likelihood (MLE) and Bayes estimators (BE) - Mean, Median and Mode - were evaluated for the LLM model with several combination of hyperparameters. The values were chosen in order to obtain the following signal-to-noise ratios: $q = 0.02$, $q = 0.50$, $q = 1.00$ and $q = 2.00$, where $q = \sigma_\eta^2/\sigma_\epsilon^2$. In this case, series of size $n = 100$ and 300 were generated with a burn-in equal to 100. For the LLT and SBM models, only one combination of hyperparameter values is presented, with series of size $n = 200$.

For the BE, a chain with 20,000 samples was generated for each parameter and, as there was no thinning, the last 4,000 samples were kept. The method of Geweke (1992) and graphics of the estimated autocorrelation functions for the chains of hyperparameters were used to assess the convergence of the MCMC algorithm. The number of MC and bootstrap replications were fixed at 500. Uniform and Jeffreys priors were considered for the hyperparameters and the prior for $\alpha_0$ was set as $N(0, 10^4)$. The level and probability of the confidence and credibility intervals, respectively, were fixed at 0.95.

## 4.1 Comparing the estimation procedures

Table 1 shows the results for the hyperparameter estimation in the LLM model. The methods are compared through the bias and mean square error

Table 1: Maximum likelihood and Bayesian estimation for the LLM

| | | | MLE | Mean | Median | Mode | Mean | Median |
|---|---|---|---|---|---|---|---|---|
| n | $q$ | $\psi$ | | Jeffreys prior | | | Uniform prior | |
| 100 | 0.02 | $\sigma_\eta^2 = 0.1$ | 0.11 | 0.15 | 0.11 | 0.06 | 0.24 | 0.19 |
| | | | (0.01) | (0.01) | (0.01) | (0.01) | (0.04) | (0.02) |
| | | $\sigma_\epsilon^2 = 5$ | 4.88 | 4.94 | 4.86 | 4.87 | 5.00 | 4.91 |
| | | | (0.57) | (0.57) | (0.57) | (0.61) | (0.59) | (0.58) |
| | 0.50 | $\sigma_\eta^2 = 2$ | 2.02 | 2.17 | 2.02 | 1.78 | 2.47 | 2.30 |
| | | | (0.60) | (0.67) | (0.60) | (0.58) | (0.93) | (0.75) |
| | | $\sigma_\epsilon^2 = 4$ | 3.94 | 4.01 | 3.93 | 3.91 | 4.05 | 3.97 |
| | | | (0.77) | (0.79) | (0.77) | (0.78) | (0.80) | (0.77) |
| | 1.00 | $\sigma_\eta^2 = 2$ | 2.03 | 2.13 | 2.02 | 1.84 | 2.35 | 2.22 |
| | | | (0.50) | (0.50) | (0.46) | (0.48) | (0.65) | (0.56) |
| | | $\sigma_\epsilon^2 = 2$ | 1.97 | 2.01 | 1.97 | 1.95 | 2.04 | 1.99 |
| | | | (0.31) | (0.30) | (0.30) | (0.48) | (0.31) | (0.30) |
| | 2.00 | $\sigma_\eta^2 = 4$ | 4.06 | 4.12 | 3.97 | 3.76 | 4.46 | 4.30 |
| | | | (1.62) | (1.37) | (1.37) | (1.57) | (1.66) | (1.55) |
| | | $\sigma_\epsilon^2 = 2$ | 1.96 | 2.07 | 2.01 | 1.95 | 2.09 | 2.03 |
| | | | (0.60) | (0.53) | (0.53) | (0.61) | (0.53) | (0.53) |
| 300 | 0.02 | $\sigma_\eta^2 = 0.1$ | 0.10 | 0.11 | 0.10 | 0.08 | 0.14 | 0.12 |
| | | | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| | | $\sigma_\epsilon^2 = 5$ | 5.00 | 5.03 | 5.00 | 5.00 | 5.04 | 5.01 |
| | | | (0.21) | (0.21) | (0.21) | (0.21) | (0.21) | (0.21) |
| | 0.50 | $\sigma_\eta^2 = 2$ | 1.99 | 2.04 | 1.99 | 1.90 | 2.13 | 2.08 |
| | | | (0.17) | (0.18) | (0.17) | (0.17) | (0.20) | (0.19) |
| | | $\sigma_\epsilon^2 = 4$ | 4.01 | 4.03 | 4.00 | 4.00 | 4.04 | 4.02 |
| | | | (0.26) | (0.27) | (0.26) | (0.26) | (0.27) | (0.27) |
| | 1.00 | $\sigma_\eta^2 = 2$ | 1.98 | 2.02 | 1.98 | 1.92 | 2.09 | 2.05 |
| | | | (0.13) | (0.13) | (0.13) | (0.14) | (0.15) | (0.14) |
| | | $\sigma_\epsilon^2 = 2$ | 2.01 | 2.02 | 2.01 | 2.00 | 2.03 | 2.01 |
| | | | (0.10) | (0.10) | (0.10) | (0.10) | (0.10) | (0.10) |
| | 2.00 | $\sigma_\eta^2 = 4$ | 3.97 | 4.03 | 3.97 | 3.87 | 4.14 | 4.07 |
| | | | (0.43) | (0.44) | (0.43) | (0.44) | (0.46) | (0.45) |
| | | $\sigma_\epsilon^2 = 2$ | 2.01 | 2.03 | 2.01 | 2.00 | 2.03 | 2.02 |
| | | | (0.19) | (0.19) | (0.18) | (0.18) | (0.19) | (0.19) |

