TEXTO PARA DISCUSSÃO Nº 575

AN ADAPTIVE RESAMPLING SCHEME FOR CYCLE ESTIMATION

Alexandra Mello Schmidt[*]
Dani Gamerman[**]
Ajax R. B. Moreira[***]

Rio de Janeiro, julho de 1998

[*] Bolsista do PNPE na Diretoria de Pesquisa do IPEA.
[**] Do IM-UFRJ.
[***] Da Diretoria de Pesquisa do IPEA.

O IPEA é uma fundação pública vinculada ao Ministério do Planejamento e Orçamento, cujas finalidades são: auxiliar o ministro na elaboração e no acompanhamento da política econômica e prover atividades de pesquisa econômica aplicada nas áreas fiscal, financeira, externa e de desenvolvimento setorial.

TEXTO PARA DISCUSSÃO tem o objetivo de divulgar resultados de estudos desenvolvidos direta ou indiretamente pelo IPEA, bem como trabalhos considerados de relevância para disseminação pelo Instituto, para informar profissionais especializados e colher sugestões.

ISSN 1415-4765

SERVIÇO EDITORIAL

Rio de Janeiro – RJ
Av. Presidente Antônio Carlos, 51 – 14º andar – CEP 20020-010
Telefax: (021) 220-5533
E-mail: editrj@ipea.gov.br

Brasília – DF
SBS Q. 1 Bl. J, Ed. BNDES – 10º andar – CEP 70076-900
Telefax: (061) 315-5314
E-mail: editbsb@ipea.gov.br

SUMÁRIO

RESUMO

Os modelos lineares dinâmicos (MLD) West and Harrison (1997) constituem instrumentos úteis para a previsão de curto prazo de séries de tempo porque são flexíveis e também podem decompor a trajetória da série em fatores relevantes que têm interpretação e dinâmica característica. Tais modelos, entretanto, dependem de quantidades desconhecidas, constantes ao longo da amostra, denominadas hiperparâmetros.

Neste artigo, um (MLD) com componentes auto-regressivas é utilizado para descrever séries que têm componentes cíclicas. A distribuição marginal para os parâmetros de estado pode ser obtida ponderando as distribuições condicionais desses parâmetros pela distribuição marginal dos hiperparâmetros. Na maioria dos casos, a distribuição conjunta dos hiperparâmetros pode ser obtida analiticamente, mas não a distribuição marginal dos componentes, o que requer integração numérica. Propomos obter amostras de hiperparâmetros utilizando uma variante do método de amostragem e reamostragem por importância (SIR). Apresentamos uma aplicação com dados simulados e duas com dados reais.

ABSTRACT

Bayesian dynamic linear models (DLM) are useful in time series modelling because of the flexibility that they present in obtaining a good forecast. They are based on a decomposition of the relevant factors which explain the behavior of the series through a series of state parameters. Nevertheless the DLM as developed in West & Harrison (1997) depend on additional quantities, such as the variance of the system disturbances, which, in practice, are unknown. These are refered as hyperparameters of the model.

In this paper, DLM with auto-regressive components are used to describe time series showing cyclic behavior. The marginal posterior distribution for state parameters can be obtained by weighing the conditional distribution of state parameters by the marginal distribution of hyperparameters. In most cases the joint distribution of the hyperparameters can be obtained analytically but the marginal distributions of the components can not, thus requiring numerical integration. We propose to obtain samples of the hyperparameters by a variant of the Sampling Importance Resampling (SIR) method. A few applications are made with simulated and real datasets.

1 Introduction

Dynamic linear models (DLM) are useful in time series modelling because of their flexibility. The decomposition of the relevant factors which explain the behaviour of the series is possible, making clear how the evolution of the series is made. Nevertheless the DLM as developed in West & Harrison (1997) depend on some quantities which, in practice, are unknown.

One possibility to solve this problem is to use some optimization algorithm in order to obtain the maximum likelihood estimator (MLE) of the hyperparameter. Then the model's analysis can proceed conditioned on its value. The estimator's uncertainty can be measured locally by the Hessian at the optimal value or globally by obtaining its distribution using for example the *bootstrap* procedure. Another possibility, which became available with the advance of computational methods, is to obtain a sample of the posterior distribution of the model's parameter vector. So a better estimation becomes available because the complete marginal distribution of interest can be approximately obtained.

Here DLM with cyclical autoregressive components are used in order to describe time series showing cycle behaviour. This component is specified as in West (1995). This model specification leads to a non linear dynamic model, because the unknown elements for the cycle component, the wavelengths and the decay terms, are at the system matrix. In addition, the variances of the system equation are also unknown. So the evolution of the state parameters depends on values which are constant over the time span of the series but are unknown. These quantities are called hyperparameters and conditional on their values the inference follows analytically as described in West & Harrison (1997).

In the Bayesian approach the uncertainty about hyperparameters and state parameters is described probabilistically. Once their prior distribution is set, the posterior can be obtained, by combining it with the data. Therefore, inference about the parameters is based on summaries of the posterior distribution. In most cases the posterior distribution does not have an explicit analytical form, so the need of some form of approximation is clear. The implementation of reasonably simple techniques had made possible the use of the Bayesian approach. The main point is to obtain the posterior distribution of the quantities of interest. If they are unavailable analytically, they can be approximated by a sample from their distribution. Many simulation techniques are available and their efficiency is highly related to the dimension of the parameter vector. Lopes, Moreira & Schmidt (1996) applied the sampling importance resampling (SIR) (Rubin, 1988) method to obtain the posterior distribution of the hyperparameters of some classes of DLM. In that paper full Bayesian analyses were compared with those conditioned on particular values of the hyperparameters. As some authors (Tachibana (1995), O'Hagan (1995)) had already stressed, the SIR method does not work well when the importance and posterior distributions are far apart. In this case, one obtains a posterior distribution which could be more concentrated than it really is. The objective is to sample a reasonable number of points on the region of interest in order to obtain a good approximation of the posterior distribution. Here a variant of the SIR method is

used. It is a simple method which samples elements from an importance function which is constructed as a first approximation of the posterior distribution.

Define $\boldsymbol{\theta}^*$ a vector whose components are $\boldsymbol{\theta}$, a $Tq \times 1$ vector of state parameters, V the observational error variance and $\boldsymbol{\psi}$ a $p \times 1$ vector of hyperparameters. When $\boldsymbol{\psi}$ is known, the marginal posterior of $\boldsymbol{\theta}$ and V can be analytically obtained. Therefore, the dimension of the problem can be reduced to that of $\boldsymbol{\psi}$. So when using SIR we can obtain exact inference about $\boldsymbol{\theta}$ and V and just make approximations about the distribution of $\boldsymbol{\psi}$. We propose here an adaptive version of SIR, which we call SIR in two steps. This method will be described in detail in section 3.

This paper is organized as follows: in section 2 we define the general DLM, then the problems in specifying the hyperparameters are discussed. In this section we also present the models used in the applications and specially the autoregressive representation of cycles on DLM. The analogy between the trigonometric and the autoregressive representations is shown. In section 3 the SIR in two steps is presented and the advantages and disadvantages of using this method are discussed. In section 4 we analyse a simulated series with one cycle and we study the performance of the method in obtaining the posterior distribution of the hyperparameters involved. Then the eletroencephalogram recordings from the scalp of an individual undergoing electroconvulsive therapy given by West (1995) is reanalysed for comparative purposes. Finally, the monthly series of Brazilian industrial production index is analysed. The last section discusses the results obtained and the relevance of these techniques.

2 Bayesian Dynamic Linear Models

2.1 General Bayesian DLM and Hyperparameters

The DLM can be written as

(a) the observation equation,

$$y_t = \mathbf{F}_t'\boldsymbol{\theta}_t + v_t \qquad \text{where} \qquad v_t \sim N[0, V_t],$$

(b) the system equation,

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \qquad \text{where} \qquad \boldsymbol{\omega}_t \sim N[0, \mathbf{W}_t].$$

$\mathbf{F}_t$ is a q-dimensional vector and $\mathbf{G}_t$ is a $q \times q$ matrix, both assumed known for all t for the moment. We assume that the disturbance terms v_t and $\boldsymbol{\omega}_t$ are independent over time and mutually independent. Usually the DLM is denoted by the quadruple $M = \{\mathbf{F}, \mathbf{G}, V, \mathbf{W}\}_t$, which is time dependent, and by the parameter's prior distribution, usually denoted by:

$$(\boldsymbol{\theta}_0 \mid D_0) \sim N(\mathbf{m_0}, \mathbf{C}_0),$$

where D_0 is the set of information available at the instant $t = 0$. When the quadruple M is completely specified and we assume normal distributions on the error terms, the inference follows exactly as in West & Harrison (1997). Nevertheless, in practice, it is common to specify models where the variance of the disturbance terms are unknown. It is also possible to define models which have unknown quantities on the system equation matrix $\mathbf{G}$. In practice the value of the observational variance V_t is also unknown. However a closed-form Bayesian analysis of the DLM with unknown constant variance V is available if a particular structure is imposed in the $\mathbf{W}_t$ sequence and on the initial prior for $\boldsymbol{\theta}_0$. This structure enables a conjugate sequential updating procedure for V, in addition to that for $\boldsymbol{\theta}_t$. The evolution variance and the prior variance are specified as multiples of the observational variance for all time, namely $\mathbf{W}_t = V\mathbf{W}_t^*$ and $\mathbf{C}_0 = V\mathbf{C}_0^*$. The conjugate analysis is based on gamma distributions for $\phi = V^{-1}$ and thus inverse gamma distributions for V, for all t (West & Harrison, 1997).

