

# Wavelets in State Space Models

<sup>1</sup>Eliana Zandonade and <sup>2</sup>Pedro A. Morettin

<sup>1</sup>Federal University of Esp rito Santo, Vit ria, Brazil

<sup>2</sup>University of S o Paulo, S o Paulo, Brazil

**Abstract.** In this paper we consider the utilization of wavelets in conjunction with state space models. Specifically, the parameters in the system matrix are expanded in wavelet series and estimated via the Kalman Filter and the EM algorithm. In particular this approach is used for switching models. Two applications are given, one to the problem of detecting the paths of targets using an array of sensors, and the other to a series of daily spreads between two Brazilian bonds.

*Kew words:* EM algorithm; Kalman filter; state space models; switching models; wavelets.

## 1 Introduction

This work proposes the utilization of wavelets in state space models. Specifically, parameters in the system matrix are expanded in wavelet series and estimated via the Kalman filter and the EM algorithm. In particular this approach can be useful in the case of switching models.

State space models are currently a powerful tool in many areas of statistics, econometrics and finance. They can model a great variety of time series structures, including missing and uneven observations and allow the use of the Kalman Filter (KF) to compute the likelihood in a simple way. Also, non-linear and non-Gaussian models can be entertained. There is a vast literature on the subject and we refer to Harvey (1990), West & Harrison (1997), Shumway & Stoffer (2000) and Kitagawa & Gersch (1996) as general references where further references and discussion can be found.

Wavelets are a contemporary tool, alternative to classical Fourier analysis. Their advantage is that they are localized in time (or space), being useful for the analysis of non-stationary processes. Good mathematical introductions to wavelets are Chui (1992) and Meyer (1993). For applications in statistics see Vidakovic (1999) and Percival & Walden (2000).

The plan of this article is as follows. In section 2 we review state space models, the EM algorithm and wavelets. In section 3 we consider the case of models with change in regimes and two applications are given. The general situation is discussed in section 4. We conclude the work with some discussion and comments.

## 2 Background

In this section we provide some background information on state space models, on the estimation of parameters and on wavelets. In most of the cases the estimation methods use maximum likelihood estimators or Bayesian methods. We do not discuss the latter here and the interested reader is referred to Harrison & Stevens (1976) and West & Harrison (1997).

### 2.1 State Space Models

Consider the model defined by the equations

$$y_t = F_t \theta_t + v_t, \quad (1)$$

$$\theta_t = G_t \theta_{t-1} + w_t, \quad (2)$$

for  $t = 1, \dots, T$ , where  $y_t$  is an  $r \times 1$  vector of observations,  $F_t$  is an  $r \times p$  matrix (called the system matrix),  $\theta_t$  is a  $p \times 1$  vector of unknown states and  $G_t$  is a  $p \times p$  transition matrix that describes how the states behave across time. The observation error  $v_t$  and the error  $w_t$  associated with the state vector are assumed to be independent Gaussian white noises, with zero means and covariance matrices  $V_t$  and  $W_t$ , respectively. For  $t = 0$  we also assume that  $\theta_0$  is normal, with mean  $\mu$  and covariance matrix  $\Sigma$ . In this way, the process  $y_t$  is completely specified by the so-called characterization vector  $\varphi = (F_t, G_t, V_t, W_t)$ , in the notation of Harrison & Stevens (1976). This vector may depend on a set of unknown parameters that will have to be estimated. For example, an ARMA(p,q) model with time-varying coefficients can be put in the framework (1)-(2). Other instances of the models are the structural models of Harvey (1984) and the dynamic linear models of Harrison & Stevens (1976).

The KF is a recursive procedure used to compute the optimal estimator of the state vector at any instant of time, having information up to time  $t$ , namely  $Y_t = (y_1, \dots, y_t)$ . The procedure may be viewed as two-staged where the first step obtains the best estimate of observation at time  $t$  using the

information up to time  $t - 1$  (corresponds to obtaining the best prediction through the prediction equations) and in the second step, the knowledge of the new available observation is used to update the prediction obtained in the previous step (this is done through the updating equations). The parameters  $\varphi, \mu$  and  $\Sigma$  are assumed to be known for all  $t$ . The mean of  $\theta_t$  obtained by the KF, under normality, is an optimal estimator in the sense of minimizing the mean square error. If we do not have normality, the KF provides the optimal linear prediction (Harvey (1990)). Even with the normality assumption the KF does not provide robust estimates. Meinhold & Singpurwalla (1989) present a method to robustify the KF.

The following notation will be used, with  $t \geq s$ :

$$x_t^s = E(\theta_t | y_1, \dots, y_s), \quad (3)$$

$$P_t^s = \text{Var}(\theta_t | y_1, \dots, y_s), \quad (4)$$

$$P_{t,t-1}^s = \text{Cov}(\theta_t, \theta_{t-1} | y_1, \dots, y_s). \quad (5)$$

Then  $x_t^t$  and  $P_t^t$  will be the estimators derived from the KF,  $x_t^T$  and  $P_t^T$ ,  $t \leq T$  are the smoothed estimators of minimum mean square error, based on all observations  $y_1, \dots, y_T$ , and  $x_t^T, P_t^T$ ,  $t > T$  are the predictors of  $\theta_t$ .

We do not present here the equations of the usual KF. Details can be found in Anderson & Moore (1979). In what follows, modified forms of the KF equations will be employed.

For references on non-linear and non-Gaussian state space models see Fahrmeir (1992), Carlin, Polson & Stoffer (1992), Carter & Kohn (1994), Kitagawa & Gersch (1996) and Durbin & Koopman (1996).

## 2.2 Maximum Likelihood Algorithms

Under the assumption of normality the log-likelihood can be written in the innovations form as

$$\log L = -\frac{T}{2} \log 2\pi - (1/2) \sum_{t=1}^T \log |C_t| - (1/2) \sum_{t=1}^T \nu_t' F_t^{-1} \nu_t, \quad (6)$$

for the observations  $y_1, \dots, y_T$ . See Brockwell and Davis (1991). In (6) we implicitly assume that  $F_t$  is invertible for all  $t$  and

$$C_t = \text{Var}(y_t) = F_t P_t^{t-1} F_t' + V_t,$$

$$\begin{aligned}\nu_t &= y_t - \hat{y}_{t|t-1}, \\ \hat{y}_{t|t-1} &= E(y_t | Y_{t-1}) = F_t x_t^{t-1}.\end{aligned}$$

Briefly, the procedure for obtaining the maximum likelihood estimates (MLE) is the following:

Step 1. Specify initial values of  $\varphi$ ,  $\mu$  and  $\Sigma$ .

Step 2. Apply the KF.

Step 3. Compute (6) and then use the EM algorithm to update  $\varphi$ ; use some other method to update  $\mu$  and  $\Sigma$  (see Harvey (1990) and Shumway, Olsen & Levy (1981) for details).

Step 4. Iterate until convergence.

The form of the EM algorithm to be used here is the one given by Shumway & Stoffer (1982) and presented in Appendix A.

### 2.3 Wavelets

We now turn to some ideas on wavelets. The novel idea on wavelets is that they are *localized* in time (and space), contrary to properties of trigonometric functions. This behavior makes them ideal to analyze nonstationary signals. Fourier bases are localized in frequency but not in time: small changes in some of the observations may induce substantial changes in almost all the components of a Fourier expansion, a property that does not carry over to a more general wavelet expansion.

It is convenient to start with a *father wavelet* or *scaling* function  $\phi$ , such that

$$\phi(t) = \sqrt{2} \sum_k \ell_k \phi(2t - k) \tag{7}$$

and usually normalized as  $\int_{-\infty}^{\infty} \phi(t) dt = 1$ . A *mother wavelet*  $\psi$  is then obtained through

$$\psi(t) = \sqrt{2} \sum_k h_k \phi(2t - k), \tag{8}$$

where  $\ell_k$  and  $h_k$  are related through

$$h_k = (-1)^k \ell_{1-k}. \tag{9}$$

The equations (7) and (8) are called *dilation equations*. The coefficients  $\ell_k, h_k$  are low-pass and high-pass filters, respectively, which appear in the so-called quadrature mirror filters, used in fast algorithms to compute the wavelet transform.