Obs.: Numbers in brackets are the mean square error. The posterior Mode is not shown for the uniform prior as, in this case, it is the same as the MLE.

($mse$) of the estimates.

Regarding the estimation of the variance of the level component, $\sigma_\eta^2$, the MLE and BE with Jeffreys prior should be preferred to the BE with uniform prior, as they present smaller $mse$. Concerning the bias, the MLE and posterior Median with Jeffreys prior are the ones that present the best performance. The posterior Mean and Median with uniform prior seem to overestimate the real value of this hyperparameter. For the variance of the disturbances, $\sigma_\epsilon^2$, the results are very similar for all methods. Increasing the sample size, bias and $mse$ decrease, as expected, but the behaviour of the estimation procedures seems to remain the same.

To verify the convergence of the M-H algorithm, Figures 1 and 2 present the estimated autocorrelation function of the generated chains for hyperparameters $\sigma_\eta^2$ and $\sigma_\epsilon^2$, with uniform and Jeffreys priors, respectively, and $n = 100$. As it can be seen, there is a quick decay in the autocorrelation functions, with values very close to zero from lag 20. The same graphs were plotted to the other models (not shown) and results were very similar. The value of the $z$ statistic from Geweke's method was not inside the 1% significance region for generated chains, meaning that the convergence of the algorithm was attained.

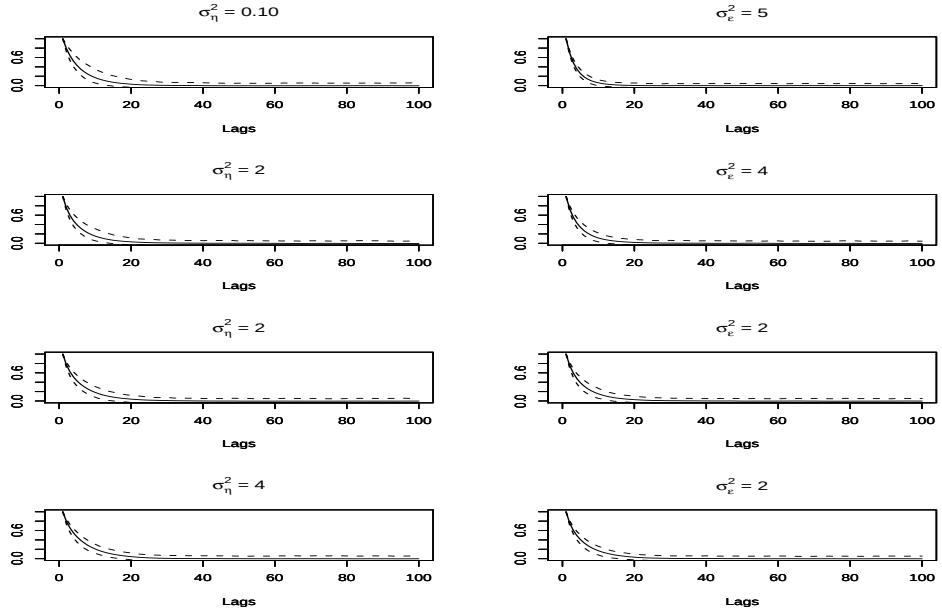


Figure 1: Estimated autocorrelation functions for the chains of hyperparameters with $n$=100 and Uniform prior, in 500 experiments Monte Carlo. The solid and dashed lines indicate the mean, the 0.05 (lower limit) and 0.95 (upper limit) percentiles, respectively.

The results for the LLT model are shown in Table 2. For the $\sigma_\eta^2$ component, the MLE and posterior Mode with Jeffreys prior are the less biased, but present larger *mse* than the other estimators. For $\sigma_\epsilon^2$ and the slope component, $\sigma_\xi^2$, there not seems to be much difference among the procedures, being all of them very close to the real value of the hyperparameter, with similar values for the *mse*.