Unknown quantities which are constant over the sample are called hyperparameters. Let $\boldsymbol{\psi}$ be the p-dimensional vector of hyperparameters and define the model that is dependent on $\boldsymbol{\psi}$ as $M(\boldsymbol{\psi}) = \{\mathbf{F}(\boldsymbol{\psi}), \mathbf{G}(\boldsymbol{\psi}), V(\boldsymbol{\psi}), \mathbf{W}(\boldsymbol{\psi})\}$. So, conditioned on the values of the vector $\boldsymbol{\psi}$, the DLM is completely specified.

There are different methods for estimating $\boldsymbol{\psi}$. One possibility is to obtain the maximum likelihood estimator using for example the Davidon-Flecher-Powell (Brian, 1984) algorithm which uses the gradient of the likelihood function to calculate in each iteration the Hessian. This method however presents inference for the parameters conditioned on just one value of the hyperparameter. It is also possible to use simulation methods such as the *bootstrap* (Efron, 1982).

Another possibility, which seems more reasonable, is to define a prior distribution for the hyperparameter and obtain its posterior distribution. In this way, the distribution of the state parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_T)$ and of V will be weighted by all possible values of the hyperparameters and can be obtained as

$$p(\boldsymbol{\theta}, V \mid D_T) = \int_{\boldsymbol{\psi}} p(\boldsymbol{\theta}, V \mid D_T, \boldsymbol{\psi}) p(\boldsymbol{\psi} \mid D_T) d\boldsymbol{\psi}, \tag{1}$$

where $D_T = \{D_0, \mathbf{y}\} = \{D_0, y_1, y_2, \cdots, y_T\}$ is the set of all observations. Following Bayes' theorem, the hyperparameter's posterior distribution is given by

$$p(\boldsymbol{\psi} \mid D_T) = \frac{L(\boldsymbol{\psi}; D_T) p(\boldsymbol{\psi})}{\int_{\boldsymbol{\psi}} L(\boldsymbol{\psi}; D_T) p(\boldsymbol{\psi}) d\boldsymbol{\psi}} \tag{2}$$

where $p(\boldsymbol{\psi})$ is the prior distribution of the hyperparameter and $L(\boldsymbol{\psi}; D_T) = p(\mathbf{y} \mid \boldsymbol{\psi}, D_0)$ is the predictive density of D_T conditioned on $\boldsymbol{\psi}$ and D_0, so

$$L(\boldsymbol{\psi}; D_T) = \int_{\boldsymbol{\theta}} \int_V p(\mathbf{y} \mid \boldsymbol{\theta}^*, D_0) p(\boldsymbol{\theta}, V \mid \boldsymbol{\psi}) d\boldsymbol{\theta} dV = E_{\boldsymbol{\theta}, V \mid \boldsymbol{\psi}}[p(\mathbf{y} \mid \boldsymbol{\theta}^*, D_0)] \tag{3}$$

whose analytical expression can be found in West & Harrison (1997).

Expression (1) shows that the posterior distribution of the model's parameter is obtained by considering all possible values of the hyperparameters, not just one

value, as in maximum likelihood methods. However usually even using simple analytical forms of the prior distribution, the integral in (2) does not have

an analytical solution. Because of this, in order to obtain the posterior distribution of the hyperparameter it is necessary to use some numerical approximation, which may provide a sample of the posterior distribution of interest. It is important to stress that because of the analytical tractability of the integral in (3), the only numerical marginalizations required are those associated with the hyperparameter. Calculations regarding any other quantity, such as state parameters, observational variance or future

observations can be performed analytically up to their conditioning on the values of $\boldsymbol{\psi}$.

2.2 Classes of DLM

In the following subsections we describe the models applied in this paper and the hyperparameters related to each. All the system's error variances are scaled by V and conjugate analyses follow. The results can be found in West & Harrison (1997) and West (1995).

2.2.1 First Order Polynomial DLM $\{1, 1, V, W\}$

The observation and system equations are respectively:

$$\begin{aligned} \text{observation}: Y_t &= \mu_t + v_t \quad v_t \sim N(0, V) \\ \text{level}: \mu_t &= \mu_{t-1} + \omega_t \quad \omega_t \sim N(0, VW). \end{aligned}$$

If the values of V and W are known exact inference follows just as in West & Harrison (1997). In practice, W is usually unknown and is the only hyperparameter of this model, i.e. $\psi = W$.

2.2.2 Second Order Polynomial DLM (Linear Growth)

The model is defined by the observation and system equations as follows :

$$\begin{aligned} \text{observation}: Y_t &= \mu_t + \nu_t \quad , \quad \nu_t \sim N[0, V] \\ \text{level}: \mu_t &= \mu_{t-1} + \beta_t + \omega_{t,1}\boldsymbol{\omega}_t = \begin{pmatrix} \omega_{t,1} & \omega_{t,2} \end{pmatrix}' \sim N[\mathbf{0}, V\mathbf{W}] \\ \text{growth}: \beta_t &= \beta_{t-1} + \omega_{t,2}. \end{aligned} \tag{4}$$

The system equation here is slightly different from that previously presented since μ_t depends now also on contemporaneous parameters. The level of the series is described by μ_t, and β_t represents the current rate of change in level. Again the inference does not proceed analytically if the value of $\boldsymbol{\psi} = (W_1, W_2)$ is unknown. The model conditioned on the hyperparameter can be specified by the quadruple below, restoring the usual DLM form.

$$M(W_1, W_2) = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, V, V \begin{pmatrix} W_1 + W_2 & W_2 \\ W_2 & W_2 \end{pmatrix} \right\}$$

2.2.3 Form-Free Seasonal Effects DLMs $\{\mathbf{E}_q, \mathbf{P}_q, V, V\mathbf{W}\}$

Modelling seasonal patterns in time series requires a periodic component. The seasonal effects model defines parameters which sum to zero. For a cycle of length q the model is defined by

$$\mathbf{E}_q = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad , \quad \mathbf{P}_q = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

where $\mathbf{E}_q$ is a q-dimensional vector, $\mathbf{P}_q$ is a $q \times q$ matrix. Here the hyperparameter $\boldsymbol{\psi}$ consists of the unknown elements of $\mathbf{W}$.

2.2.4 Cycle Representation in DLM

Cyclical phenomena can be defined as processes that repeat themselves with an not anticipated regular pattern. The simplest such example is the function defined by the harmonic

$$y_t = a\cos(\theta t) + b\sin(\theta t) = r\cos(\theta t + \phi)$$

where $r^2 = a^2 + b^2$ and $\cos(\theta) = a/r$. The constant r is the amplitude of the series, ϕ the phase, θ the frequency and $\lambda = 2\pi/\theta$ the wavelength.

Consider the second order autoregressive form, $y_t - a_1 y_{t-1} - a_2 y_{t-2} = 0$. If $a_1^2 + 4a_2 < 0$, $y_t = \beta_1 r^t \cos(\theta t + \beta_2)$ where β_1 and β_2 are arbitrary constants, $r = (-a_2)^{1/2}$ and θ satifies $\cos(\theta) = a_1/[2(-a_2)^{1/2}]$.

It is possible to represent cycle behaviour using second order autoregressive models. As described in West (1995), when $\xi_t = a_1\xi_{t-1} + a_2\xi_{t-2}$, the model can be put in the state-space form, where writing $\boldsymbol{\gamma}_t = (\xi_t, \xi_{t-1})'$,

$$\xi_t = (1,0)\boldsymbol{\gamma}_t \quad \text{and} \quad \boldsymbol{\gamma}_t = \mathbf{G}\boldsymbol{\gamma}_{t-1}, \text{with } \mathbf{G} = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix} \quad .$$

Given fixed initial conditions $\boldsymbol{\gamma}_1 = (\xi_1, \xi_0)'$, the implied solution $\xi_t = (1,0)\mathbf{G}^{t-1}\boldsymbol{\gamma}_1$ has functional form in t determined by the eigen-structure of the state matrix $\mathbf{G}$ (West & Harrison, 1997, chapter 5). In the case of a pair of complex conjugate eigen-structure we have exactly the solution of a second order autoregressive model.

The stochastic version of the difference equation, $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \omega_t$, where the ω_t are a sequence of zero mean, independent and usually normally distributed innovations is a standard AR(2) model with state space representation (see West (1995) for details). It implies that a time series y_t, exhibiting stochastic but persistent cycles of wavelength λ and decay r may be modelled as

$$\begin{aligned} y_t &= \xi_t + v_t \quad v_t \sim N(0, V) \\ \xi_t &= a_1\xi_{t-1} + a_2\xi_{t-2} + \epsilon_t \quad \epsilon_t \sim N(0, w), \end{aligned} \tag{5}$$

The special case $\epsilon_t = 0$ for all t implies that the y_t are noisy observations of a cosine wave of constant amplitude and phase; the other extreme special case of $v_t = 0$ for all t implies observations on the pure cyclical AR(2) processes, $y_t = \xi_t$, with time varying amplitude and phase.