We assume that these functions generate an orthonormal system of  $L_2(\mathfrak{R})$ , which we call  $\{\phi_{j_0,k}(t)\} \cup \{\psi_{j,k}(t)\}_{j \geq j_0,k}$ , with  $\phi_{j_0,k}(t) = 2^{j_0/2} \phi(2^{j_0}t - k)$ ,  $\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)$ , for  $j \geq j_0$  and  $j_0$  is the coarsest scale. Some properties may hold for these wavelets, as the admissibility condition  $\int_{-\infty}^{\infty} \psi(t) dt = 0$ , or that the first  $(r - 1)$  moments of  $\psi$  vanish, for some fixed  $r \geq 1$ . The degree of smoothness of  $\psi$  is then provided by  $r$ .

For any  $f \in L_2(\mathfrak{R})$ , we may then consider the expansion

$$f(t) = \sum_{k=-\infty}^{\infty} \alpha_k \phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_{k=-\infty}^{\infty} \beta_{j,k} \psi_{j,k}(t), \quad (10)$$

for some coarsest scale  $j_0$ , and where the true wavelet coefficients are given by

$$\alpha_k = \int_{-\infty}^{\infty} f(t) \phi_{j_0,k}(t) dt, \quad \beta_{j,k} = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt, \quad (11)$$

following the orthonormality.

An estimate will take the form

$$\hat{f}(t) = \sum_{k=-\infty}^{\infty} \hat{\alpha}_k \phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_{k=-\infty}^{\infty} \hat{\beta}_{j,k} \psi_{j,k}(t), \quad (12)$$

where the  $\hat{\alpha}_k, \hat{\beta}_{j,k}$  are estimates of  $\alpha_k, \beta_{j,k}$  respectively.

Several issues are of interest here:

- (i) the choice of the wavelet basis;
- (ii) the choice of a thresholding policy;
- (iii) the choice of the parameters appearing in the thresholding scheme;
- (iv) the estimation of the scale parameter (noise level).

We discuss briefly (i) now. For further details on (ii)-(iv) see Morettin (1997).

Concerning the choice of the wavelet basis, some possibilities are the Haar, compactly supported wavelet bases (Daubechies (1992)), complex

wavelets (Morlet, or modulated Gaussian), Mexican hat (second derivative of Gaussian), Shannon, Meyer, etc.

The form of the signal to be analysed may lead to a particular basis, for example the Haar which is useful for categorical-type signals. It is based on

$$\phi(t) = 1, 0 \leq t < 1 \quad (13)$$

and

$$\psi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \end{cases} \quad (14)$$

and the expansion is then

$$f(t) = \alpha_0 + \sum_{j=0}^J \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk}(t) \quad (15)$$

In practice the time period of observation will be rescaled to the unit interval.

In this paper we will use the Haar, Morlet and Shannon wavelets. These are generated in order to get orthonormal systems. The Morlet wavelet has mother function

$$\psi(t) = e^{i\omega t} e^{-t^2/2},$$

where  $i = \sqrt{-1}$  and  $\omega$  is some fixed frequency. The Shannon wavelet has mother function

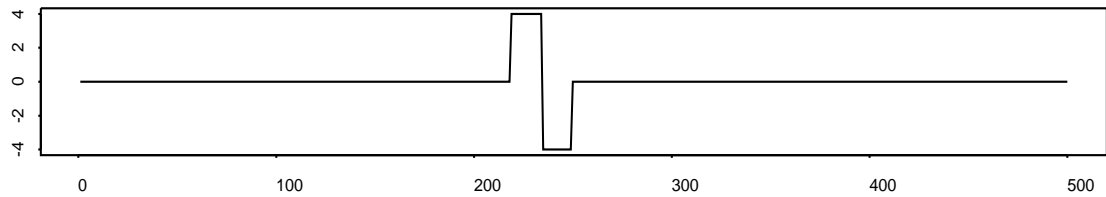
$$\psi(t) = \frac{\sin(\pi t/2)}{\pi t/2} \cos(3\pi t/2).$$

Figure 1 shows the plots of these wavelets. Several software packages are now available for wavelets analysis in statistical problems. We mention the module S+WAVELETS of SPlus (Bruce & Gao (1996)), WaveLab (Donoho, Huo, Duncan & Levi-Tsabari (2000)) and Wavethresh (Nason & Silverman (1994)). We have developed our own programs in S and Fortran 90 to perform the applications of the paper.

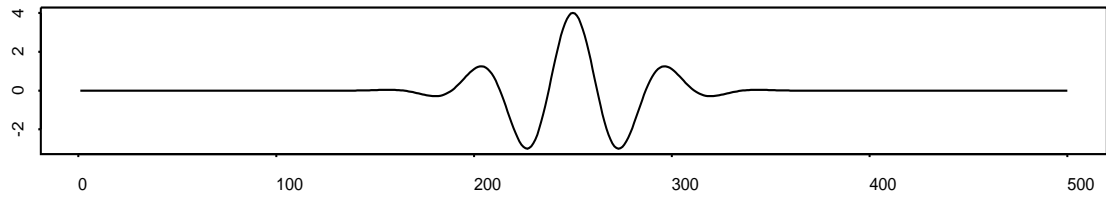
## 3 Models with Switching

### 3.1 The model

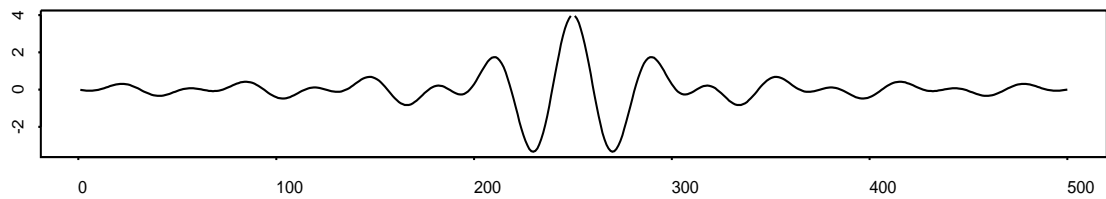
One way of modeling changes of behavior in time is through the use of dynamical linear systems in the form of state space models. We will focus on a proposal by Shumway & Stoffer (1991). These authors consider the



(a)



(b)



(c)

Figure 1: Some wavelets (a) Haar (b) Morlet (c) Shannon

setup (1)-(2) plus the assumptions made therein. To incorporate a switching structure in the system matrix  $F_t$ , it was assumed that there are  $m$  possible configurations, which are the states of an independent non-stationary process governed by the probabilities

$$\pi_\ell(t) = P(F_t = M_\ell), \ell = 1, 2, \dots, m, t = 1, \dots, T, \quad (16)$$

independently of  $F_1, \dots, F_{t-1}$  and past history  $y_1, \dots, y_{t-1}$ .

Important piece of information about the current state is given by the filtered probabilities for the state  $\ell$ ,

$$\pi_\ell(t|t) = P(F_t = M_\ell | Y_t), \quad (17)$$

for  $Y_s = (y_1, \dots, y_s)$ ,  $s = 1, \dots, T$ .

Shumway & Stoffer (1991) obtain estimates for the state vector and for the probabilities associated to the form of  $F_t$ , using the KF and a pseudo-EM algorithm for the parameter estimation. Kim (1994) extended this idea, introducing a dependence in the process of change of regime and allowing for a variation to occur in  $F_t$  as well as in the transition matrix  $G_t$ . Another reference is Hamilton (1989).

Following the approach of Shumway & Stoffer (1991) we propose a model for the probabilities  $\pi_\ell(t)$  using wavelets, which describe the variation of these probabilities along time. In order to guarantee that the estimated probabilities lie in  $(0, 1)$  we use the logistics-type parametrization

$$\pi_\ell(t) = \frac{\exp\{r_\ell(t)\}}{\sum_{i=1}^m \exp\{r_i(t)\}}, \ell = 1, \dots, m, \quad (18)$$

where

$$r_\ell(t) = \sum_{n=1}^s \beta_n^{(\ell)} \psi_n(t/T), \ell = 1, \dots, m-1, r_m(t) = 0, \quad (19)$$

or, equivalently,

$$r_\ell(t) = \begin{cases} \frac{\exp\{\sum_{n=1}^s \beta_n^{(\ell)} \psi_n(t/T)\}}{1 + \sum_{i=1}^{m-1} \exp\{\sum_{n=1}^s \beta_n^{(i)} \psi_n(t/T)\}}, & l = 1, \dots, m-1, \\ \frac{1}{1 + \sum_{i=1}^{m-1} \exp\{\sum_{n=1}^s \beta_n^{(i)} \psi_n(t/T)\}}, & l = m. \end{cases} \quad (20)$$

Here we have denoted by  $s$  the total number of parameters  $\beta_{j,k}$  associated to the wavelets  $\psi_{j,k}(t/T)$ , according to the notation in section 2.2.