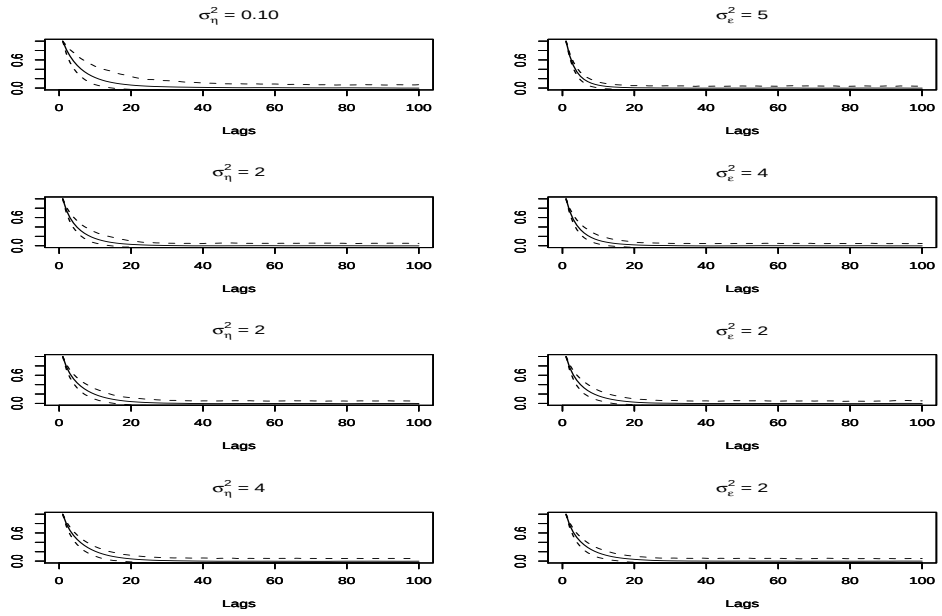


Figure 2: Estimated autocorrelation functions for the chains of hyperparameters with $n$=100 and Jeffreys prior, in 500 experiments Monte Carlo. The solid and dashed lines indicate the mean, the 0.05 (lower limit) and 0.95 (upper limit) percentiles, respectively.

Results for the SBM are also presented in Table 2. It seems that the introduction of the seasonal component in the model brings some changes in the estimation methods with respect to the conclusions drawn for the other two models. For example, in the case of the level component, $\sigma_\eta^2$, the BE with a Uniform prior presents now smaller *mse* than the other methods, although they overestimate the value of the parameter. The MLE is still the less biased. For $\sigma_\xi^2$ and $\sigma_\omega^2$, the BE with a uniform prior possess larger bias than the other procedures, although the *mse* is very similar for all of them. For the error component, $\sigma_\epsilon^2$, MLE and posterior Median with Jeffreys prior show smaller bias, but larger *mse* than the BE with a uniform prior.

## 4.2 Confidence and credibility intervals

It is well known that the credible and coverage probabilities have different meanings and interpretations. Citing Casella and Berger (2002), *"The first reflects the experimenter´s subjective beliefs, while the later reflects the un-*

Table 2: Maximum likelihood and Bayesian estimation for the LLT and SBM ($n$=200)

| | | MLE | Mean | Median | Mode | Mean | Median |
|---|---|---|---|---|---|---|---|
| Model | $\psi$ | | Jeffreys prior | | | Uniform prior | |
| LTL | $\sigma_\eta^2 = 0.5$ | 0.520 | 0.590 | 0.550 | 0.520 | 0.606 | 0.563 |
| | | (0.110) | (0.078) | (0.077) | (0.118) | (0.079) | (0.076) |
| | $\sigma_\xi^2 = 0.1$ | 0.098 | 0.100 | 0.094 | 0.084 | 0.118 | 0.110 |
| | | (0.002) | (0.001) | (0.001) | (0.002) | (0.002) | (0.035) |
| | $\sigma_\epsilon^2 = 1.0$ | 0.990 | 0.970 | 0.968 | 0.970 | 0.992 | 0.986 |
| | | (0.044) | (0.034) | (0.035) | (0.045) | (0.034) | (0.035) |
| SBM | $\sigma_\eta^2 = 0.5$ | 0.496 | 0.437 | 0.400 | 0.334 | 0.553 | 0.524 |
| | | (0.079) | (0.069) | (0.081) | (0.167) | (0.057) | (0.057) |
| | $\sigma_\xi^2 = 0.03$ | 0.030 | 0.038 | 0.034 | 0.032 | 0.041 | 0.037 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| | $\sigma_\omega^2 = 0.1$ | 0.103 | 0.109 | 0.101 | 0.084 | 0.132 | 0.122 |
| | | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| | $\sigma_\epsilon^2 = 1.0$ | 0.998 | 1.104 | 1.094 | 1.174 | 0.981 | 0.971 |
| | | (0.063) | (0.075) | (0.074) | (0.144) | (0.052) | (0.053) |

Obs.: Numbers in brackets are the mean square error. The posterior Mode is not shown for the uniform prior as, in this case, it is the same as the MLE.