The quadruple which defines this model is defined by $M(\boldsymbol{\psi}) = \{\mathbf{F}, \mathbf{G}(a_1, a_2), V, V\mathbf{W}(\omega)\}$, where

$$\mathbf{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{G}(\boldsymbol{\psi}) = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{W}(\boldsymbol{\psi}) = \begin{pmatrix} w & 0 \\ 0 & 0 \end{pmatrix}, \tag{6}$$

with $\boldsymbol{\psi} = (a_1, a_2, w)$, $w = V(\epsilon_t)$, $r = (-a_2)^{1/2}$ and $\lambda = 2\pi / cos^{-1}(a_1/2r)$.

Observe that this model has (fixed) unknown components in the system matrix equation $\mathbf{G}$. Because of this, the model is non linear, so its inference procedures does not follow directly. They can be performed conditioned on the values of $\boldsymbol{\psi}$ or by obtaining their posterior distribution. Another approximation which can be done is that proposed by Migon & Harrison (1984) where a basic linearisation is suggested for a_1 and a_2. Beyond that, in practice, the parameter of the disturbance variance term is also unknown.

The applications of this paper use models which are superpositions of the model components described in the sections above. In studying series which have more than one cycle then as many multiple cycle components can be modelled by superposition of one cycle models.

3 Methodology

Practical implementation of Bayesian methods usually requires substantial computation. This computational requirement is essentially to calculate summaries of the posterior distribution.

Given data information D_T obtained under a parametric model indexed by the parameter $\boldsymbol{\psi}$, the Bayesian learning process is based on (2). One immediate computational problem is to solve the integral in the denominator. Some approximating methods available to obtain the posterior distribution of the variable of interest are Gaussian quadrature (Naylor & Smith, 1982), Gibbs sampling (Gelfand & Smith, 1990) and Metropolis-Hastings (Hastings, 1970) algorithms. The performance of the methods are strongly influenced by the dimension of the problem. Our aim here is to use an adaptive version of the SIR method proposed by Rubin (1988). Apart from being an easy method to implement, it is used because $\boldsymbol{\theta}$ and V are analytically treated. We only require approximations for the posterior distribution of $\boldsymbol{\psi}$, so that the problem's dimension is reduced from $(Tq + p + 1)$ to p.

3.1 Sampling Importance Resampling

Rubin (1988) proposed a method based on an importance function to obtain an approximated sample of the posterior distribution. Suppose that $p(\boldsymbol{\psi})$ is the specified prior and $L(\boldsymbol{\psi}; \mathbf{y})$ is the likelihood and suppose that $h(\boldsymbol{\psi})$ is a probability density

function which ideally should be a close approximation of $p(\boldsymbol{\psi}|\mathbf{y})$. The SIR method is as follows :

(i) generate $\boldsymbol{\psi}_1$, $\boldsymbol{\psi}_2$, $\cdots$, $\boldsymbol{\psi}_n$ from $h(\boldsymbol{\psi})$;

(ii) resample $\boldsymbol{\psi}_i^*$, $i = 1, \cdots, m$ with probability

$$\omega_i = \frac{f_{\mathbf{y}}(\boldsymbol{\psi}_i)}{\sum_{i=1}^n f_{\mathbf{y}}(\boldsymbol{\psi}_i)} \quad , i = 1, ..., n$$

where

$$f_{\mathbf{y}}(\boldsymbol{\psi}_i) = \frac{L(\boldsymbol{\psi}_i; \mathbf{y}) p(\boldsymbol{\psi}_i)}{h(\boldsymbol{\psi}_i)}.$$

It can be shown that $p(\boldsymbol{\psi}_i^* < a) \overset{m,n\to\infty}{\longrightarrow} \int_{-\infty}^a p(\boldsymbol{\psi}|\mathbf{y}) d\boldsymbol{\psi}$. In other words, $\boldsymbol{\psi}_i^*$ is approximately a sample from the posterior $p(\boldsymbol{\psi}|\mathbf{y})$. Smith & Gelfand (1992) suggest taking $h(\boldsymbol{\psi}_i)$ as the prior distribution. In this particular case the weights ω_i are simply given by

$$\omega_i = \frac{L(\boldsymbol{\psi}_i; \mathbf{y})}{\sum_{i=1}^m L(\boldsymbol{\psi}_i; \mathbf{y})} \quad , i = 1, ..., n. \tag{7}$$

Applying this procedure to the DLM problem after observing the entire time series, i.e. $D_T = \{D_0, y_1, \cdots, y_T\}$, we will obtain an approximation for the hyperparameter posterior distribution. So, for example, the posterior mean of the parameter vector,

$$\begin{aligned} E(\boldsymbol{\theta}|D_T) &= \int_{\boldsymbol{\theta}} \boldsymbol{\theta} p(\boldsymbol{\theta}|D_T) d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\psi}} \left\{ \int_{\boldsymbol{\theta}} \boldsymbol{\theta} p(\boldsymbol{\theta}|D_T, \boldsymbol{\psi}) d\boldsymbol{\theta} \right\} p(\boldsymbol{\psi}|D_T) d\boldsymbol{\psi} \\ &= \int_{\boldsymbol{\psi}} E(\boldsymbol{\theta}|\boldsymbol{\psi}, D_T) p(\boldsymbol{\psi}|D_T) d\boldsymbol{\psi}, \end{aligned} \tag{8}$$

which, using SIR, can be approximated by

$$\sum_{i=1}^m E(\boldsymbol{\theta}|\boldsymbol{\psi}_i, D_T) q_i^*. \tag{9}$$

If instead, one is interested on the marginal posterior of V, $p(V|D_T)$, this can similarly be approximated by

$$\sum_{i=1}^n p(V|\boldsymbol{\psi}_i, D_T) q_i^*.$$

The term q_i^* is just the proportion of times that the hyperparameter value $\boldsymbol{\psi}_i$ was resampled and should reproduce characteristics of the posterior distribution $p(\boldsymbol{\psi}|D_T)$.

3.2 Sampling Importance Resampling in Two Steps

The main problem with the SIR method is the appropriate choice of the importance function. When using the SIR method as suggested by Smith & Gelfand (1992), i.e. with the prior distribution as the importance function, some problems can occur. Tachibana (1995) points out that if the region where the likelihood has significant values is small compared to the region of the importance function, many points will receive very small weights ω_i and will not be resampled. So, observations of the importance function which are too far from the relevant region of the likelihood will be discarded and those which are nearest will be heavily resampled, generating a sample with less variability than the posterior. In this case, the weights will be concentrated on a few observations of the prior and the result is a poor approximation for the posterior distribution. Lopes, Moreira & Schmidt (1996) found these problems in obtaining the hyperparameter's posterior distributions in some applications. In some cases it was concentrated on very few values.

In order to obtain a relevant approximation of the posterior distribution a variant of the SIR method is used which will be called SIR in two steps. Given a prior distribution and a likelihood function,

1. use SIR method as proposed by Smith & Gelfand (1992): first obtain a sample from the prior distribution and then form the weights as in (7);

2. define the parameters of an importance density function of a given form according to the approximation of the posterior given above. Here these parameters are calculated according to the maximum and minimum values of each of the coordinates of the sample obtained in the first approximation of the posterior distribution. The importance density is taken in a product form;

3. obtain a sample of the importance function with parameters defined in step 2 and resample with probability given by (**??**). This last sample is considered an approximate sample of the posterior distribution.

It is expected that the sample obtained in this way will provide a better approximation of the posterior distribution. The first step is made just to define a rough approximation of the relevant region of the posterior distribution. This is a way to overcome problems due to discrepancies between the posterior and the importance densities. At the second step the sample will hopefully be obtained in a region of more relevance of the posterior distribution.

A natural option for defining the parameters of the importance density function of the second step is to use the sample mean and variance of the first approximation. This approach is suggested by Evans & Swartz (1995). We used this approach in some examples and it resulted in an underdispersed importance density relative to the posterior distribution. As a result,

the importance weights were heavily distorted with low posterior density points having a very large resampling weight.

When using resampling techniques, it is important that the parameter region with relevant posterior probability is adequately covered by the importance density

function. Therefore, it is better to have an overdispersed importance density. This can easily be achieved by taking more robust sample statistics to define the parameters of the importance density. For the applications in this paper, the extreme order statistics from the sample were used. In particular, when univariate normal importance density functions were considered, their mean μ and variance σ^2 were set as

$$\mu = \frac{\max + \min}{2} \quad \text{and} \quad \sigma = \frac{\max - \min}{2}, \tag{10}$$

where min and max are the minimum and maximum values of the sample from the first step as described above. Observe that μ is just the middle point of the sample range. The standard deviation can be specified according to the probability mass that one wants to allocate to the tails of the distribution. In our case we have given around 30% probability to hyperparameter values beyond those included in the first step sample range. In other words we have considered that $2\sigma = \max - \min$, meaning that we have given chances of about 15% of sampling values that are over the maximum value and under the minimum value from the first approximation.

4 Applications

The aim of this paper is to apply the methodology described in the previous section to models obtained by superposition of the components described in section 2. The first dataset analysed was simulated with one wavelength and no decay. The second dataset analysed was the series of eletroencephalogram (EEG) recordings analysed by West (1995), using Gibbs sampling. Here a comparison between his approach and the one using SIR in two steps is made. Then we analysed the Brazilian General Industrial Production index.

In general for all the models analysed below the prior distribution for the hyperparameters of the cycles components i.e. a_1 and a_2 of equation (5) had been chosen according to the region of stability conditions for the homogenous solution. As can be seen in Enders (1995) this region is given by $a_1 \in [-2, 2]$ and $a_2 \in [-1, 0]$. The prior distribution for these hyperparameters were chosen as uniforms over the region of stability.