### 3.2 Estimation of Parameters

The estimation of parameters of the model proceeds as follows.

Stage 1. Estimation of the state vector  $\theta_t$

In this step an extension of the classical KF, given in Shumway & Stoffer (1991), is employed and uses estimates of the filtered probabilities  $\pi_\ell(t|t)$ . See Appendix B for details.

Stage 2. Estimation of  $\beta_n^{(\ell)}, V_t, W_t$

Here the EM algorithm is used. Compute the log-likelihood of  $\theta_0, \dots, \theta_T, y_1, \dots, y_T$ , with  $(\theta_0|D_0) \sim N(\mu, \Sigma)$ , where  $D_0$  is the information set at time  $t = 0$ :

$$\begin{aligned}
 \log L &= -(1/2) \log |\Sigma| - (1/2)(\theta_0 - \mu)' \Sigma^{-1} (\theta_0 - \mu) - (T/2) \log |W_t| \\
 &\quad - (1/2) \sum_{t=1}^T (\theta_t - G_t \theta_{t-1})' W_t^{-1} (\theta_t - G_t \theta_{t-1}) \\
 &\quad + \sum_{t=1}^T \sum_{\ell=1}^m I(F_t = M_\ell) \log(\pi_\ell(t)) - (T/2) \log |V_t| \\
 &\quad - (1/2) \sum_{t=1}^T \sum_{\ell=1}^m I(F_t = M_\ell) (y_t - F_t \theta_t)' V_t^{-1} (y_t - F_t \theta_t),
 \end{aligned} \tag{21}$$

where  $I(\cdot)$  is the indicator function. See Appendix C for the details of steps E and M of the algorithm.

### 3.3 Applications

Consider first an example of Shumway & Stoffer (1991), where the problem consists of detecting the paths of targets ( $\theta_t$ ) using an array of sensors ( $y_t$ ). It is not known, at a given instant of time, which target was detected by a given sensor. Here  $F_t$  will represent the changes in position of a target. The authors considered  $m = 10$  possible matrices  $F_t$ , shown in Table 1. For example, for  $M_1$  we have

$$F_t = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

Table 1: Possible matrices  $F_t$

Matrix	$y_{t_1}$	$y_{t_2}$	$y_{t_3}$
$M_1$	$\theta_{t_1}$	$\theta_{t_1}$	$\theta_{t_1}$
$M_2$	$\theta_{t_1}$	$\theta_{t_1}$	$\theta_{t_2}$
$M_3$	$\theta_{t_1}$	$\theta_{t_2}$	$\theta_{t_1}$
$M_4$	$\theta_{t_2}$	$\theta_{t_1}$	$\theta_{t_1}$
$M_5$	$\theta_{t_1}$	$\theta_{t_3}$	$\theta_{t_2}$
$M_6$	$\theta_{t_1}$	$\theta_{t_2}$	$\theta_{t_3}$
$M_7$	$\theta_{t_2}$	$\theta_{t_1}$	$\theta_{t_3}$
$M_8$	$\theta_{t_2}$	$\theta_{t_3}$	$\theta_{t_1}$
$M_9$	$\theta_{t_3}$	$\theta_{t_2}$	$\theta_{t_1}$
$M_{10}$	$\theta_{t_3}$	$\theta_{t_1}$	$\theta_{t_2}$

meaning that all three sensors detected target  $\theta_{t_1}$ .

In this situation we take  $G_t \equiv G$ ,  $V_t \equiv V$  and  $W_t \equiv W$  as the matrices

$$G = \begin{bmatrix} 1.005 & 0 & 0 \\ 0 & 0.990 & 0 \\ 0 & 0 & 1.000 \end{bmatrix}, \quad V = \begin{bmatrix} 0.0625 & 0 & 0 \\ 0 & 0.0625 & 0 \\ 0 & 0 & 0.0625 \end{bmatrix},$$

$$W = \begin{bmatrix} 0.0025 & 0 & 0 \\ 0 & 0.0025 & 0 \\ 0 & 0 & 0.0025 \end{bmatrix}.$$

Also, let  $\mu = (5, 5, 5)'$ , and  $T = 100$ . As an example, we take  $M_5$ , for  $1 \leq t \leq 30$  and  $50 \leq t \leq 69$  and  $M_9$ , for  $31 \leq t \leq 49$  and  $70 \leq t \leq 100$ . We present the estimation for four models: The model entertained by Shumway & Stoffer (1991), which considers the probabilities associated to the matrices  $F_t$  as 0.5, and the others which use Haar, Morlet and Shannon wavelets to model these probabilities. Here we also take  $s = 3$ , that is, considered

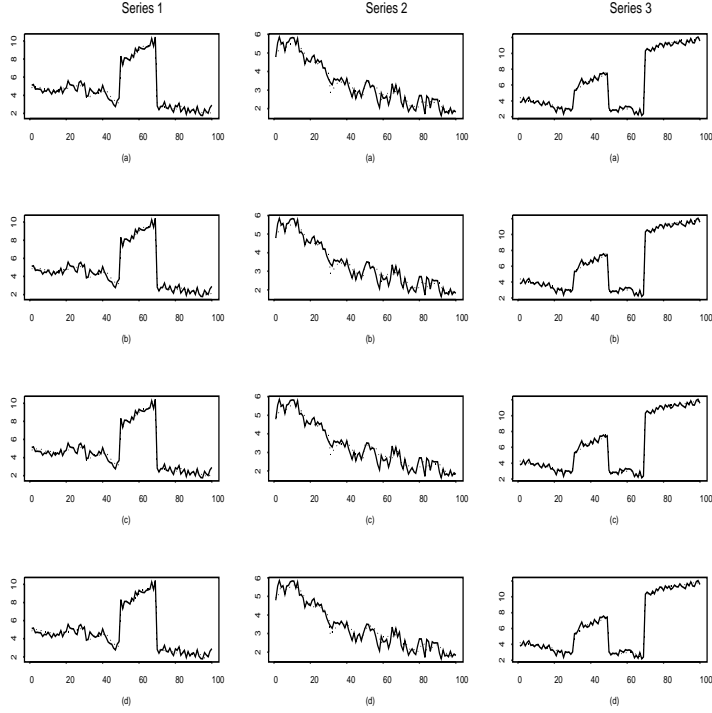


Figure 2: Plots of the series and respective estimates for the several models. (a) Shumway and Stoffer (b) Haar (c) Morlet (d) Shannon

the first three wavelets with scale factors  $j = 1, 2$  and translation factor  $k = 1, \dots, 2^{j-1}$ . To arrive at the estimates of the probabilities  $\pi_\ell(t)$  a non-linear system with  $s = 3$  equations had to be solved.

In Figure 2 we have plots of the series and predictions from the four models indicated above. Figure 3 plots states and respective estimates. We can see that the results are quite good. Figure 4 shows the estimates of the probabilities  $\pi(t)$  and  $\pi(t|t)$ . We see that the Morlet and Shannon wavelet models captured the differences in  $M_5$  and  $M_9$  manifested in  $\pi_5(t)$  and  $\pi_9(t)$ . Table 2 presents the estimation results obtained.

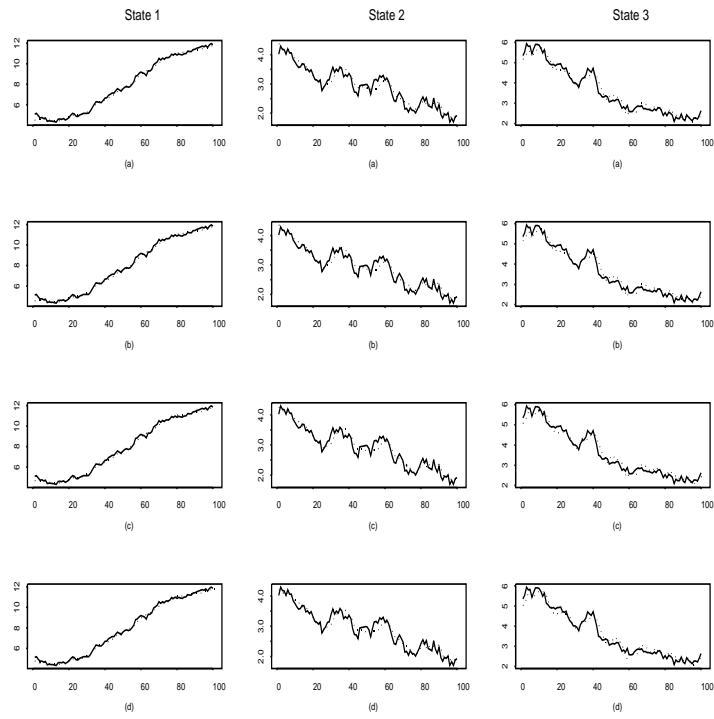


Figure 3: Plots of the states  $\theta_{t_1}$ ,  $\theta_{t_2}$  and  $\theta_{t_3}$  and respective estimates for the several models. (a) Shumway and Stoffer (b) Haar (c) Morlet (d) Shannon