*certainty in the sampling procedure"* . Even though, in this work an attempt to compare the results obtained using these two procedures will be done.

Bootstrap and asymptotic intervals are built under the classical paradigm. The bootstrap is used here based on the work of Franco et al. (2008), which shows that the bootstrap mimics very closely the behavior of the MLE, which enables this technique to be used as a tool for building confidence intervals in this case. Under the Bayesian paradigm, credibility intervals using the Jeffreys and uniform priors are calculated. The percentage of times that the built intervals contain the true value of the hyperparameters are obtained based on a large number of Monte Carlo simulations.

Table 3 presents the results for the LLM. It can be seen that the intervals present coverage rates close to the 0.95 fixed point, except the asymptotic interval for $n = 100$ and $\sigma_\eta^2$ for $n = 300$. In general, the best combination of small width and coverage rate close to 0.95 is attained by the Jeffreys credibility interval, but the bootstrap intervals show a good performance specially for $q \geq 1$ and $\sigma_\epsilon^2$ when $q = 0.50$. The asymptotic interval improves its performance when the sample size increases, being close to the other procedures, except for $q = 2.00$. The uniform credibility interval presents the larger widths, for all cases.

In Table 4 are the results for the LLT and SBM. For the LLT, it can be seen that coverage rates are always above the 95% level for the $\sigma_\eta^2$ component, for all methods. For the $\sigma_\epsilon^2$ and $\sigma_\xi^2$ components, the bootstrap intervals

Table 3: Confidence and credibility intervals for the LLM with nominal level of 95%.