4.1 Simulated Data

A time series was simulated according to (5) with wavelength $\lambda = 9.76\,(a_1 = 1.6)$, no decay ($a_2 = -1$) and variance of the system disturbance $\omega = 1$. The model used to analyse this dataset assumed the first term of the autoregressive term (a_1) and the variance of the system disturbances (ω) as unknown quantities, and so $\boldsymbol{\psi} = (a_1, \omega)$. The value of a_2 was taken as known. The prior distribution for ω was taken as $Gamma(1, 1)$ and a_1 and ω were assumed to be prior independent.

When using SIR in two steps the importance function for a_1 was chosen as normal distribution. For ω, the importance function was taken as the prior distribution over both steps.

At first, this data was analysed using SIR, as proposed in section 3.1 with the prior distributions as the importance function, to compare the results with those of SIR in two steps and to verify the gain of using the second approach. The first step used a sample of size 2000 and the resample had size 500. For the second approach we first obtained a sample of size 500 using the prior distributions as the importance function to calculate the parameters for the importance function of the second step. The parameters obtained following the methodology of the previous section for the normal distribution which defines the importance function were $\mu = 1.595$ and $\sigma = 0.012$. At the end the resample size was also equal to 500. The summaries of the posterior distribution of a_1 and ω can be seen in the Table 1 below.

Table 1: Quantiles of the posterior distribution obtained using SIR in two steps

Hyperparameter	Quantiles		
	25%	50%	75%
a_1	1.5978	1.6026	1.6062
ω	1.0118	1.2285	1.3902

From table 1 it can be seen that the median of a_1 is 1.6026, which is very close to its real value 1.6. It also can be noticed that the interquantile interval is really small. For ω the interquantile interval does not contain the true value of the hyperparameter, but its range is bigger than that of a_1 probably because the region where a_1 was sampled was more concentrated than that for ω. This can also be seen from Figure 1 with the histograms of the hyperparameter's posterior distributions. The gain in doing SIR in two steps can be clearly seen.

The comparison between the two approaches SIR and SIR in two steps can also be observed in Tables 2 and 3 below.

Table 2: Summaries of the posterior distribution obtained using SIR

Group	LLik	Accumulated Probability	Number of Points at the Group
0	-349.506	0.43067	1
1	-349.506	0.43067	1
2	-349.989	0.69625	2
3	-350.223	0.90649	3

Group 0 has the hyperparameter pair which produces the posterior mode. Groups 1, 2 and 3 contain the points that add up to 25%, 50% and 75% of probability mass, respectively without those from previous group. The probabilities referred to here are given by the weights w_i given in (**??**). LLik stands for the logarithm of the maximum value of the predictive likelihood $L(\boldsymbol{\psi}; D_T)$ at each group. The fourth column shows the number of points sampled in each group.

Figure 1. Histograms for the hyperparameters a_1 and w for the simulated data: top row - standard SIR; bottom row - SIR in 2 steps.

Table 3: Summaries of the posterior distribution obtained using SIR in two steps

Group	LLik	Accumulated Probability	Number of Points at the Group
0	-348.535	0.00953	1
1	-348.793	0.25242	30
2	-349.334	0.50216	76
3	-350.124	0.75095	164

Comparing tables 2 and 3 we can see that with the same sample size the first approximation (table 2) has far less points in each group than those of the second one, probably because in the second approach we sampled from a more concentrated region. Observe that at group 3 of both procedures we have almost the same LLik but on Table 3 there are many more points in this group, indicating that the approximation is better than that from Table 2. Also interesting from Table 3, there are in all 271 points accounting for a posterior mass probability of 75% and they are all within 2 points from the maximum likelihood in terms of logarithm of likelihood, showing again how the points were sampled around the region of maximum likelihood.

4.2 Eletroencephalogram Data

Here the analysis of $n = 110$ electroencephalogram (EEG) recordings from the scalp of an individual undergoing electroconvulsive therapy is made. This dataset was introduced and analysed in West (1995) and same models are used here. Initially the data is fitted by the model which is the superposition of model (4) and one which has two wavelengths, which is itself the superposition of two model forms (5).

The approximation used in West (1995) to obtain the parameter's posterior distribution was based on the Gibbs sampling. Here, we use SIR in two steps to obtain an approximation of hyperparameter's posterior distribution only. Following our previous discussion, there is no need to sample the state parameters and the observational variance.

Initially we assumed conditionally uniform prior distributions for the wavelength coefficients a_1 induced by constraints on wavelengths of $2 \leq \lambda_j \leq 25$, and a minimum separation of one wavelength, $\lambda_2 > \lambda_1 + 1$, exactly as in West (1995). Then, on the second step of the adaptive SIR method, we used truncated normal distributions with parameters according to (10) but also with the separation constraint of one wavelength. The priors for the system's standards deviations were set as uniform distributions on the interval $(0, 1)$. In our models, all the disturbances' variances are scaled by V and therefore the values of cyclic component variances W_{c1} and W_{c2} are generally smaller than 1. Larger values can easily be accommodated if deemed necessary.

The importance density functions on the second step for the W_j's were adopted as normal distributions truncated to their relevant interval, the $(0, 1)$ region. Their parameters were obtained again according to (10). In the first step of the adaptive

SIR procedure we took a sample size of 1000, on the second step the sample size was of 5000 and the resampling size was of 1000. The results of this analysis can be seen in table 4 below.

Figure 2 shows the histograms of the approximation of the hyperparameter's posterior distribution obtained using SIR in two steps which can be compared with Figures 9 to 13 of West (1995). It must be stressed that the variances terms of our models are not comparable with those from West (1995) because here our variances are scaled by V, the observational variance. It is clear that using the same prior distribution we have obtained different wavelengths. In this analysis the first cycle has wavelength's mode around 4.3, while in West's analysis the mode of the wavelength of the first cycle is around 6.3. This is similar to our mode of the second cycle's wavelength whereas West's second cycle wavelength has mode around 10. We evaluated the posterior density at the modal points obtained with the SIR and Gibbs sampling methodologies. It was observed that the mode obtained with the former is more than 2 units in the log scale larger than the latter.

Table 4: Quantiles of the posterior distribution obtained using SIR in two steps - EEG with 2 wavelentghs

Hyperparameter	Quantiles		
	25%	50%	75%
λ_1	4.1908	4.2940	4.3886
λ_2	6.2768	6.4533	6.5303
W_l	0.7089	0.8658	0.9011
W_{c1}	0.1844	0.2514	0.3470
W_{c2}	0.3293	0.5973	0.7943

where λ_i is the i-th wavelength and W_{ci} is the variance of the i-th wavelength parameter.

Figure 2. Histogram of the hyperparameters associated with the two cycle components for the EEG data with unrestricted prior on wavelengths.

We proceeded to analyse the same exercise but now using a prior distribution for the wavelengths constrained to the interval $[6, 25]$. The results are shown in figure 3. As can be observed, constraining the prior distribution to the interval $[6, 25]$ we have obtained similar results as those from West (1995). On figure 4a and 4b are the estimated mean components for both models, the first with wavelenghts constrained to interval $[2, 25]$ and the second to $[6, 25]$.

Following West (1995) we proceed with the analysis of $k = 3$ wavelengths. The separation constraint of one wavelength was used again, i.e. $\lambda_3 > \lambda_2 + 1 > \lambda_1 + 1$. The results can be seen in the table below. Again we obtained a different result from West (1995) when the wavelengths components were constrained to $[2, 25]$.

Figure 3. Histogram of the hyperparameters associated with the two cycle components for the EEG data with a restricted prior on wavelengths.

Figure 4. Mean estimated components of model with two cycle components for the EEG data with: first column - unrestricted prior; second column - restricted prior. The dataset appear as dots to be compared with the mean response estimates.

Table 5: Quantiles of the posterior distribution obtained using SIR in two steps - EEG with 3 wavelengths

Hyperparameter	Quantiles		
	25%	50%	75%
λ_1	3.9783	4.0615	4.0615
λ_2	6.3714	6.3714	6.3983
λ_3	10.039	10.290	10.290
W_l	0.6645	0.6645	0.9104
W_{c1}	0.3133	0.3133	0.6334
W_{c2}	0.1949	0.2400	0.5399
W_{c3}	0.5558	0.5558	0.5831

Figure 5. Histogram of the hyperparameters associated with the three cycle components for the EEG data with unrestricted prior on wavelengths.

Figure 6. Histogram of the hyperparameters associated with the three cycle components for the EEG data with a restricted prior on wavelengths.

Figures 5 and 6 shows the hyperparameter's posterior distributions of the SIR in two steps analysis which can be compared to the results reported by West (1995).

The analysis was repeated but now the prior distribution was constrained to the region $[6, 25]$ and again we obtained similar results as those from West (1995). Here it is clear from the histograms that the approximation obtained is not very good showing that the method might not work very well with a large number of hyperparameters or larger number of steps are required.