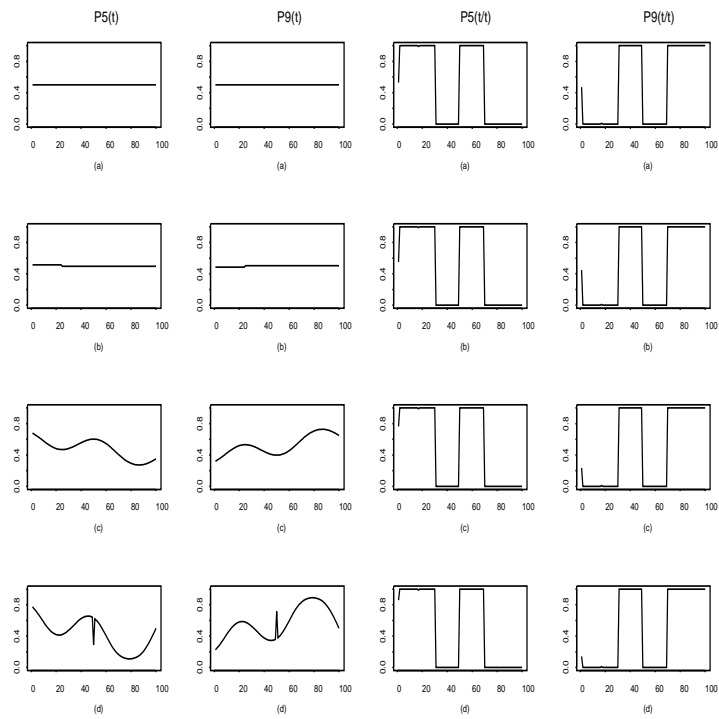


Figure 4: Plots of the probabilities and filtered probabilities for the several models (a) Shumway and Stoffer (b) Haar (c) Morlet (d) Shannon

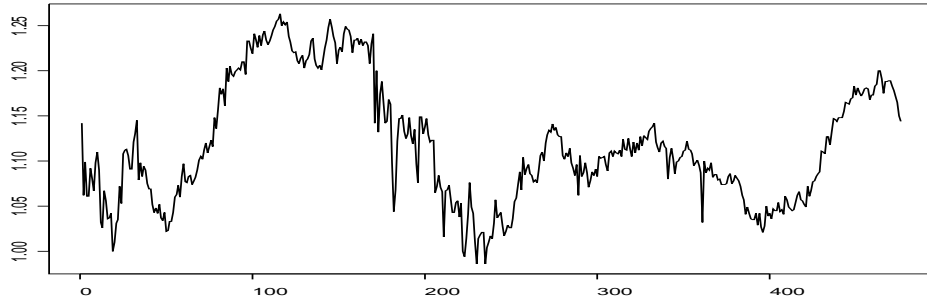


Figure 5: Plot of the daily spread

We consider next an application of the switching model to a series with autoregressive dynamics, namely the daily spread between Brazilian C-bond and Par-bond.

Let  $c_t$  be the closing price of the C-bond on day  $t$  and  $p_t$  the corresponding price of the Par-bond. The daily spread is defined to be  $y_t = c_t/p_t$ . We have  $T = 476$  observations from January 21, 1994 and March 13, 1996. Figure 5 plots the series. Previous analysis of the series suggests that it oscillates around a mean level of about 1,100, displaying autocorrelation in its dynamics, in the sense that the series drifts around this mean level through short drifts. Market factors that influence both series cause this behavior. Basically three regimes are noticed: the first, when the series oscillates at a level inferior to the global mean level (of short duration); a second regime, for which the series oscillates around the global mean level (longer duration); and finally a third regime, when the oscillations are around a level superior to the global average (short duration).

Take two states for  $F_t$ :  $F_t = 1$  with probability  $P_1(t)$  and  $F_t = 0$  with probability  $P_2(t)$ . The probability  $P_1(t)$  will be modelled by a wavelet expansion with three coefficients,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$ . We use the same wavelets as before. In order to obtain the mean and standard error of the estimates we use a bootstrap procedure as developed in Stoffer and Wall (1991). We refer to this work for details. For each model we simulate 1,000 bootstrap

Table 2: Results for the example using the models Shumway & Stoffer (1991), Haar, Morlet and Shannon wavelets, respectively. (\*) Model with larger log-likelihood.

Parameters	Shumway & Stoffer	Haar	Morlet	Shannon
Iteration	51	51	51	51
log-likel.	-98.4054	-97.4826	-88.1812 (*)	-95.4079
$G_{11}$	1.0035	1.0036	1.0043	1.0046
$G_{22}$	0.9376	0.9380	0.9401	0.9398
$G_{33}$	0.8732	0.8729	0.8686	0.8655
$V_{11}$	0.0930	0.0927	0.0887	0.0866
$V_{22}$	0.0487	0.0487	0.0473	0.0460
$V_{33}$	0.0685	0.0683	0.0662	0.0646
$W_{11}$	0.0164	0.0166	0.0185	0.0198
$W_{22}$	0.0184	0.0184	0.0189	0.0194
$W_{33}$	0.0387	0.0389	0.0419	0.0440
$\beta_1$	-	0.0206	0.6694	1.4428
$\beta_2$	-	0.0291	0.3199	-0.0876
$\beta_3$	-	0	0.5935	1.0337

replicates. In Figure 6 we have the estimated series, states and probabilities. For the Morlet and Shannon wavelets the probabilities  $P_1(t)$  and  $P_2(t)$  show a variation across time which mimic changes of the series. The Haar wavelet estimates both probabilities around 0.5. For all models the filtered probabilities are equal to 1 for  $P_1(t)$  and 0 for  $P_2(t)$ .

Figure 7 shows the histograms for the 1,000 bootstrap samples for each parameter and each model. We see that there was a concentration of sampled values around a single interval for  $G_t$ ,  $V_t$  and  $W_t$ . There was a great dispersion for the wavelet coefficients, with isolated groups. Further work is needed here for a more detailed study of the distribution of these estimates.

## 4 A General Formulation

In order to model change in the system matrix  $F_t$  we can consider its expansion in wavelets. Assume  $F_t = [F_t^{(u,v)}]$ , the model (1)-(2) as described before and assume further that  $r = p = 1$ ,  $G_t \equiv G$ ,  $V_t \equiv V$  and  $W_t \equiv W$ . Therefore

$$F_t = \sum_j \sum_k \gamma_{j,k} \psi_{j,k}(t/T). \quad (22)$$

Replacing as before the indices  $j, k$  by a unique index  $\ell = 1, \dots, s$  where  $s$  is the total number of  $\gamma_{j,k}$ 's, we are led to solve  $s$  equations in the form

$$\sum_{t=1}^T [y_t x_t^T \psi_\ell(t/T)] - \sum_{t=1}^T [C_t (\sum_{i=1}^s (\hat{\gamma}_i \psi_i(t/T)) \psi_\ell(t/T)] = 0, \quad (23)$$

for  $\ell = 1, 2, \dots, s$ , where  $x_t^T$  and  $P_t^T$  are as in (3)-(4) and  $C_t = P_t^T + x_t^T (x_t^T)'$ .

We can write the system as  $Z\hat{\gamma} = U$ , where the  $(i, j)$ -th element of  $Z$  is  $\sum_{t=1}^T C_t \psi_i(t/T) \psi_j(t/T)$ ,  $i, j = 1, \dots, s$ ,  $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_s)'$  and the  $i$ -th element of the vector  $U$  is  $\sum_{t=1}^T y_t x_t^T \psi_i(t/T)$ ,  $i = 1, \dots, s$ .

If the inverse of  $Z$  exists then  $\hat{\gamma} = Z^{-1}U$ .

(i) Estimate of  $G$ :

$$\hat{G} = \left( \sum_{t=1}^T A_t \right)^{-1} \left( \sum_{t=1}^T B_t \right),$$

where  $A_t = P_{t-1}^T + x_{t-1}^T x_{t-1}^{T'}$ ,  $B_t = P_{t,t-1}^T + x_t^T x_{t-1}^{T'}$ , and  $x_t^T$ ,  $P_t^T$  and  $P_{t,t-1}^T$  are obtained using the Kalman smoother.