| | | | | **n=100** | | | |
|---|---|---|---|---|---|---|---|
| $q$ | $\psi$ | Asymp | Cred.Unif | Cred.Jeff | Perc | BC | BCa |
| 0.02 | $\sigma_\eta^2 = 0.1$ | [-0.05; 0.27] | [0.04; 0.77] | [0.02; 0.49] | [0.01; 0.36] | [0.01; 0.35] | [0.01; 0.35] |
| | | (0.32) | (0.73) | (0.47) | (0.35) | (0.34) | (0.34) |
| | | **0.83** | **0.92** | **0.96** | **0.90** | **0.90** | **0.90** |
| | $\sigma_\epsilon^2 = 5$ | [3.42; 6.34] | [3.57; 6.88] | [3.58; 6.73] | [3.49; 6.37] | [3.59; 6.51] | [3.59; 6.51] |
| | | (2.92) | (3.31) | (3.15) | (2.88) | (2.92) | (2.92) |
| | | **0.92** | **0.96** | **0.96** | **0.91** | **0.92** | **0.92** |
| 0.50 | $\sigma_\eta^2 = 2$ | [0.56; 3.49] | [1.08; 4.81] | [0.95;4.25] | [0.80;3.96] | [0.77; 3.92] | [0.77; 3.92] |
| | | (2.93) | (3.73) | (3.30) | (3.16) | (3.15) | (3.15) |
| | | **0.89** | **0.95** | **0.96** | **0.92** | **0.93** | **0.93** |
| | $\sigma_\epsilon^2 = 4$ | [2.22; 5.66] | [2.29; 6.24] | [2.33; 6.12] | [2.16; 5.74] | [2.27; 5.88] | [2.27; 5.88] |
| | | (3.44) | (3.95) | (3.79) | (3.58) | (3.61) | (3.61) |
| | | **0.94** | **0.96** | **0.96** | **0.94** | **0.95** | **0.95** |
| 1.00 | $\sigma_\eta^2 = 2$ | [0.74; 3.32] | [1.15; 4.23] | [1.04; 3.84] | [0.92; 3.71] | [0.83; 3.56] | [0.83; 3.56] |
| | | (2.58) | (3.08) | (2.80) | (2.79) | (2.73) | (2.73) |
| | | **0.91** | **0.96** | **0.95** | **0.94** | **0.94** | **0.94** |
| | $\sigma_\epsilon^2 = 2$ | [0.88; 3.05] | [0.94; 3.39] | [0.96; 3.30] | [0.82; 3.11] | [0.95; 3.27] | [0.95; 3.28] |
| | | (2.17) | (2.45) | (2.34) | (2.29) | (2.32) | (2.33) |
| | | **0.94** | **0.96** | **0.95** | **0.95** | **0.95** | **0.96** |
| 2.00 | $\sigma_\eta^2 = 4$ | [1.74; 6.38] | [2.39; 7.42] | [2.21; 6.86] | [2.03; 6.83] | [1.79; 6.48] | [1.79; 6.48] |
| | | (4.64) | (5.03) | (4.65) | (4.80) | (4.69) | (4.69) |
| | | **0.92** | **0.96** | **0.95** | **0.94** | **0.93** | **0.93** |
| | $\sigma_\epsilon^2 = 2$ | [0.46; 3.46] | [0.67; 3.90] | [0.70; 3.80] | [0.49; 3.54] | [0.69; 3.88] | [0.69; 3.89] |
| | | (3.00) | (3.23) | (3.10) | (3.05) | (3.19) | (3.20) |
| | | **0.94** | **0.96** | **0.96** | **0.93** | **0.95** | **0.96** |
| | | | | **n=300** | | | |
| $q$ | $\psi$ | **Asymp** | **Cred.Unif** | **Cred.Jeff** | **Perc** | **BC** | **BCA** |
| 0.02 | $\sigma_\eta^2 = 0.1$ | [0.01; 0.19] | [0.05;0.29] | [0.04; 0.24] | [0.03; 0.22] | [0.03; 0.21] | [0.03; 0.21] |
| | | (0.18) | (0.24) | (0.20) | (0.19) | (0.18) | (0.18) |
| | | **0.89** | **0.95** | **0.95** | **0.92** | **0.91** | **0.91** |
| | $\sigma_\epsilon^2 = 5$ | [4.14; 5.86] | [4.20; 6.03] | [4.20; 6.01] | [4.16; 5.88] | [4.20; 5.93] | [4.20; 5.93] |
| | | (1.72) | (1.83) | (1.81) | (1.72) | (1.73) | (1.73) |
| | | **0.93** | **0.96** | **0.96** | **0.93** | **0.94** | **0.94** |
| 0.50 | $\sigma_\eta^2 = 2$ | [1.15; 2.83] | [1.35; 3.19] | [1.30; 3.05] | [1.23; 2.95] | [1.23; 2.95] | [1.23; 2.95] |
| | | (1.68) | (1.84) | (1.75) | (1.72) | (1.72) | (1.72) |
| | | **0.95** | **0.96** | **0.96** | **0.95** | **0.95** | **0.95** |
| | $\sigma_\epsilon^2 = 4$ | [3.00; 5.02] | [3.06; 5.17] | [3.06; 5.15] | [3.01; 5.05] | [3.04; 5.08] | [3.04; 5.08] |
| | | (2.02) | (2.11) | (2.09) | (2.04) | (2.04) | (2.04) |
| | | **0.95** | **0.97** | **0.96** | **0.95** | **0.95** | **0.95** |
| 1.00 | $\sigma_\eta^2 = 2$ | [1.24; 2.73] | [1.41; 2.99] | [1.36; 2.90] | [1.31; 2.85] | [1.28; 2.81] | [1.28; 2.81] |
| | | (1.49) | (1.58) | (1.54) | (1.54) | (1.53) | (1.53) |
| | | **0.96** | **0.95** | **0.95** | **0.96** | **0.97** | **0.97** |
| | $\sigma_\epsilon^2 = 2$ | [1.38; 2.64] | [1.40; 2.72] | [1.41; 2.70] | [1.36; 2.66] | [1.40; 2.71] | [1.40; 2.71] |
| | | (1.26) | (1.32) | (1.29) | (1.30) | (1.31) | (1.31) |
| | | **0.96** | **0.96** | **0.96** | **0.96** | **0.97** | **0.97** |
| 2.00 | $\sigma_\eta^2 = 4$ | [2.64; 5.31] | [2.90; 5.72] | [2.82; 5.58] | [2.77; 5.51] | [2.65; 5.35] | [2.65; 5.35] |
| | | (2.67) | (2.82) | (2.76) | (2.74) | (2.70) | (2.70) |
| | | **0.95** | **0.96** | **0.96** | **0.95** | **0.96** | **0.96** |
| | $\sigma_\epsilon^2 = 2$ | [1.14; 2.88] | [1.17; 2.98] | [1.17; 2.96] | [1.13; 2.87] | [1.21; 3.01] | [1.21; 3.01] |
| | | (1.74) | (1.81) | (1.79) | (1.74) | (1.80) | (1.80) |
| | | **0.96** | **0.96** | **0.97** | **0.95** | **0.96** | **0.96** |

Obs.: Numbers in square brackets are the limits of the intervals, in round brackets are the width and in bold are the coverage rates of the intervals.

Table 4: Confidence and credibility intervals for the LLT and SBM with nominal level of 95% ($n$=200)