4.3 General Industrial

In this section the analysis of the Brazilian general industrial production index is made. This is a Laspeyres index with fixed basis measuring the output of the production of the industrial sector in Brazil. It is calculated by the Official Statistics office of Brazil (IBGE) based on a sample of 944 products. It roughly covers two thirds of the output of the industrial sector and is the single index that most reliably measures the economic activity in this sector of the economy. The data set used in this section covers the period from January of 1981 to February of 1997. The model used is

$$\begin{aligned} y_t &= \mu_t + \phi_{t,0} + \xi_t + \alpha\ days + v_t \quad v_t \sim N(0, V_t) \\ \mu_t &= \mu_{t-1} + \beta + \omega_{t,l} \quad \omega_{t,l} \sim N(0, W_l) \\ \xi_t &= a_1\xi_{t-1} + a_2\xi_{t-2} + \omega_{t,c} \quad \omega_{t,c} \sim N(0, W_c). \end{aligned}$$

It is obtained by the superposition of the components described in sections 2.2.2 to 2.2.4. The regressor variable, *days*, is the number of working days per month and its coefficient α is static. Observe that the parameters $\{\beta, a_1, a_2\}$ are also static.

Figure 7. Brazilian general industrial production index: data and fitted model

So the hyperparameters of this particular model are $\boldsymbol{\psi} = (a_1, a_2, W_l, W_c)$. As outlined above the prior for the hyperparameters with respect to the cycle representation were stabilished as $a_1 \sim U(1, 2)$ and $a_2 \sim U(-1, 0)$, the region where the solution is stable. The prior for a_1 has the restriction of being at the interval $(1, 2)$ because a cycle with wavelength smaller than 6 months has no economic meaning. Normal importance functions for both these hyperparameters were stabilished with mean and standard deviation obtained according to (10).

Table 6: Summaries of the posterior distribution obtained using SIR in two steps - General Industrial Production in Brazil

Group	LLik	Accumulated Probability	Number of Points at the Group
0	-609.737	0.06681	1
1	-610.958	0.29033	16
2	-611.406	0.52847	31
3	-612.278	0.79407	79

Table 7: Quantiles of the posterior distribution obtained using SIR in two steps

Hyperparameter	Quantiles		
	25%	50%	75%
a_1	1.2290	1.3387	1.3568
a_2	-0.62654	-0.5501	-0.4810
W_l	1.0468	1.1067	1.2287
W_c	1.0074	1.039	1.1370

For W_l and W_c, the hyperparameters of the disturbances variances, the prior was set as a product of two log-normal distributions. These were chosen in such a way that the associated normal distribution has probability 90% in the interval $(-2, 2)$. The importance functions were also log-normals and the mean and the variance were calculated for the associated normal which has mean and standard deviation stabilished as in (10). For the first step of SIR a sample of size 1000 was obtained, then the parameters of the importance functions of the second step were obtained and another sample of size 5000 was used to finally provide an approximation of the posterior distributions of interest. Tables 6 and 7 and figure 8 summarize some results for the hyperparameter's posterior distributions. In addition, figure 7 presents the fitted model, obtained by the posterior means of the mean responses at each time.

The complicated nature of these likelihoods is apparent when results are compared with classical estimation. Standard software for MLE of hyperparameters[1] led to a value with likelihood -612.424, clearly a local mode. Another interesting result can be analysed when observing table 7. The intervals for a_1 and a_2 are very wide. They include values that result in different interpretations for the wavelength and the decay components suggesting a possible identification problem.

Figure 8. Histogram of the hyperparameters of the Brazilian industrial production index data.

5 Conclusion

The exact analysis of time series using DLM is only possible when all the components of the model are known. This rarely occurs in practice. Some classes of models were discussed and the representation of cycle behaviour using an autoregressive model of second order was chosen.

In order to obtain the hyperparameter's posterior distribution we used an adaptive method of the SIR algorithm proposed by Rubin (1988). We called this method SIR in two steps, because we first obtained a sample of the posterior distribution using the hyperparameter's prior distribution as the importance function. Then, based on this sample, we calculated the parameters of an improved importance function of the second step. Finally using this function we obtained the final sample of the posterior distribution.

As shown in section 4.1, when we have a likelihood that is very concentrated, SIR in two steps approximation works better than that SIR using the prior as the importance function, as suggested by Smith & Gelfand (1992). In section 4.2 we reproduced the analysis made by West (1995) using Gibbs sampling and found different wavelengths using the same prior distribution. When we constrained the prior distribution we had nearly the same results as those from West (1995). The limitation of the SIR in two steps method, appear when working in highly dimensional space of hyperparameters. Possibly further steps may be beneficial in these cases.

It should be stressed that there are other numerical methods which could be used in order to obtain the hyperparameter's posterior distribution, such as MCMC methods. The idea here was just to verify that for low dimensions of the hyperparameter we can obtain a good approximation of its posterior distribution using SIR in two steps. The method is simple to implement. For low dimensional $\boldsymbol{\psi}$ it was shown that the resulting sample may provide a reasonable approximation of hyperparameter's posterior distribution.

References

Brian, B. (1984) Basic Optimization Methods. Technical Report, University of Bradford, School of Mathematical Science.

[1]The software used for MLE was STAMP (Koopman et al., 1995)

Efron, B. (1982) *The Bootstrap, Jacknife and Other Resampling Plans*, Society of Industrial and Applied Mathematics, Philadelfia.

Enders, W. (1995) *Applied Econometric Time Series*, First Edition, John Wiley & Sons, New York.

Gelfand, A. E., and Smith, A. F. M. (1990) 'Sampling based approaches to calculating marginal densities'. *J. Amer. Statist. Ass.*, **85**, 398-405.

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

Lopes, H., Moreira, A. R. B. and Schmidt, A. M. (1996) Hyperparameter estimation in forecasting models. Under revision for CS&DA.

Koopman, S. J., Harvey, A. C., Doornik, J. A. and Shephard, N. (1995). *STAMP 5.0: Structural Time Series Analyser, Modeller and Predictor*, Chapman & Hall, London.

Migon, H. S. and Harrison, P. J. (1984) An application of non-linear Bayesian forecasting to television advertising. In Bayesian Statistics 2 (eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A.F.M. Smith), North-Holland, Amsterdam.

Naylor, J.C. and Smith, A.F.M. (1982) Application of a method for the efficient computation of posterior distributions. *Appl. Statist.*, **31**, 214-225.

O'Hagan, A. (1994) *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference*, Edward Arnold, London.

Rubin, D. B. (1988) Using the SIR Algorithm to Simulate Posterior Distributions. In Bayesian Statistics 3 (eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A.F.M. Smith), pp. 395-402, Oxford University Press.

Smith, A. F. M. and Gelfand, A. E. (1992) Bayesian statistics without tears: a sampling-resampling perspective, *Amer. Statist.*, **46**, 84-88.

Tachibana, V. (1995) Approximate methods in Bayesian models of randomized response and logistic regression. Unpublished PhD thesis, COPPE-UFRJ, 1995 (in Portuguese).

West, M. and Harrison, P. J. (1997) *Bayesian Forecasting and Dynamic Models*, Second Edition, Springer, New York.

West, M. (1995) Inference on Cycles in Time Series, *J. Amer. Statist. Ass.*, **90**, 1301-1312.

Captions for the figures

Figure 1. Histograms for the hyperparameters a_1 and w for the simulated data: top row - standard SIR; bottom row - SIR in 2 steps.

Figure 2. Histogram of the hyperparameters associated with the two cycle components for the EEG data with unrestricted prior on wavelengths.

Figure 3. Histogram of the hyperparameters associated with the two cycle components for the EEG data with a restricted prior on wavelengths.

Figure 4. Mean estimated components of model with two cycle components for the EEG data with: first column - unrestricted prior; second column - restricted prior. The dataset appear as dots to be compared with the mean response estimates.

Figure 5. Histogram of the hyperparameters associated with the three cycle components for the EEG data with unrestricted prior on wavelengths.

Figure 6. Histogram of the hyperparameters associated with the three cycle components for the EEG data with a restricted prior on wavelengths.

Figure 7. Brazilian general industrial production index: data and fitted model

Figure 8. Histogram of the hyperparameters of the Brazilian industrial production index data.