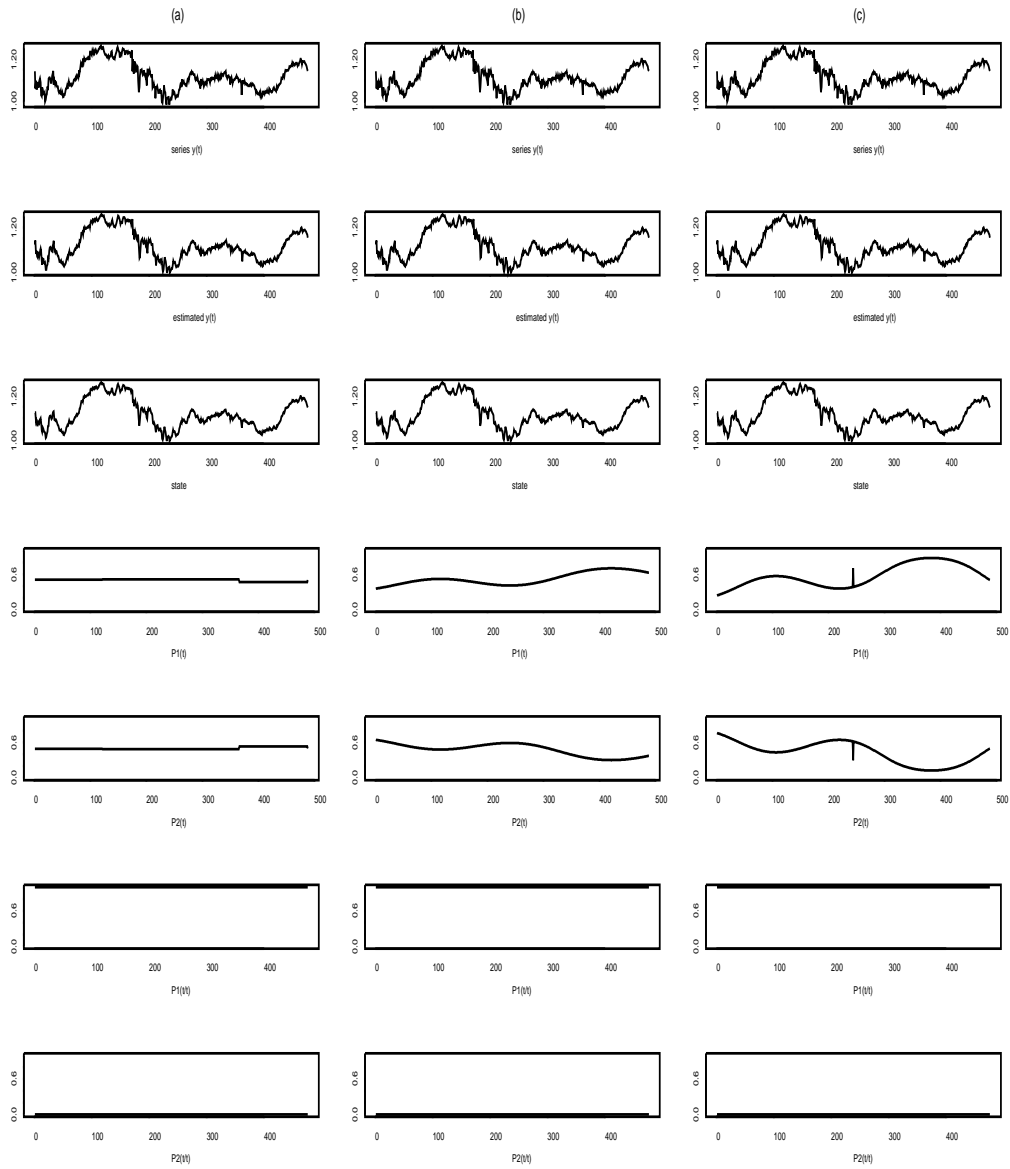


Figure 6: Plots of estimates of series, states, probabilities and filtered probabilities for the models: (a) Haar (b) Morlet (c) Shannon

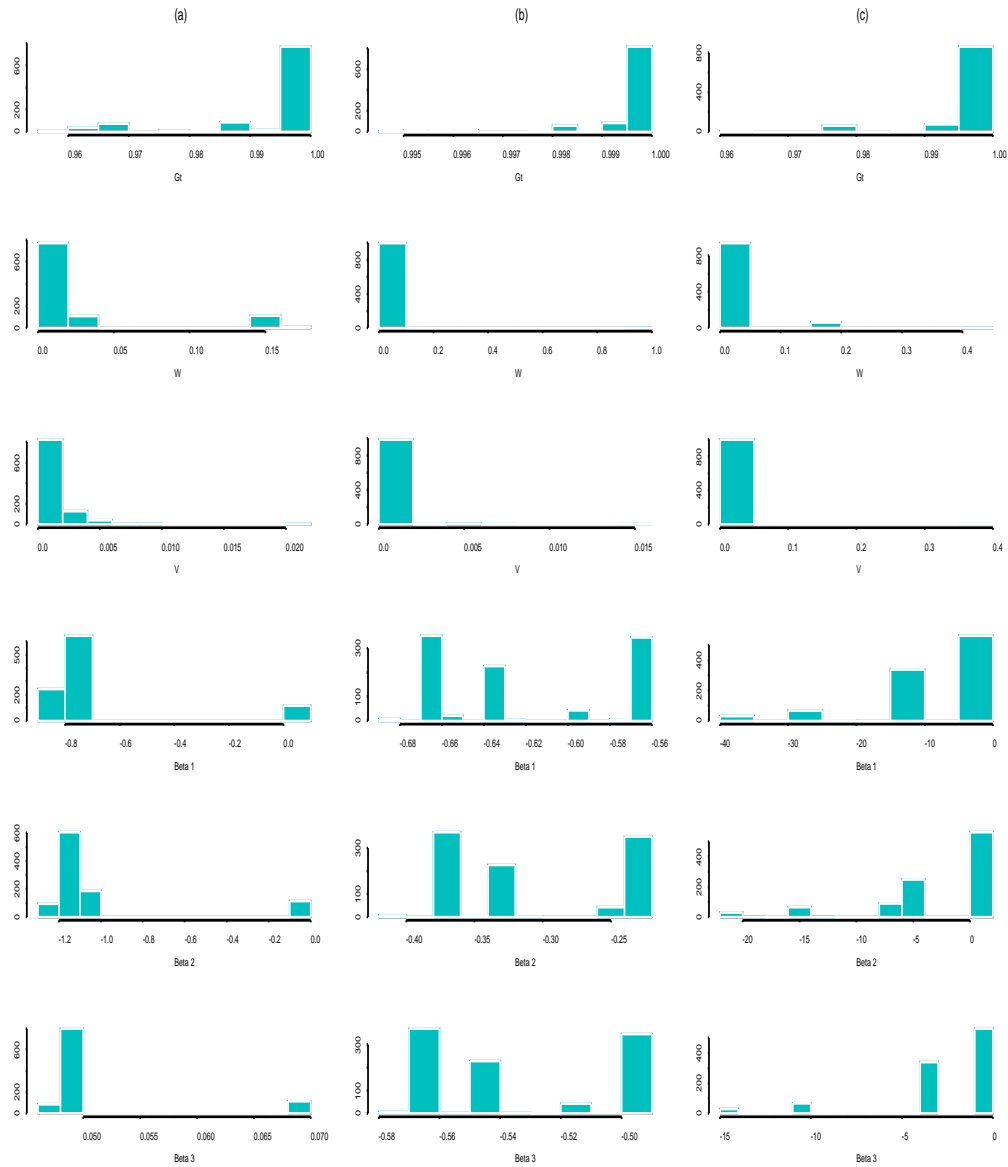


Figure 7: Histograms for the bootstrap samples of  $G$ ,  $W$ ,  $V$ ,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  for the models with : (a) Haar (b) Morlet (c) Shannon

(ii) Estimate of  $W$ :

$$\hat{W} = (1/T) \sum_{t=1}^T [C_t - 2B_t \hat{G}_t + A_t \hat{G}_t^2].$$

(iii) Estimate of  $V$ :

$$\hat{V} = (1/T) \sum_{t=1}^T [(y_t - \hat{F}_t x_t^T)^2 + P_t^T \hat{F}_t^2],$$

where

$$\hat{F}_t = \sum_{n=1}^s \hat{\gamma}_n \psi_n(t/T).$$

For this situation the application of the EM algorithm is done in two stages, through a profile likelihood approach (see Richards (1961), for details). Here,  $W$  is the perturbation parameter to be estimated in the first stage of the procedure and the remaining parameters are estimated in the second stage.