| Model | $\psi$ | Asymp | Cred.Unif | Cred.Jeff | Perc | BC | BCa |
|---|---|---|---|---|---|---|---|
| LTL | $\sigma^2_\eta = 0.5$ | [-0.16; 1.20] | [0.09; 1.38] | [ 0.10; 1.33] | [0.05; 1.30] | [0.05; 1.46] | [0.05; 1.43] |
| | | (1.36) | (1.29) | (1.23) | (1.25) | (1.41) | (1.38) |
| | | **0.97** | **0.99** | **0.98** | **0.98** | **0.99** | **0.99** |
| | $\sigma^2_\xi = 0.1$ | [ 0.02; 0.18] | [0.05; 0.23] | [0.04; 0.19] | [0.04; 0.19] | [0.04;0.19] | [0.04;0.19] |
| | | (0.16) | (0.18) | (0.15) | (0.15) | (0.15) | (0.15) |
| | | **0.90** | **0.97** | **0.94** | **0.94** | **0.95** | **0.95** |
| | $\sigma^2_\epsilon = 1.0$ | [0.57; 1.40] | [0.59; 1.43] | [0.58; 1.40] | [0.55; 1.39] | [0.54; 1.38] | [0.54; 1.40] |
| | | (0.83) | (0.84) | (0.82) | (0.84) | (0.84) | (0.86) |
| | | **0.96** | **0.97** | **0.97** | **0.96** | **0.95** | **0.95** |
| SBM | $\sigma^2_\eta = 0.50$ | [-0.17;1.20] | [0.12;1.16] | [0.06;1.03] | [0.02;1.45] | [0.04;1.77] | [ 0.04; 1.75] |
| | | (1.37) | (1.04) | (0.97) | (1.43) | (1.73) | (1.71) |
| | | **0.93** | **0.97** | **0.93** | **0.96** | **0.99** | **0.99** |
| | $\sigma^2_\xi = 0.03$ | [-0.01;0.07] | [0.01;0.09] | [0.01;0.08] | [0.00;0.08] | [0.00;0.10] | [0.00;0.10] |
| | | (0.08) | (0.08) | (0.07) | (0.08) | (0.09) | (0.10) |
| | | **0.82** | **0.96** | **0.95** | **0.89** | **0.98** | **0.98** |
| | $\sigma^2_\omega = 0.10$ | [-0.01;0.21] | [0.06;0.26] | [0.05;0.22] | [0.01;0.23] | [0.02;0.26] | [ 0.02; 0.27] |
| | | (0.22) | (0.20) | (0.17) | (0.22) | (0.24) | (0.25) |
| | | **0.87** | **0.95** | **0.96** | **0.89** | **0.95** | **0.95** |
| | $\sigma^2_\epsilon = 1.00$ | [0.32;1.60] | [0.50;1.52] | [0.60;1.66] | [0.28; 1.82] | [0.24;1.81] | [0.23; 1.77] |
| | | (1.28) | (1.02) | (1.06) | (1.54) | (1.57) | (1.54) |
| | | **0.91** | **0.97** | **0.94** | **0.96** | **0.96** | **0.96** |

Obs.: Numbers in square brackets are the limits of the intervals, in round brackets are the width and in bold are the coverage rates of the intervals.

are generally better. It can be also noted that the asymptotic interval gives negative values for the lower limits of the intervals, as in the $\sigma^2_\eta$ case, which is not desirable once they are variance components. For the SBM, the credibility intervals present, in general, the best combination of smaller width for the intervals and coverage rates closer to the 0.95 fixed point. In some cases the BC and BCa intervals are close to the 0.95 point, with acceptable width, as in the $\sigma^2_\omega$ case. Once again, the asymptotic intervals present negative values for the inferior limits, except for $\sigma^2_\epsilon$.

# 5 Real data application

The Local Level Model (LLM) was fitted to an electric time series in the Northeast region of Brazil. These data were obtained from a large study concerning the quantity of energy necessary to answer the maximum demand in the peak interval (from 6:00 to 9:00 pm). The series are monthly observations of electric consumption from CHESF (São Francisco Hydroelectric Company), in the period from Jan 1983 to Feb 1997 ($n = 170$).

The data are shown in Figure 3 and the series does not seem to present seasonality nor a change in the level, so that LLM can be selected to model the data.