Figure 1

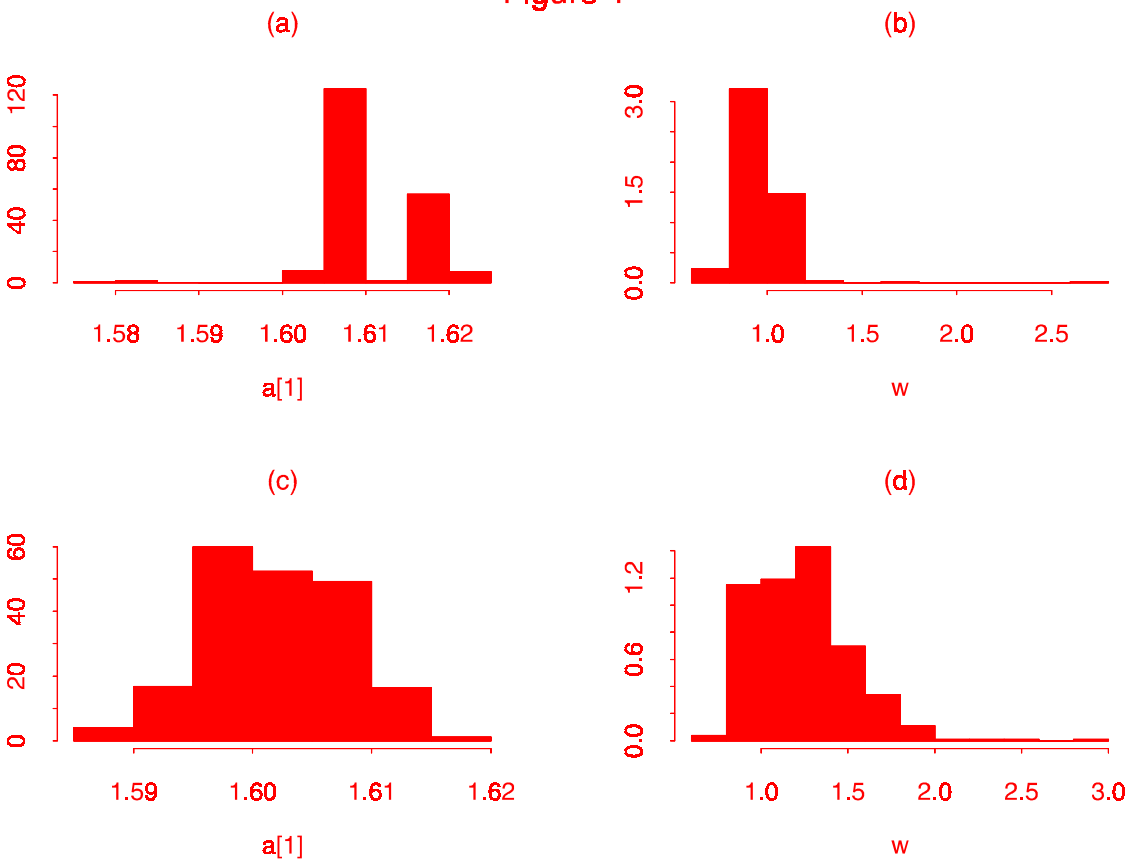

(a)
0
40
80
120
1.58
1.59
1.60
1.61
1.62
a[1]
(b)
0.0
1.5
3.0
1.0
1.5
2.0
2.5
w
(c)
0
20
40
60
1.59
1.60
1.61
1.62
a[1]
(d)
0.0
0.6
1.2
1.0
1.5
2.0
2.5
3.0
w

Figure 2

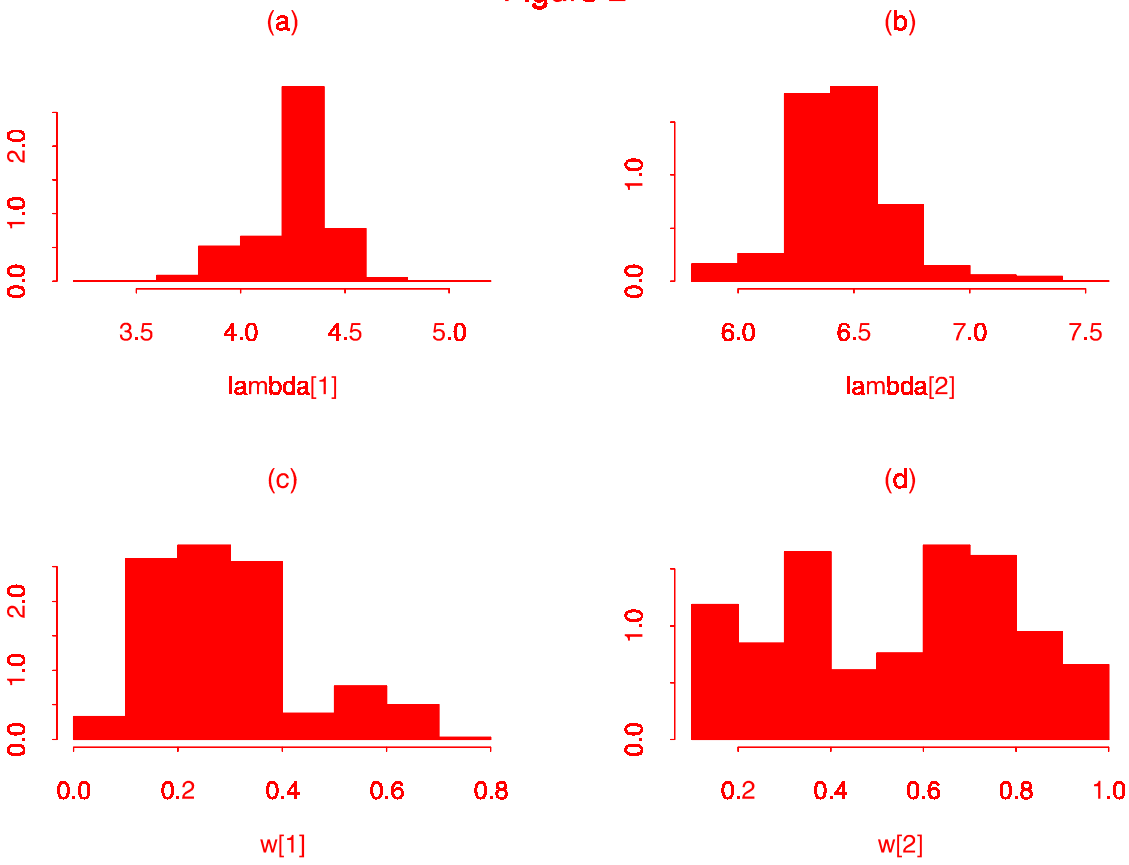

(a)
0.0
1.0
2.0
3.5
4.0
4.5
5.0
lambda[1]
(b)
0.0
1.0
6.0
6.5
7.0
7.5
lambda[2]
(c)
0.0
1.0
2.0
0.0
0.2
0.4
0.6
0.8
w[1]
(d)
0.0
1.0
0.2
0.4
0.6
0.8
1.0
w[2]

Figure 3

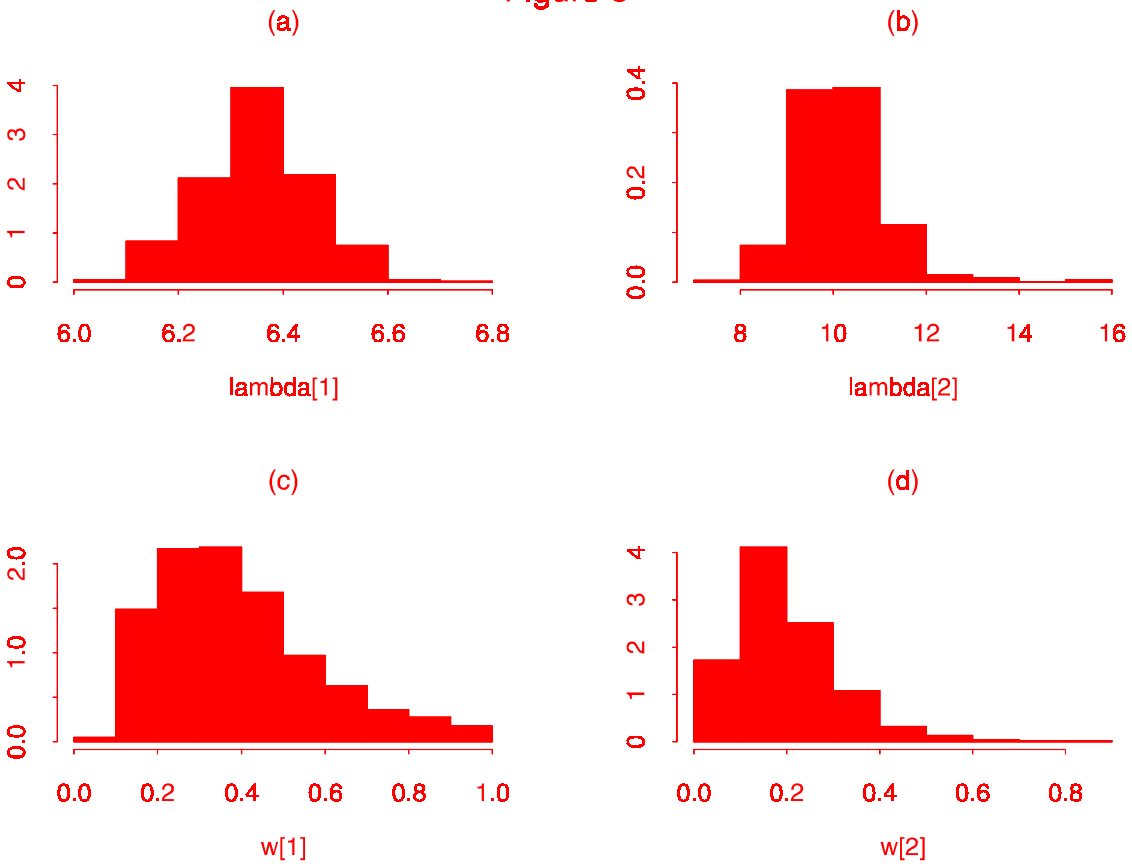

(a)
0
1
2
3
4
6.0
6.2
6.4
6.6
6.8
lambda[1]
(b)
0.0
0.2
0.4
8
10
12
14
16
lambda[2]
(c)
0.0
1.0
2.0
0.0
0.2
0.4
0.6
0.8
1.0
w[1]
(d)
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
w[2]