To illustrate the methodology, we have done a simulation using Haar wavelets in (22). The estimation depends on the initial values given for  $W$  in stage one. This being the case, we estimated models that assumed four initial values for  $W$ : 0.001, 0.01, 0.1 and 1. We generated  $T = 256$  values and fitted models with the largest scale  $J$  varying from 2 to 6. We fixed the following values for the parameters:  $G = 1, W = 0.16, V = 0.25$ . We report in Table 3 and Figure 8 the case with  $J = 2$ , that is, using three Haar wavelets. We see that the results are very reasonable, except for the initial value  $W = 1$ .

## 5 Comments

In this work we have proposed the use of wavelets to model parameters present in the system matrix  $F_t$  of a state space model. In particular we

Table 3: Results obtained for the example using models with Haar wavelets,  $J = 2$ , for several initial values of  $W$ . (\*) is the best model according to BIC.

$W$	Iteration	$-2\log L$	W	V	$G_t$	BIC
0.001	11	276.3841	0.1145	0.5673	0.9974	288.4254
0.01	11	275.9883	0.1645	0.5589	0.9972	288.0296
0.1	11	275.8095	0.1976	0.5563	0.9971	287.8507 (*)
1	11	439.0225	24.6767	0.8286	0.9896	451.0637

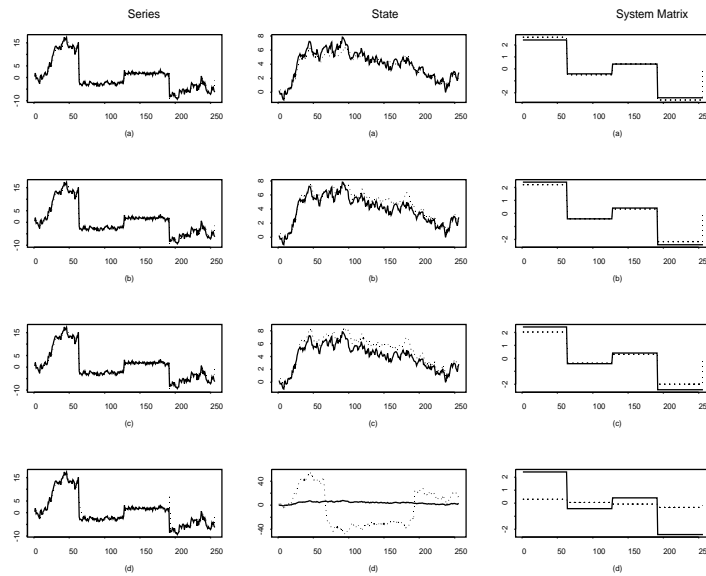


Figure 8: Plots of the series  $y_t$ , state  $\theta_t$  and system matrix  $F_t$  (solid line) and respective estimates (dashed line) for the four models, using Haar wavelets with  $J = 2$  and 4 initial values. (a)  $W = 0.001$  (b)  $W = 0.01$  (c)  $W = 0.1$  (d)  $W = 1$

have dealt with the case of time series with switching, for which certain probabilities can be expanded in wavelets. Another possibility is to use wavelets to model the transition matrix  $G_t$ . This situation will be considered elsewhere. Simulations made suggest that the proposal leads to useful results. The use of the Kalman filter associated with the EM algorithm has proved to be a good combination for the purpose of parameter estimation. We have considered also, through a real example, the problem of obtaining standard errors for the estimates. This can be done using the bootstrap procedure as suggested by Stoffer & Wall (1991).

**Acknowledgement.** We would like to thank a referee for comments that led to the improvement of the paper.

## References

- Anderson, B. D. O. & Moore, J. B. (1979). *Optimal Filtering*, Prentice Hall, Englewood Cliffs, N. J. .
- Brockwell, P. & Davis, R.A. (1991 ). *Time Series: Theory and Methods*, Springer, New York.
- Bruce, A. & Gao, H. (1996). *S+ WAVELETS Users Manual*, MathSoft Inc., Seattle.
- Carlin, B. P., Polson, N. G. & Stoffer, D. S. (1992). A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling, *Journal of the American Statistical Association* **87**(418): 493–500.
- Carter, C. K. & Kohn, R. (1994). On Gibbs Sampling for State-Space Models, *Biometrika* **81**(3): 541–553.
- Chui, C. K. (1992). *An Introduction to Wavelets*, Academic Press, San Diego.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, SIAM , Philadelphia.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Series B* **39**: 1–38.
- Donoho, D., Huo, X., Duncan, M. & Levi-Tsabari, O. (2000). *About WaveLab*, Stanford University, <http://www-stat.stanford.edu/~wavelab>.

- Durbin, J. & Koopman, S. J. (1996). Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models, *Technical report*, London School of Economics and Political Science, London.
- Fahrmeir, L. (1992). *State Space Modelling and Conditional Mode Estimation for Categorical Time Series*, Springer, New York, chapter New Directions in Time Series Analysis, Part I, pp. 87–109.
- Hamilton, J. D. (1989). A new Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica* **57**(2): 357–384.
- Harrison, P. J. & Stevens, C. F. (1976). Bayesian Forecasting, *Journal of the Royal Statistical Society, Series B* **38**: 205–247.
- Harvey, A. C. (1984). A Unified View of Statistical Forecasting Procedure, *Journal of Forecasting* **3**: 245–275.
- Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- Kim, C. J. (1994). Dynamic Linear Models with Markov-Switching, *Journal of Econometrics* **60**: 1–22.
- Kim, C. & Nelson, C. (1999). *State Space Models With Regime Switching*, The MIT Press, Cambridge, Massachusetts.
- Kitagawa, G. & Gersch, W. (1996). *Smoothness Priors Analysis of Time Series. Lecture Notes in Statistics, 116*, Springer Verlag, New York.
- Meinhold, R. J. & Singpurwalla, N. D. (1989). Robustification of Kalman Filter Models, *Journal of the American Statistical Association* **84**(406): 479–486.
- Meyer, Y. (1993). *Wavelets: Algorithms and Applications*, SIAM, Philadelphia.
- Morettin, P. (1997). Wavelets in Statistics, *Resenhas IME-USP* **3**(2): 211–272.
- Nason, G. & Silverman, B. (1994). The Discrete Wavelet Transform in S, *J. Computational and Graphical Statistics* **3**: 163–191.
- Percival, D. & Walden, A. T. (2000). *Wavelet Methods for Time Series*, Oxford University Press, Oxford.

- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in Fortran 77, The Art of Scientific Computing, Vol. 1*, Cambridge University Press, Cambridge.
- Richards, F. S. G. (1961). A Method of Maximum Likelihood Estimation, *Journal of the Royal Statistical Society, Series B* **23**: 469–475.
- Shumway, R. H., Olsen, D. E. & Levy, L. J. (1981). Estimation and Tests of Hypotheses for the Initial Mean and Covariance in the Kalman Filter Model, *Communications Statistics-Theory Methods* **10**(16): 1625–1641.
- Shumway, R. H. & Stoffer, D. S. (1982). An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm, *Journal of Time Series Analysis* **3**(4): 253–264.
- Shumway, R. H. & Stoffer, D. S. (1991). Dynamic Linear Models with Switching, *Journal of the American Statistical Association* **86**(415): 763–769.
- Shumway, R. & Stoffer, D. (2000). *Time Series Analysis and Its Applications*, Springer-Verlag, New York.
- Stoffer, D. S. & Wall, K. D. (1991). Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter, *J. Amer. Statist. Association* **86**(416): 1024–1033.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, J. Wiley and Sons, New York.
- West, M. & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Linear Models, Second Edition*, Springer-Verlag, New York.

## Appendix A: The EM Algorithm

The Expectation-Maximization algorithm (or EM algorithm) is a non-linear optimization algorithm appropriate for applications involving non-observed components or observations irregularly spaced in time. It is an iterative procedure for computing the maximum likelihood estimator when a subset of the total set of data is missing.

Dempster, Laird & Rubin (1977) demonstrated the applicability of the EM algorithm and popularized the method in statistics. In the usual formulation of the algorithm, the vector of complete data is made of the observed series ( $Y$ ) and of the non-observed process ( $\theta$ ). In several applications  $\theta$  consists of a “latent” process.