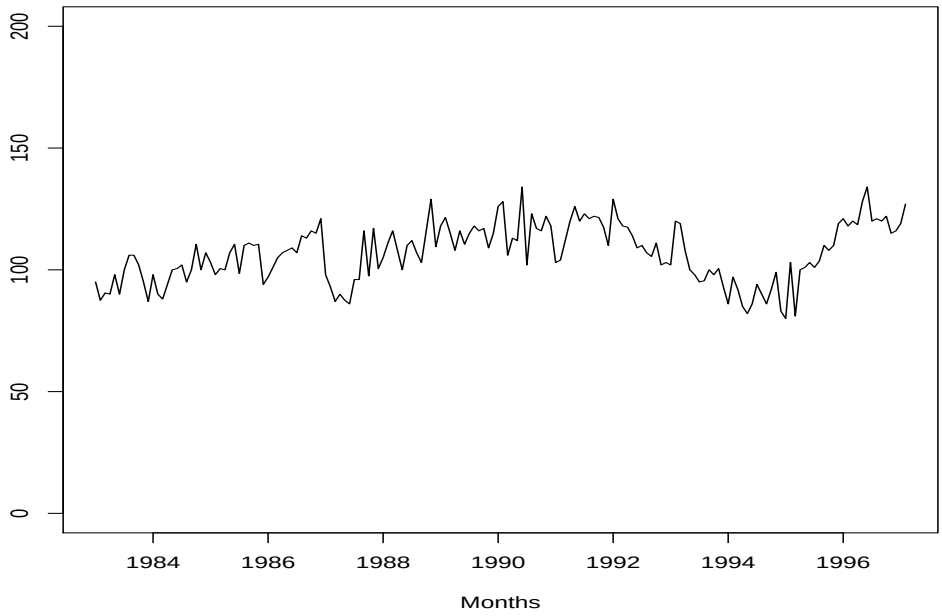


Figure 3: Electric Energy Consumption of CHESF. Values were divided by 10.

The hyperparameters of the LLM were estimated using the software Ox (Doornik (1999)). The estimates under classical and Bayesian perspectives are in Table 5. According to the simulations performed in Section 4, better results for the $\sigma_\eta^2$ component are obtained with the MLE and posterior median with Jeffreys prior, while there is no substantial difference among the methods for $\sigma_\epsilon^2$. Thus, these hyperparameters can be estimated as $\hat{\sigma}_\eta^2 \cong 12.4$ and $\hat{\sigma}_\epsilon^2 \cong 32$, providing a signal-to-noise ratio estimate equal to 0.4. A residual analysis was carried out and no evidence of correlation across time in the error term was found.

Table 6 shows the interval estimates for hyperparameters $\sigma_\eta^2$ and $\sigma_\epsilon^2$. Concerning $\sigma_\eta^2$, the simulation results showed that the Jeffreys credibility interval should be prefered, thus the estimated interval for $\sigma_\eta^2$ can be calculated as $[6.30; 22.96]$. With respect to $\sigma_\epsilon^2$, BC and BCa intervals were slightly better, so an interval for $\sigma_\epsilon^2$ can be set as $[22.17; 42.06]$. As the zero is not included in the intervals, the LLM can be an adequate model for this series.

Table 5: Maximum likelihood and Bayesian estimates for the hyperparameters of CHESF series.

| $\hat{\psi}$ | MLE | BE-Median | BE-Mode | BE-Mean | BE-Median | BE-Mean |
|---|---|---|---|---|---|---|
| | | | Jeffreys prior | | Uniform prior | |
| $\hat{\sigma}^2_\eta$ | 12.48 | 12.34 | 11.44 | 12.85 | 13.75 | 14.36 |
| $\hat{\sigma}^2_\epsilon$ | 31.85 | 31.51 | 31.83 | 32.32 | 31.69 | 32.08 |

Table 6: Confidence and credibility intervals for the hyperparameters of CHESF series with nominal level of 95%.

| Intervals | $\sigma^2_\eta$ | $\sigma^2_\epsilon$ |
|---|---|---|
| Asymp | [5.03; 19.93] | [21.76; 41.94] |
| Cred.Unif | [7.29; 24.92] | [22.23; 43.51] |
| Cred.Jeff | [6.30; 22.96] | [22.54; 44.55] |
| Perc | [6.05; 20.60] | [22.27; 42.63] |
| BC | [6.54; 21.47] | [22.17; 42.06] |
| BCa | [6.53; 21.45] | [22.17; 42.05] |

# 6 Concluding remarks

The modelling of a time series using non-observable components can be done mainly using two distinct approaches. Following the classical point of view, the method is called structural model and the maximum likelihood (MLE) procedure is used to estimate the hyperparameters. Under the Bayesian approach, the procedure is called dynamic linear model and Markov Chain Monte Carlo methods are used to approximate the posterior distribution of the hyperparameters.

In this work, the performances of these two procedures have been empirically investigated, by estimating the variances of the errors of the non-observable components in structural models. Through some Monte Carlo experiments, the bias and mean square error (*mse*) of the estimators were calculated. For the Bayesian estimators, the posterior mean, median and mode were the chosen estimators. The results showed that, in general, the maximum likelihood approach and the posterior mode are less biased and they have smaller *mse*.

Confidence (bootstrap and asymptotic) and credibility intervals for the hyperparameters were also built and compared with respect to the width and coverage percentages. The bootstrap confidence intervals used were the percentile, the bias corrected and the bias corrected and accelerated. The

credibility intervals showed, in general, a better combination of width and coverage rate. It should be stressed that the asymptotic intervals can present boundary problems, leading to negative lower limits, what should not be expected, as in this case the hyperparameters are the variances of the error terms.

The methodology was also applied to a real series of the São Francisco Hydroeletric Company (CHESF) from Brazil. Maximum likelihood and Bayesian estimators were calculated, as well as confidence and credibility intervals. The results led to the conclusion that the series follow a LLM.

Future research includes the construction of hypothesis tests to verify the significance of the components in the structural model.

**Acknowledgements**

The author G.C. Franco was partially supported by CNPq-Brazil and also by *Fundação de Amparo à Pesquisa no Estado de Minas Gerais* (FAPEMIG Foundation). The author T.R. Santos was supported by CNPq-Brazil and CAPES-Brazil. The authors also thank Thaize M. Vieira who provided some simulation presented in the paper.

# References

Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter*, University Press, Cambridge.

West, M. and Harrison, J. (1997).*Bayesian forecasting and dynamic models*, Springer, New York.

Kalman R.E. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering,Transactions ASME*, 82: 35-45.

Harvey A.C., Koopman S.J., Shephard N. (2004). *State space and unobserved component models: Theory and applications*, University of Cambridge, Cambridge.

Durbin J., Koopman S.J. (2001). *Time series analysis by state space methods*, Oxford University Press, Oxford.

Harvey A.C., Proietti T. (2005). *Readings in unobserved components models*, Oxford University Press, Cambridge.

Gamerman D., Lopes H.F. (2006). *Markov Chain Monte Carlo. Stochastic simulation for bayesian inference*, Chapman and Hall, London.

Carvalho C., Johannes M., Lopes H.F., Polson N. (2009). Particle learning and smoothing. *Technical Report*. University of Chicago Booth School of Business.

Carter C.K., Kohn R. (1994). On Gibbs sampling for state space models, *Biometrika*, 81: 541-533.

Lopes H.F., Moreira A.R.B., Schmidt A.M. (1999). Hyperparameter estimation in forecast models, *Computational Statistics and Data Analysis*, 29: 387-410.

Reis E.A., Salazar E., Gamerman D. (2006). Comparison of sampling schemes for dynamic linear models, *International Statistical Review*, 74: 203-214.

Schmidt A.M.,Gamerman D., Moreira A.R.B. (1999). An adaptive resampling scheme for cycle estimation, *Journal of Applied Statistics*, 26: 619-641.

Harvey A.C., Trimbur T., Dijk H. v. (2007). *Trends and cycles in economic time series: a Bayesian approach*, *Journal of Econometrics*, 140/2, 618-649.

Efron B., Tibshirani R. (1993). *An introduction to the bootstrap*, Chapman and Hall, London.

Franco G.C., Santos T.R., Ribeiro J.A., Cruz F.R.B. (2008). Confidence intervals for hyperparameters in structural models, *Communications in Statistics*, 33: 486-497.

Franco G.C.,Souza R.C. (2002). A comparison of methods for bootstrapping in the local level model, *Journal of Forecasting*, 21: 27-38.

Pfeffermann D., Tiller R. (2005). Bootstrap approximation to prediction MSE for state-space models with estimated parameters, *Journal of Time Series Analysis*, 26: 893-916.

Stoffer D.S., Wall K.D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter, *Journal of the American Statistical Association*, 86: 1024-1033.

Wall K.D., Stoffer D.S. (2002). A state space approach to bootstrapping conditional forecasts in ARMA models, *Journal of Time Series Analysis*, 23: 733-751.

Rodriguez, A., Ruiz E. (2009). Bootstrap prediction intervals in state-space models, *Journal of Time Series Analysis*, 30: 167-178.

Efron B.,Tibshirani R. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy, *Statistical Science*, 1: 54-77.

Proietti, T. (2000). Comparing seasonal components for structural time series models, *International Journal of Forecasting*, 16: 247-260.

Harvey, A. C., Todd, P. H. J. (1983). Forecasting economic time series with structural and Box-Jenkins models: a case study, *Journal of Business and Economic Statistics*, 1: 229-307.

Migon H.S., Gamerman D. (1999). *Statistical inference: An integrated approach*, Arnold, London.

Hastings W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57: 97-109.

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E. (1953). Equations of state calculations by fast computing machine, *Journal of Chemical and Physics*, 21: 1087-1091.

Chib S., Greenberg E. (1995). Understanding the Metropolis-Hastings algorithm, *American Statistician*, 49: 327-335.

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

Migon H.S, Gamerman D., Lopes H. F., Ferreira M. A. R. (2005). *Dynamic Models*. In: Editors: C.R. Rao, Dipak K. Dey, Elsevier B. V. (Eds.). Handbook of Statistics,25, p. 553-588.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, J.M. et al. (Eds.). *Bayesian Statistics*. New York, USA: Oxford University, p. 625-631.

Plummer M. et al. (2005). *Output analysis and diagnostics for MCMC*. The Coda Package. Software R. Avaliable at http://www.fis.iarc.fr/coda/.

Doornik J. A. (1999). *Ox: An Object-Oriented Matrix Language*. 3.ed. London: Timberlake Consultants Press.

Casella G., Berger R.L. (2002). *Statistical inference*, Duxbury, Pacific Grove.