Figure 4

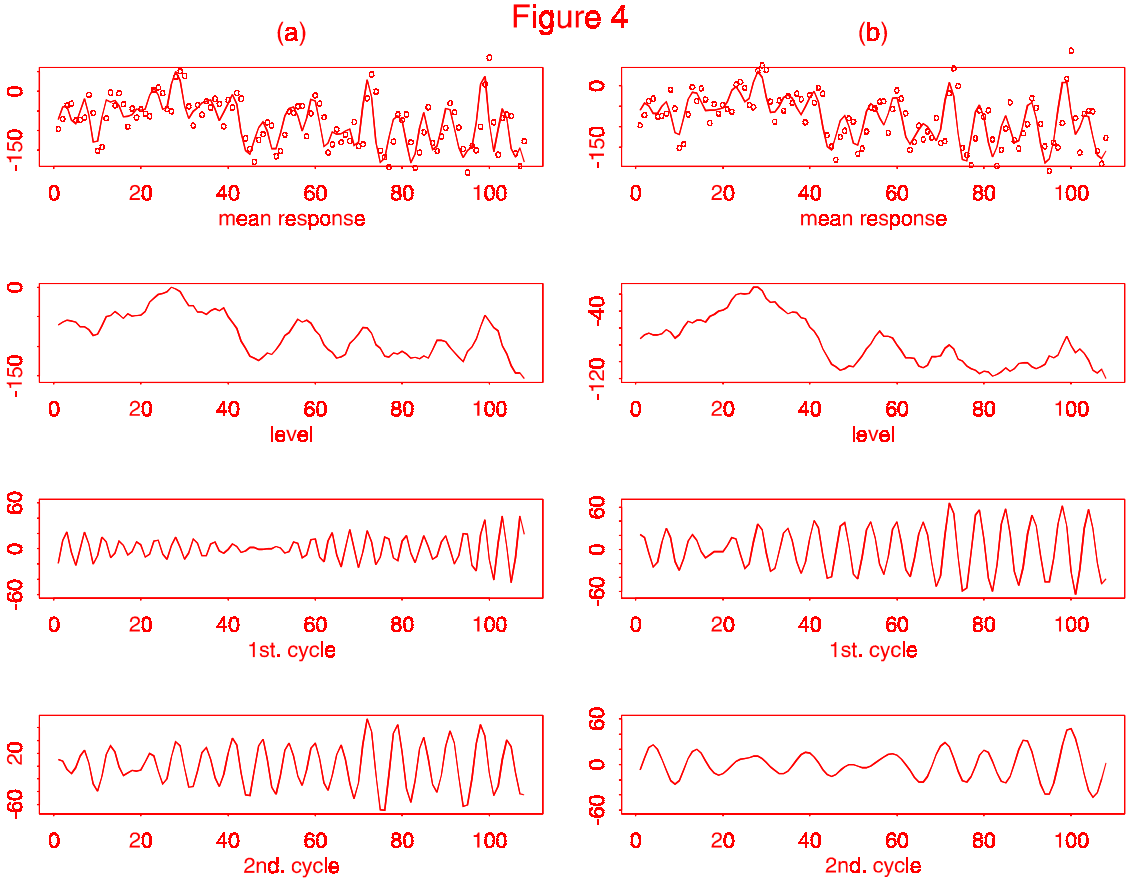


Figure 5

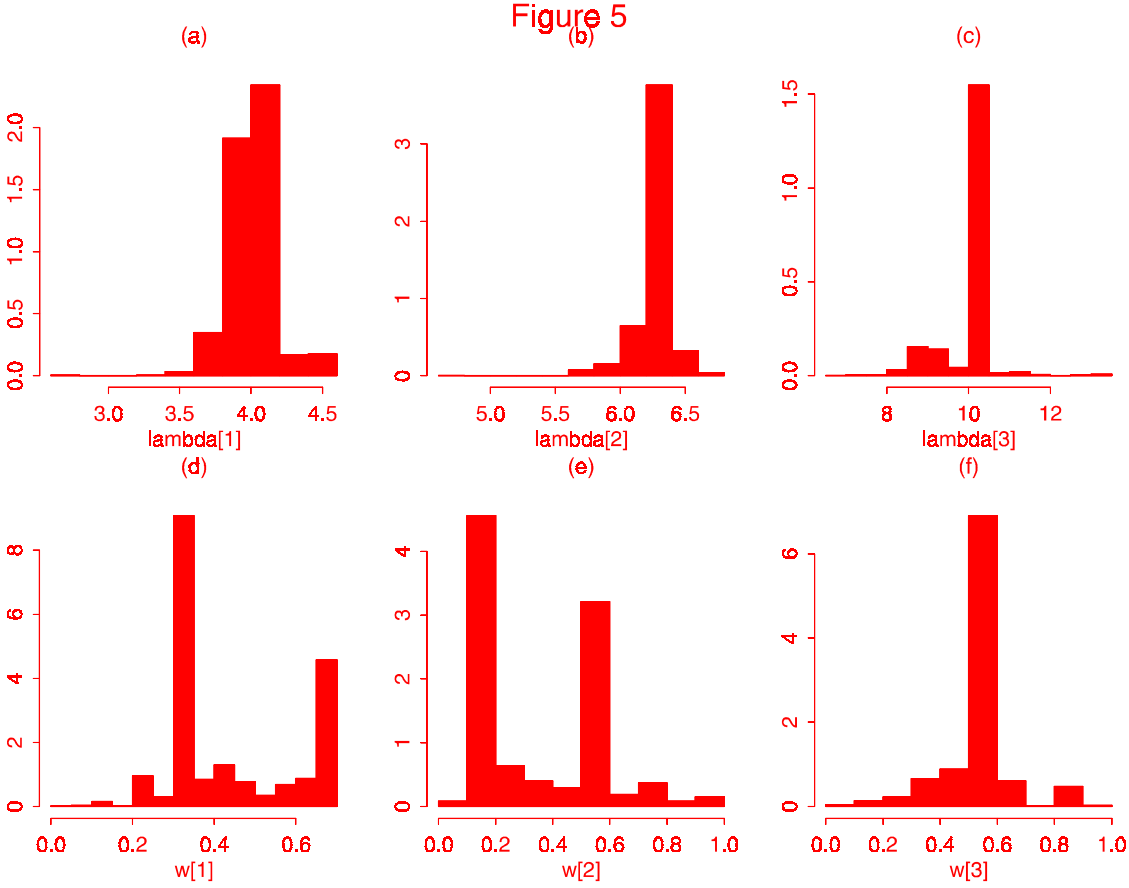


Figure 6

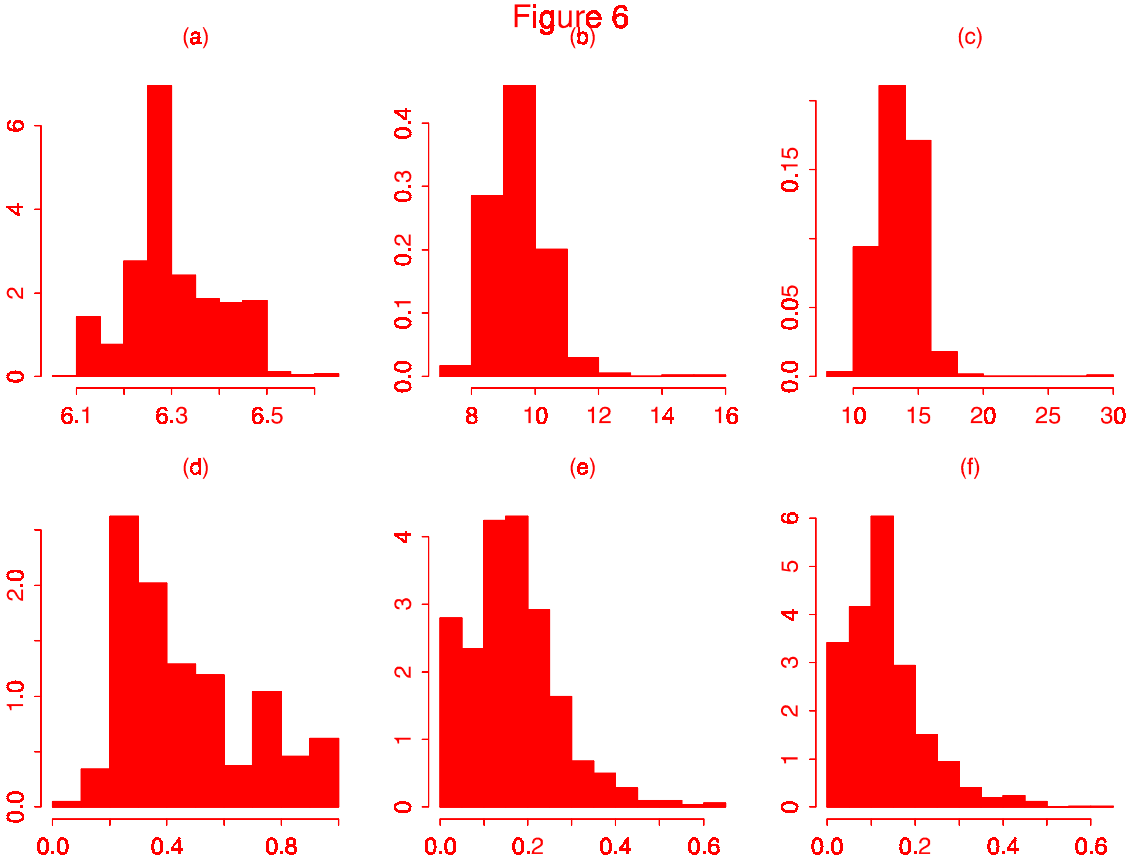

(a)
(b)
(c)
(d)
(e)
(f)
0
2
4
6
6.1
6.3
6.5
0.0
0.1
0.2
0.3
0.4
8
10
12
14
16
0.0
0.05
0.15
10
15
20
25
30
0.0
1.0
2.0
0.0
0.4
0.8
0
1
2
3
4
0.0
0.2
0.4
0.6
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6