In the state space model, the data  $y_1, \dots, y_T$  form the observed vector and  $\theta_1, \dots, \theta_T$  is the non-observed state vector. The application of the EM algorithm for this model requires the minimum mean square error estimate of the state vector, based on the set of all observations  $x_t$ ,  $t = 1, \dots, T$ . This estimate will be denoted by  $x_t^T$ ,  $t = 1, \dots, T$  and its covariance matrix by  $P_t^T$ ,  $t = 1, \dots, T$ . These are obtained by the smoothing algorithm, requiring one step forward and another backwards (Kalman filter and smoother, respectively).

Following Shumway & Stoffer (1982), the EM algorithm for the state space model can be put as follows:

E STEP : Compute the expectation of the joint log-likelihood of  $\theta_0, \dots, \theta_T$  and  $y_1, \dots, y_T$ , given  $y_1, \dots, y_T$ , obtaining

$$E(\log L|y_1, \dots, y_T) \doteq H(\mu, \Sigma, F, G, V, W).$$

M STEP : Maximize  $H(\mu, \Sigma, F, G, V, W)$  with respect to the parameters  $\varphi = (F, G, V, W)$ . The solutions are

$$\begin{aligned} F &= DC^{-1}, \\ G &= BA^{-1}, \\ V &= T^{-1} (C - BA^{-1}B'), \\ W &= T^{-1} \sum_{t=1}^T \left[ (y_t - Fx_t^T) (y_t - Fx_t^T)' + FP_t^T F' \right], \end{aligned}$$

where

$$\begin{aligned} A &= \sum_{t=1}^T (P_{t-1}^T + x_{t-1}^T x_{t-1}^{T'}), \\ B &= \sum_{t=1}^T (P_{t,t-1}^T + x_t^T x_{t-1}^{T'}), \\ C &= \sum_{t=1}^T (P_t^T + x_t^T x_t^{T'}), \\ D &= \sum_{t=1}^T x_t^T y_t' \quad \text{and} \end{aligned}$$



$P_{t-1}^T$ ,  $P_{t,t-1}^T$ ,  $P_t^T$ ,  $x_t^T$  are obtained through the Kalman smoother.

Under the Bayesian viewpoint, the EM algorithm is an iterative method to compute the mode of the posterior probability distribution. Let  $\varphi^i$  be the mode (in the iteration  $i$ ) of the observed posterior distribution  $p(\varphi | Y)$ . Let  $p(\varphi | Y, \theta)$  the complete posterior and  $p(\theta | \varphi^i, Y)$  the conditional predictive distribution of the latent process  $\theta$ , conditional to the mode at iteration  $i$  and to the data. The step E consists in computing

$$Q(\varphi, \varphi^i) = \int_{\theta} \log [p(\varphi | Y, \theta)] p(\theta | \varphi^i, Y) d\theta;$$

that is, the expectation of a  $\log [p(\varphi | Y, \theta)]$  relative to  $p(\theta | \varphi^i, Y)$ .

In the M step the function  $Q$  is maximized with respect to  $\varphi$  to obtain  $\varphi^{i+1}$ .

## Appendix B: Estimation of the Non-Observed State Vector $\theta_t$

We will utilize the KF equations and smoother that yield least squares estimators. These equations consider a weighted combination of the  $m$  possible realizations of  $F_t$  and are an extension of the classical equations of the KF. The derivation of the equations can be found in Shumway & Stoffer (1991) and uses estimators for the filtered probabilities  $\pi_{\ell}(t | t)$  given by the model.

Consider the conditional probability  $\pi_{\ell}(t | t)$  given by Bayes theorem (posterior probability):

$$\pi_{\ell}(t | t) = P(F_t = M_{\ell} | Y_t), \quad \text{with } Y_t = \{y_1, \dots, y_t\},$$

$$\pi_{\ell}(t | t) = \frac{\pi_{\ell}(t) f_{\ell}(t | t - 1)}{\sum_{i=1}^m \pi_i(t) f_i(t | t - 1)},$$

where  $f_{\ell}(t | t - 1)$  is the conditional density of  $y_t$  given  $F_t = M_{\ell}$  and the past. For the case of the multivariate normal it is assumed that

$$f_{\ell}(t | t - 1) \sim N(M_{\ell} x_t^{t-1}; \Sigma_{t\ell} = M_{\ell} P_t^{t-1} M_{\ell}' + V_t),$$

where

- (i)  $x_t^{t-1} = E(\theta_t | Y_{t-1})$  and
- (ii)  $P_t^{t-1} = \text{cov}(\theta_t | Y_{t-1})$ , obtained by the KF.

The following equations are obtained by the KF and smoother:

$$x_t^{t-1} = E(\theta_t | Y_{t-1}) = G_t x_{t-1}^{t-1}, \quad (23)$$

$$P_t^{t-1} = \text{Var}(\theta_t | Y_{t-1}) = G_t P_{t-1}^{t-1} G_t' + W_t,$$

$$x_t^t = E(\theta_t | Y_t) = x_t^{t-1} + \sum_{\ell=1}^m \pi_\ell(t|t) K_{t\ell} (y_t - M_\ell x_t^{t-1})$$

and

$$P_t^t = \text{Var}(\theta_t | Y_t) = \sum_{\ell=1}^m \pi_\ell(t|t) (I - K_{t\ell} M_\ell) P_t^{t-1},$$

where  $Y_t = \{y_1, y_2, \dots, y_t\}$ ,  $K_{t\ell} = P_t^{t-1} M_\ell' (M_\ell P_t^{t-1} M_\ell' + V_t)^{-1}$  is the filter gain,  $x_0^0 = \mu$  and  $P_0^0 = \Sigma$ .

**Smoothing equations:** for  $t = T, T-1, \dots, 1$ :

$$x_{t-1}^T = x_{t-1}^{t-1} + J_{t-1} (x_t^T - x_t^{t-1})$$

and

$$P_{t-1}^T = P_{t-1}^{t-1} + J_{t-1} (P_t^T - P_t^{t-1}) J_{t-1}',$$

where

$$J_{t-1} = P_{t-1}^{t-1} G_t' (P_t^{t-1})^{-1},$$

$$P_{t-1, t-2}^T = P_{t-1}^{t-1} J_{t-2}' + J_{t-1} (P_{t, t-1}^T - G_t P_{t-1}^{t-1}) J_{t-2}'$$

and

$$P_{T, T-1}^T = \sum_{\ell=1}^m \pi_\ell(T|T) (I - K_{T\ell} M_\ell) G_t P_{T-1}^{T-1},$$

I = identity matrix.

## Appendix C: Estimation of parameters related to the wavelets $\beta_n^{(\ell)}$ , the covariance matrix of errors, $V_t$ , $W_t$ and transition matrix $G_t$

Use the EM algorithm. Initially, compute the joint complete log-likelihood of  $\theta_0, \dots, \theta_T, y_1, \dots, y_T$  with  $(\theta_0 | D_0) \sim N(\mu, \Sigma)$  and  $D_0$  the information set at  $t = 0$ :

$$\begin{aligned} \log L &= -(1/2) \log |\Sigma| - (1/2)(\theta_0 - \mu)' \Sigma^{-1} (\theta_0 - \mu) - (T/2) \log |W_t| \\ &\quad - (1/2) \sum_{t=1}^T (\theta_t - G_t \theta_{t-1})' W_t^{-1} (\theta_t - G_t \theta_{t-1}) \\ &\quad + \sum_{t=1}^T \sum_{\ell=1}^m I(F_t = M_\ell) \log(\pi_\ell(t)) - (T/2) \log |V_t| \\ &\quad - (1/2) \sum_{t=1}^T \sum_{\ell=1}^m I(F_t = M_\ell) (y_t - F_t \theta_t)' V_t^{-1} (y_t - F_t \theta_t) \end{aligned}$$

where  $I(F_t = M_\ell)$  is an indicator function.

Since  $\log L$  depends of the series  $\theta_t, t = 1, \dots, T$  which are non-observed, we apply the EM algorithm conditional on  $y_1, \dots, y_T$ .

E STEP:

Let  $H = E(\log L | y_1, \dots, y_T)$ , that is, the conditional expectation of the function  $\log L$  and let  $E(I(F_t = M_\ell) | Y_T) = \pi_\ell(t | T)$ .

Then

$$\begin{aligned} H &= -(1/2) \log |\Sigma| - (1/2) \text{tr}[\Sigma^{-1} (P_0^T + (x_0^T - \mu)(x_0^T - \mu)')] \\ &\quad - (T/2) \log |W_t| - (1/2) \text{tr}[W_t^{-1} (C - BG_t' - G_t B' + G_t A G_t')] \\ &\quad + \sum_{t=1}^T \sum_{\ell=1}^m \pi_\ell(t | T) \log(\pi_\ell(t)) - (T/2) \log |V_t| \\ &\quad - (1/2) \text{tr}[\sum_{t=1}^T \sum_{\ell=1}^m V_t^{-1} \pi_\ell(t | T) ((y_t - M_\ell x_t^T)(y_t - M_\ell x_t^T)' + M_\ell P_t^T M_\ell')], \end{aligned}$$

where  $A = \sum_{t=1}^T (P_{t-1}^T + x_{t-1}^T x_{t-1}^{T'})$ ,  $B = \sum_{t=1}^T (P_{t,t-1}^T + x_t^T x_{t-1}^{T'})$ ,  $C = \sum_{t=1}^T (P_t^T + x_t^T x_t^{T'})$  and  $x_t^T, P_t^T$ , and  $P_{t,t-1}^T$  are obtained by the Kalman smoother.

Shumway & Stoffer (1991) use what they call “pseudo-EM” due to the replacement of  $\pi_\ell(t|T)$  by  $\pi_\ell(t|t)$ . The backwards recursion of the filtered probabilities involves an integration of mixtures of normals and this may be difficult. The authors still suggest the utilization of Monte Carlo integration techniques as the Gibbs sampler, though they use the mentioned approximation.

M STEP:

Maximize  $H$  with respect to  $G_t, V_t, W_t$  and  $\beta_r^{(\ell)}$  which is equivalent to solving the following equations, for  $t = 1, \dots, T$ :

$$\frac{\partial H}{\partial G_t} = 0, \quad \frac{\partial H}{\partial W_t} = 0, \quad \frac{\partial H}{\partial V_t} = 0 \quad \text{and} \quad \frac{\partial H}{\partial \beta_r^{(\ell)}} = 0.$$

(i) Estimate of  $G_t$ :

$$\hat{G}_t = BA^{-1}.$$

(ii) Estimate of  $W_t$ :

$$\hat{W}_t = (1/T)(C - B\hat{G}_t' - \hat{G}_t B' + \hat{G}_t A \hat{G}_t').$$

(iii) Estimate of  $V_t$ :

$$\hat{V}_t = (1/T) \sum_{t=1}^T \sum_{\ell=1}^m \pi_\ell(t|t) [(y_t - M_\ell x_t^T)(y_t - M_\ell x_t^T)' + M_\ell P_t^T M_\ell'].$$

(iv) Estimation of  $\beta_n^{(\ell)}$ :

To Maximize  $H$  with respect  $\pi_\ell(t)$ , consider

$$\pi_\ell(t) = \frac{\exp(\sum_{n=1}^s \beta_n^{(\ell)} \psi_n(t/T))}{1 + \sum_{i=1}^{m-1} \exp(\sum_{n=1}^s \beta_n^{(i)} \psi_n(t/T))}, \quad \ell = 1, \dots, m-1$$

and

$$\pi_m(t) = \frac{1}{1 + \sum_{i=1}^{m-1} \exp(\sum_{n=1}^s \beta_n^{(i)} \psi_n(t/T))}.$$

Then

$$\log(\pi_\ell(t)) = \left[ \sum_{n=1}^s \beta_n^{(\ell)} \psi_n(t/T) \right] - \log\left(1 + \sum_{i=1}^{m-1} \exp\left\{ \sum_{n=1}^s \beta_n^{(i)} \psi_n(t/T) \right\}\right),$$

$\ell = 1, \dots, m-1$  and

$$\log(\pi_m(t)) = -\log\left[1 + \sum_{i=1}^{m-1} \exp\left(\sum_{n=1}^s \beta_n^{(i)} \psi_n(t/T)\right)\right].$$

The partial derivative of  $H$  with respect to  $\beta_r^{(\ell)}$ , for  $\ell = 1, \dots, m-1$  and  $r = 1, \dots, s$ , is given by

$$\frac{\partial H}{\partial \beta_r^{(\ell)}} = \sum_{t=1}^T \left[ \frac{\partial}{\partial \beta_r^{(\ell)}} \sum_{\ell=1}^m \pi_\ell(t|t) \log(\pi_\ell(t)) \right],$$

where

$$\begin{aligned} \sum_{t=1}^T \sum_{\ell=1}^m \pi_\ell(t|t) \log(\pi_\ell(t)) &= \sum_{t=1}^T \left[ \sum_{\ell=1}^m \pi_\ell(t|t) \sum_{n=1}^s (\beta_n^{(\ell)} \psi_n(t/T)) - \right. \\ &\quad \left. \log\left(1 + \sum_{\ell=1}^{m-1} \exp\left(\sum_{n=1}^s \beta_n^{(\ell)} \psi_n(t/T)\right) \sum_{\ell=1}^m \pi_\ell(t|t)\right) \right] \end{aligned}$$

and

$$\frac{\partial}{\partial \beta_r^{(\ell)}} \sum_{\ell=1}^m \pi_\ell(t|t) \log(\pi_\ell(t)) = \pi_\ell(t|t) \psi_r(t/T) - \left(1 + \sum_{\ell=1}^{m-1} \exp\left(\sum_{n=1}^s \beta_n^{(\ell)} \psi_n(t/T)\right)\right)^{-1}$$

$$\times \exp\left(\sum_{n=1}^s \beta_n^{(\ell)} \psi_n(t/T)\right) \psi_r(t/T).$$

It is necessary to solve a non-linear system with  $(m-1)s$  equations for  $\ell = 1, \dots, m-1$  and  $r = 1, \dots, s$ , where

$$\hat{\beta}^{(\ell)} = \begin{pmatrix} \hat{\beta}_1^{(\ell)} \\ \hat{\beta}_2^{(\ell)} \\ \vdots \\ \hat{\beta}_s^{(\ell)} \end{pmatrix}, \quad X_t = \begin{pmatrix} \psi_1(t/T) \\ \psi_2(t/T) \\ \vdots \\ \psi_s(t/T) \end{pmatrix},$$

$$\sum_{t=1}^T \left[ \frac{(\exp(X_t' \hat{\beta}^{(\ell)})) \psi_r(t/T)}{(1 + \sum_{i=1}^{m-1} \exp(X_t' \hat{\beta}^{(i)}))} - \pi_\ell(t|t) \psi_r(t/T) \right] = 0.$$

In order to solve this system we have to use a numerical procedure.

**Special Case:**  $m = 2, \ell = 1$ .

$$\pi_1(t) = \pi(t) = \frac{\exp(\sum_{n=1}^s \beta_n \psi_n(t/T))}{1 + \exp(\sum_{n=1}^s \beta_n \psi_n(t/T))},$$

$$\begin{aligned} \pi_2(t) &= 1 - \pi_1(t) \\ &= 1 - \frac{\exp(\sum_{n=1}^s \beta_n \psi_n(t/T))}{1 + \exp(\sum_{n=1}^s \beta_n \psi_n(t/T))} \\ &= \frac{1}{1 + \exp(\sum_{n=1}^s \beta_n \psi_n(t/T))}, \end{aligned}$$

$$\pi_1(t|t) = \pi(t|t) = \frac{\pi(t) f_1(t|t-1)}{\pi(t)(f_1(t|t-1) - f_2(t|t-1)) + f_2(t|t-1)},$$

and

$$\pi_2(t|t) = \frac{(1 - \pi(t)) f_2(t|t-1)}{\pi(t)(f_1(t|t-1) - f_2(t|t-1)) + f_2(t|t-1)}.$$

The system to be solved is:

$$\sum_{t=1}^T \left[ \frac{\exp(\sum_{n=1}^s \beta_n \psi_n(t/T)) \psi_r(t/T)}{1 + \exp(\sum_{n=1}^s \beta_n \psi_n(t/T))} - \pi(t|t) \psi_r(t/T) \right] = 0,$$

$$r = 1, \dots, s.$$

Considering  $X_t$  as before and  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_s)'$ , we have

$$\sum_{t=1}^T \left( \frac{\exp(X_t' \hat{\beta}) \psi_1(t/T)}{1 + \exp(X_t' \hat{\beta})} \right) = \sum_{t=1}^T \pi(t|t) \psi_1(t/T),$$

...

...

$$\sum_{t=1}^T \left( \frac{\exp(X_t' \hat{\beta}) \psi_s(t/T)}{1 + \exp(X_t' \hat{\beta})} \right) = \sum_{t=1}^T \pi(t|t) \psi_s(t/T).$$