Figure 7

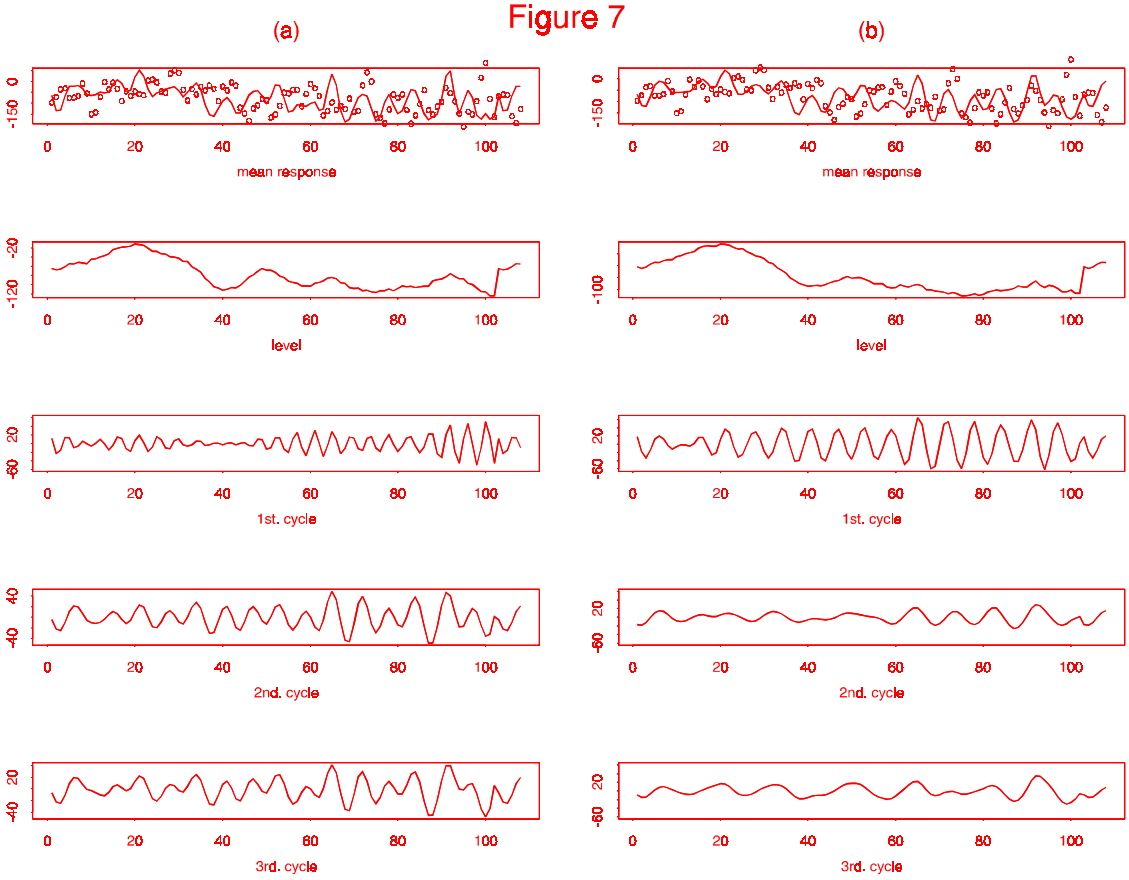
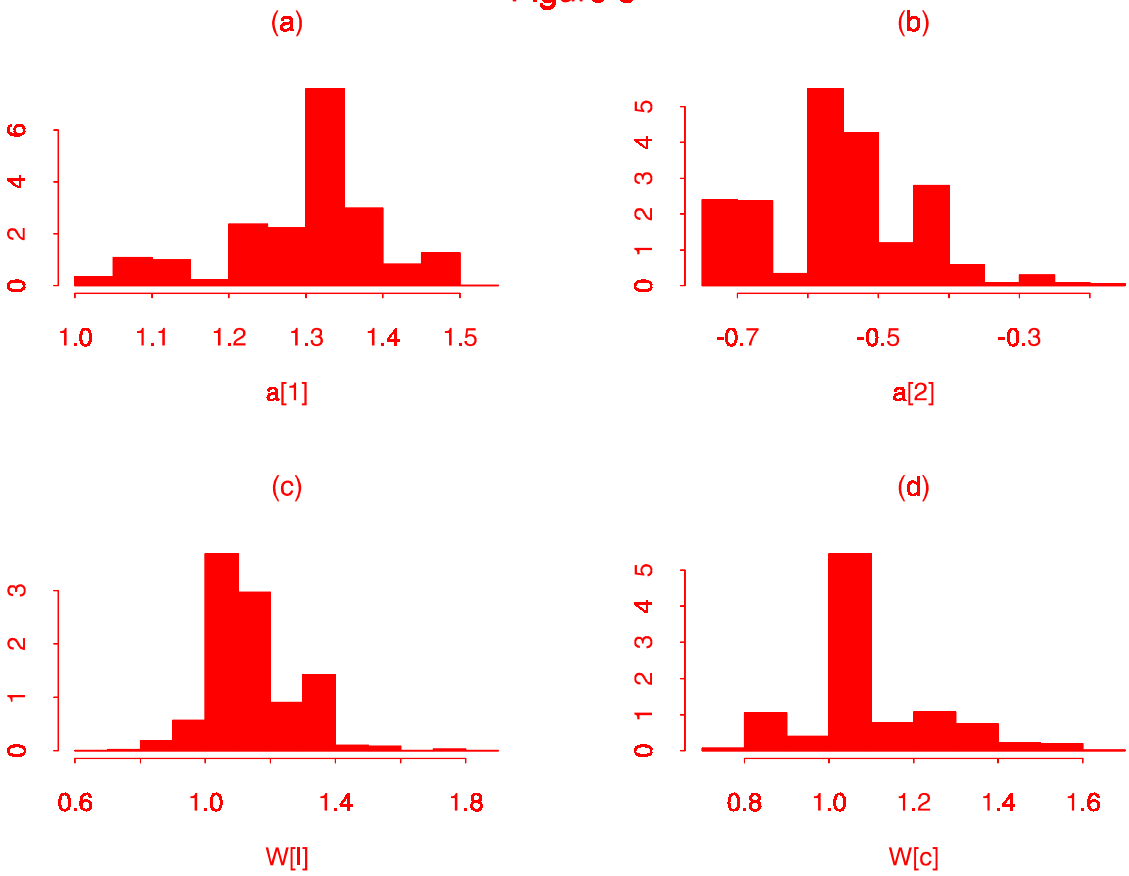


Figure 8


(a)
0 2 4 6
1.0 1.1 1.2 1.3 1.4 1.5
a[1]
(b)
0 1 2 3 4 5
-0.7 -0.5 -0.3
a[2]
(c)
0 1 2 3
0.6 1.0 1.4 1.8
W[l]
(d)
0 1 2 3 4 5
0.8 1.0 1.2 1.4 1.6
W[c]

BIBLIOGRAPHY

BRIAN, B. *Basic optimization methods*, University of Bradford/School of Mathematical Science, 1984.

EFRON, B. *The bootstrap, jacknife and other resampling plans*, Philadelfia: Society of industrial and Applied Mathematics, 1982.

ENDERS, W. *Applied econometric time series*, First Edition, John Wiley & Sons, Canada, 1995.

GELFAND, A. E., SMITH, A. F. M. Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Ass.*, v. 85, p. 398-409, 1990.

HASTINGS, W. K., Monte Carlo sampling methods using Markov Chain and their applications, *Biometrika*, v. 57, p. 97-109, 1970.

LOPES, H., MOREIRA, A. R. B., SCHMIDT, A. M. Estimação de hiperparâmetros em modelos de previsão. *Technical Report*, 449, IPEA, Brasil, 1996.

KENDALL, M. ORD, J. K. *Time Series.* 3rd Edition Edward Arnold, 1990.

KOOPMAN, S. J. HARVEY, A. C., DOORNIK, J. A., SHEPARD, N. STAMP 5.0: *Structural time series analyser.* Modeller and Predictor. Chapman & Hall, London, 1995.

MIGON, H. S., HARRISON, P. J. An application of non-linear Bayesian forecasting to television advertising, in Bayesian Statistics 2, (eds.) BERNARDO, J. M., DEGROOT, M. H., LINDLEY, D. V., SMITH, A. F. M. North-Holland, Amsterdan, 1984.

O'HAGAN, A. *Kendall's advanced theory of statistics*. Volume 2B Bayesian Inference, Edward Arnold, London, 1994.

RUBIN, D. B. Using the SIR algorithm to simulate posterior distributions, in Bayesian Statistics 3, (eds.) BERNARDO, J. M., DEGROOT, M. H., LINDLEY, D. V., SMITH, A. F. M. Cambridge, MA: Oxford University Press, p. 395-402, 1998.

SMITH, A. F. M., GELFAND, A. E. Bayesian statistics without tears: a sampling-resampling perspective. *J. Amer. Statist. Ass.*, v. 46, p.84-88, 1992.

TACHIBANA, V. *Métodos aproximados em modelos Bayesianos de resposta aleatorizada e regressão logística*. PhD thesis, COPPE-UFRJ, 1995.

WEST, M. HARRISON, P. J. Bayesian forecasting and dynamic models, Second Edition, Springer, New York, 1997.

WEST, M. Inference on cycles in time Series, *J. Amer. Statist. Ass.*, v. 90, p. 1301-1312, 1995.

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo