D S E

Dipartimento   **Scienze**   **Economiche**

Working Paper

Department of Economics

Ca' Foscari University of Venice

Monica Billio

Roberto Casarin

Domenico Sartore

Bayesian Inference on Dynamic Models with Latent Factors

Bayesian Inference on Dynamic Models with Latent Factors

Monica Billio
University of Venice

Roberto Casarin
University of Brescia

Domenico Sartore
University of Venice

Abstract

In time series analysis, latent factors are often introduced to model the heterogeneous time evolution of the observed processes. The presence of unobserved components makes the maximum likelihood estimation method more difficult to apply. A Bayesian approach can sometimes be preferable since it permits to treat general state space models and makes easier the simulation based approach to parameters estimation and latent factors filtering. The paper examines economic time series models in a Bayesian perspective focusing, through some examples, on the extraction of the business cycle components. We briefly review some general univariate Bayesian dynamic models and discuss the simulation based techniques, such as Gibbs sampling, adaptive importance sampling and finally suggest the use of the particle filter, for parameter estimation and latent factor extraction.

Address for correspondence:
Monica Billio
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
Phone: (++39) 041 2349170
Fax: (++39) 041 2349176
e-mail: billio@unive.it

1   Introduction

The analysis of dynamic phenomena is a common problem to many fields like engineering, physics, biology, statistics and econometrics. A time varying system can be represented through a *dynamic model*, defined by an observable component and an unobservable state. The hidden state vector represents the desired information that we want to extrapolate from the observations.

In the literature on business cycle analysis, dynamic models are used to capture two well known features of the economic cycle: comovement and asymmetry. Asymmetry denotes an heterogeneous dynamics of the economic variable. If the behavior of the economic time series depends on the phase of the economic cycle, then asymmetry arises. To capture asymmetry Goldfeld and Quandt (1973) introduced Markov Switching (MS) models for serially uncorrelated data, while Hamilton (1989) applies MS to serially correlated time series. In these models parameters are allowed to depend on the hidden state, representing the economic cycle. The state usually assumes two values, interpreted as positive growth trend and negative growth trend.

All above cited approaches and in particular the original work of Hamilton (1989), have been successively extended in many directions. Kim (1994) applies MS to dynamic linear model, Kim and Nelson (1999) analyze general MS dynamic models and provide Bayesian inference tools together with Markov chain Monte Carlo (MCMC) simulation techniques. Kim and Murray (2002) and Anas and Ferrara (2002) suggest to divide the business cycle in three phases: recession, high-growth and normal-growth. Another kind of extension concerns the duration of the phases of the business cycle, for example Sichel (1991), Watson (1994) and Diebold and Rudebusch (1996) assume that the transition probability of the Markov chain depends on the duration of the current phase of the cycle. Finally, multivariate extensions have been suggested by Diebold and Rudebusch (1996) and Krolzig (1997).

The seminal work of Kalman (1960) and Kalman and Bucy (1960) introduces filtering techniques (Kalman-Bucy filter) for continuous valued, linear and Gaussian dynamic systems. Harvey (1989) extensively studies state space representation of dynamic models for time series and treats the use of Kalman filter for states and parameters estimation. Hamilton (1989) introduces a filter (Hamilton-Kitagawa filter) for discrete time and discrete valued dynamic system with a finite number of state. Bauwens, Lubrano and Richard (1999) compare maximum likelihood inference with Bayesian inference on static and dynamic econometric models. Harrison and West (1997) treat the problem of the dynamic model estimation in a Bayesian perspective. Kim and Nelson (1999) analyze Monte Carlo simulation methods for non-linear discrete valued model (MS models). Recently, Durbin and Koopman (2001) propose an updated review on MCMC methods for the estimation of general dynamic models, with both a Bayesian and a maximum likelihood approach.

The main aim of this paper is to suggest the use of sequential simulation methods for filtering and smoothing in business cycle dynamic models. These methods have been recently developed to overcome some problems of the traditional MCMC methods. As pointed out by Liu and Chen (1998), the Gibbs sampler is less attractive when considering on-line data processing. Furthermore, Gibbs sampler may be inefficient when simulated states are very sticky and the sampler has difficulties to move in the state space. In these situations, the use of sequential Monte Carlo techniques and in particular of particle fil-

ter algorithms may result more efficient. Doucet, Freitas and Gordon (2001) provide the state of the art on sequential Monte Carlo methods and discuss both applications and theoretical convergence results for these algorithms, with special attention to particle filters.

The paper is structured as follows. Section 2 introduces the general representation of a dynamic model in a Bayesian framework and deals with conditionally normal linear models. Section 3 reviews simulation based methods. In particular section 3.1 reviews MCMC methods, section 3.2 presents an adaptive importance sampling algorithm. Section 3.3 discusses particle filter algorithms. Finally, section 4 provides an application of the particle filter to business cycle models and section 5 concludes.

2   Bayesian Dynamic Models

We consider a quite general formulation of a probabilistic *dynamic model* and review some fundamental relations for Bayesian inference on it. This definition includes time series models analyzed in Kalman (1960), Hamilton (1989), Harrison and West (1997) and in Doucet, Freitas and Gordon (2001).

We denote $\{\mathbf{x}_t; t \in \mathbb{N}\}$, $\mathbf{x}_t \in \mathcal{X}$, the hidden state vectors of the system, $\{\mathbf{y}_t; t \in \mathbb{N}_0\}$, $\mathbf{y}_t \in \mathcal{Y}$, the observable variables and $\theta \in \Theta$ the parameter vector. We assume that the state space, observation space and parameter space are $\mathcal{X} \subset \mathbb{R}^{n_x}$, $\mathcal{Y} \subset \mathbb{R}^{n_y}$ and $\Theta \subset \mathbb{R}^{n_\theta}$, respectively. n_x, n_y and n_θ represent the dimensions of the state, observable variable and parameter vectors.

This general Bayesian state space representation accounts also for nonlinear and non-Gaussian components and is given by an *initial distribution* $p(\mathbf{x}_0|\theta)$, a *measurement density* $p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)$ and a *transition density* $p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}, \theta)$,

$$\mathbf{y}_t \sim p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta) \tag{1}$$

$$\mathbf{x}_t \sim p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}, \theta) \tag{2}$$

$$x_0 \sim p(\mathbf{x}_0|\theta), \quad \text{with } t = 1, \dots, T \tag{3}$$

where $p(\mathbf{x}_0|\theta)$ can be interpreted as the prior distribution on the initial state of the system. By $\mathbf{x}_{0:t} \stackrel{\Delta}{=} (\mathbf{x}_0, \dots, \mathbf{x}_t)$ we denote the collection of vectors up to time t, while by $\mathbf{x}_{-t} \stackrel{\Delta}{=} (\mathbf{x}_0, \dots, \mathbf{x}_{t-1}, \mathbf{x}_{t+1}, \dots, \mathbf{x}_T)$ we denote the collection of all the vectors without the t-th element.

If the transition density depends on the past only through the last value of the hidden state vector, the dynamic model is defined *Markovian* of the first-order, i.e.

$$(\mathbf{y}_t|\mathbf{x}_t) \sim p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta) \tag{4}$$

$$(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta) \tag{5}$$

$$x_0 \sim p(\mathbf{x}_0|\theta), \quad \text{with } t = 1, \dots, T. \tag{6}$$

To assume the first-order Markov property is not restrictive because a Markov model of order p can always be rewritten as a first-order Markovian model.

Let us now see the three main issues necessary for inference: filtering, prediction and smoothing.

2.1 State Estimation

We are interested in estimating the density $p(\mathbf{x}_t|\mathbf{y}_{1:s}, \theta)$ when parameters are known. If $t = s$ the density of interest is called *filtering density*, if $t < s$ it is called *smoothing density* and if $t > s$ it is called *prediction density*. We assume that at time t the density $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}, \theta)$ is known[2]. By applying the Chapman-Kolmogorov transition density, we obtain the one-step-ahead *prediction density*

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}, \theta)d\mathbf{x}_{t-1} \tag{7}$$

When a new observation $\mathbf{y}_t$ becomes available, using the Bayes theorem, it is possible to update the prediction density and to filter the current state of the system. The *filtering density* is:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta) = \frac{p(\mathbf{y}_t, \mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta)} = \frac{p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)}{\int_{\mathcal{X}} p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)d\mathbf{x}_t} \tag{8}$$

where $p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)$ is the prediction density obtained at the previous step.

At each date t, it is possible to determine the *K-steps-ahead prediction density* of the state vector, conditional on the available information $\mathbf{y}_{1:t}$. It can be evaluated iteratively, as follows:

first step $\qquad\qquad p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_{1:t}, \theta)p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta)d\mathbf{x}_t \qquad$ (9)

k-th step $\qquad p(\mathbf{x}_{t+k}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_{t+k}|\mathbf{x}_{t+k-1}, \mathbf{y}_{1:t}, \theta)p(\mathbf{x}_{t+k-1}|\mathbf{y}_{1:t}, \theta)d\mathbf{x}_{t+k-1} \qquad$ (10)

where $\quad p(\mathbf{x}_{t+k}|\mathbf{x}_{t+k-1}, \mathbf{y}_{1:t}, \theta) = \int_{\mathcal{Y}^{k-1}} p(\mathbf{x}_{t+k}|\mathbf{x}_{t+k-1}, \mathbf{y}_{1:t+k-1}, \theta)p(d\mathbf{y}_{t+1:t+k-1}|\mathbf{y}_{1:t}, \theta)$

and $\mathcal{Y}^k = \otimes_{i=1}^k \mathcal{Y}_i$ is the k-times Cartesian product of the state space, with $k = 1, \ldots, K$. Similarly, the *K-steps-ahead prediction density* of the observable variable $\mathbf{y}_{t+K}$ conditional on the available information is determined as follows:

$$p(\mathbf{y}_{t+K}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{Y}} p(\mathbf{y}_{t+K}|\mathbf{x}_{t+K}, \mathbf{y}_{1:t+K-1}, \theta)p(d\mathbf{y}_{t+1:t+K-1}|\mathbf{y}_{1:t}, \theta)p(d\mathbf{x}_{t+K}|\mathbf{y}_{1:t}, \theta) \tag{11}$$

With general dynamics, due to the high number of integrals that must be solved, previous densities may be very difficult to evaluate. From a numerical point of view, simulation methods, like MCMC algorithms or particle filters, allow us to overcome these difficulties. From a modelling point of view to obtain analytical relations we need to introduce some simplifying hypothesis on the dynamics of the variables. For example if we assume that the evolution of the dynamic model does not depend on the past values of the observable variable $\mathbf{y}_{1:t}$, then equations (4), (5) and (6) become:

$$(\mathbf{y}_t|\mathbf{x}_t) \sim p(\mathbf{y}_t|\mathbf{x}_t, \theta) \tag{12}$$

$$(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}, \theta) \tag{13}$$

$$x_0 \sim p(\mathbf{x}_0|\theta), \quad \text{with } t = 1, \ldots, T. \tag{14}$$

---

[2]Observe that if $t = 1$ the density $p(\mathbf{x}_0|\mathbf{y}_0, \theta) = p(\mathbf{x}_0|\theta)$ is the initial distribution of the dynamic model.
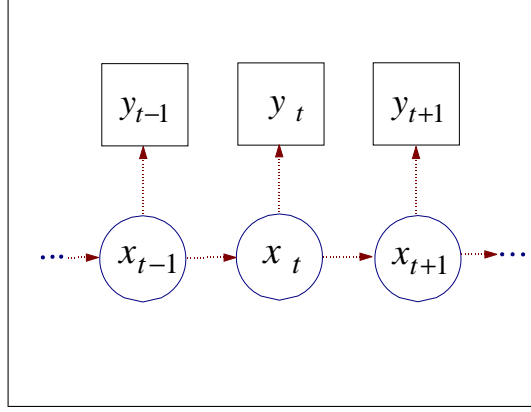
Figure 1: The causality structure of a Markov dynamic model with hidden states. A
indicates the variable is known, while a circle indicates a hidden variable.

The causality structure of this model is represented through the Directed Acyclic Graph
(DAG) of Fig. 1. Under the previous assumptions the filtering and prediction densities
simplify as follows:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1},\theta) \quad = \quad \int_{\mathcal{X}} p(\mathbf{x}_t|\mathbf{x}_{t-1},\theta)p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1},\theta)d\mathbf{x}_{t-1} \tag{15}$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t},\theta) \quad = \quad \frac{p(\mathbf{y}_t|\mathbf{x}_t,\theta)p(\mathbf{x}_t|\mathbf{y}_{1:t-1},\theta)}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1},\theta)} \tag{16}$$

$$p(\mathbf{x}_{t+K}|\mathbf{y}_{1:t},\theta) \quad = \quad \int_{\mathcal{X}} p(\mathbf{x}_{t+K}|\mathbf{x}_{t+K-1},\theta)p(\mathbf{x}_{t+K-1}|\mathbf{y}_{1:t},\theta)d\mathbf{x}_{t+K-1} \tag{17}$$

$$p(\mathbf{y}_{t+K}|\mathbf{y}_{1:t},\theta) \quad = \quad \int_{\mathcal{X}} p(\mathbf{y}_{t+K}|\mathbf{x}_{t+K},\theta)p(\mathbf{x}_{t+K}|\mathbf{y}_{1:t},\theta)d\mathbf{x}_{t+K}. \tag{18}$$

We conclude this section with two important recursive relations. Both these relations
can be proved starting from the definition of joint smoothing density and assuming that
the Markov property holds. The first relation is the sequential filtering equation:

$$p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T},\theta) = p(\mathbf{x}_{0:T-1}|\mathbf{y}_{1:T-1},\theta)\frac{p(\mathbf{y}_T|\mathbf{x}_T,\theta)p(\mathbf{x}_T|\mathbf{x}_{T-1},\theta)}{p(\mathbf{y}_T|\mathbf{y}_{1:T-1},\theta)}. \tag{19}$$

which is particularly useful when processing data sequentially and it is fundamental in
implementing particle filter algorithms. The second relation provides the factorization of
the *smoothing density* of the state vectors given the information available at time T:

$$p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T},\theta) = p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta) \prod_{t=0}^{T-1} p(\mathbf{x}_t|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\theta). \tag{20}$$

Note that the density $p(\mathbf{x}_t|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\theta)$, which appears in the joint smoothing density,
can be represented through the filtering and the prediction densities[3]:

$$p(\mathbf{x}_t|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\theta) = \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}_{1:t},\theta)p(\mathbf{x}_t|\mathbf{y}_{1:t},\theta)}{p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t},\theta)}. \tag{21}$$

---

[3]For proofs see Billio, Casarin and Sartore (2004).

This factorization of the smoothing density is also relevant when inference is carried out through simulation methods[4].

We introduce now an important class of dynamic models, which does not admit a tractable analytical representation of the filtering, prediction and smoothing densities. These are *conditional normal linear models* and are widely used in business cycle analysis (see Kim and Nelson (1999)).

2.2   Conditionally Gaussian Linear Models

Usually, models used in business cycle analysis belong to the class of the conditionally normal dynamic linear models, defined as follows:

$$
\begin{aligned}
\mathbf{y}_t &= F(s_t)\mathbf{x}_t + V(s_t)\epsilon_t & \epsilon_t \sim N(0, I) \\
\mathbf{x}_{t+1} &= G(s_t)\mathbf{x}_t + W(s_t)\eta_t & \eta_t \sim N(0, I)
\end{aligned}
\tag{22}
$$

where ϵ_t is independent of η_t and s_t is a sequence of random variables. Harrison and West (1997) call this model multi-process model: in their classification if $s_t = s_{t-1} = s$, $\forall t$ the model is a multi-process of the first kind, while if s_t is a stochastic process, the model is a multi-process of second kind. Note that if s_t is a discrete time and finite state Markov chain with known transition probabilities, the model is also called *jump Markov linear system* or *Markov switching linear model* with parameters evolving over time.

Stochastic latent factor model with Markov switching - Economic phases can be represented through a Markov switching hidden process. Let y_t be the observable variable and x_t the latent factor. The switching model is:

$$
\begin{aligned}
y_t &= \alpha x_t + \sigma_\epsilon \epsilon_t & \epsilon_t &\sim N(0, 1) & \text{(23)} \\
x_{t+1} &= \mu(s_{t+1}) + \rho\, x_t + \sigma_\eta \eta_{t+1} & \eta_{t+1} &\sim N(0, 1) & \text{(24)} \\
s_t &\sim Markov(\mathbb{P}), \quad \text{with} & s_t &\in \{0, 1\} & \text{(25)}
\end{aligned}
$$

where $\mu(s_t) = \mu + \nu s_t$, $\mathbb{P}$ is the transition matrix and ε_t is independent of $\eta_t\, \forall t$.

This kind of model can be found in Kim and Nelson (1999). The absence of analytical filtering densities makes Bayesian simulation based inference a possible solution to the filtering problem. Fig. 2 exhibits simulation paths of $1,000$ observations[5] of the Markov switching process, the latent factor and the observable variable, respectively.

3   Simulation Based Filtering

In the following, we focus on the Bayesian approach and on simulation based methods for nonlinear and non-Gaussian models. First, MCMC methods are reviewed, then some

---

[4]See for example the multi-move Gibbs sampler of Carter and Köhn (1994) and the particle filter algorithms. Only in some well known cases, these filtering densities admit an analytical form. For the normal linear dynamic model, filtering and smoothing density are given by the Kalman filter. See also Harrison and West (1997) for a Bayesian representation of the Kalman filter.

[5]We use parameters estimated in Kim and Nelson (1999) to simulate the MS model.
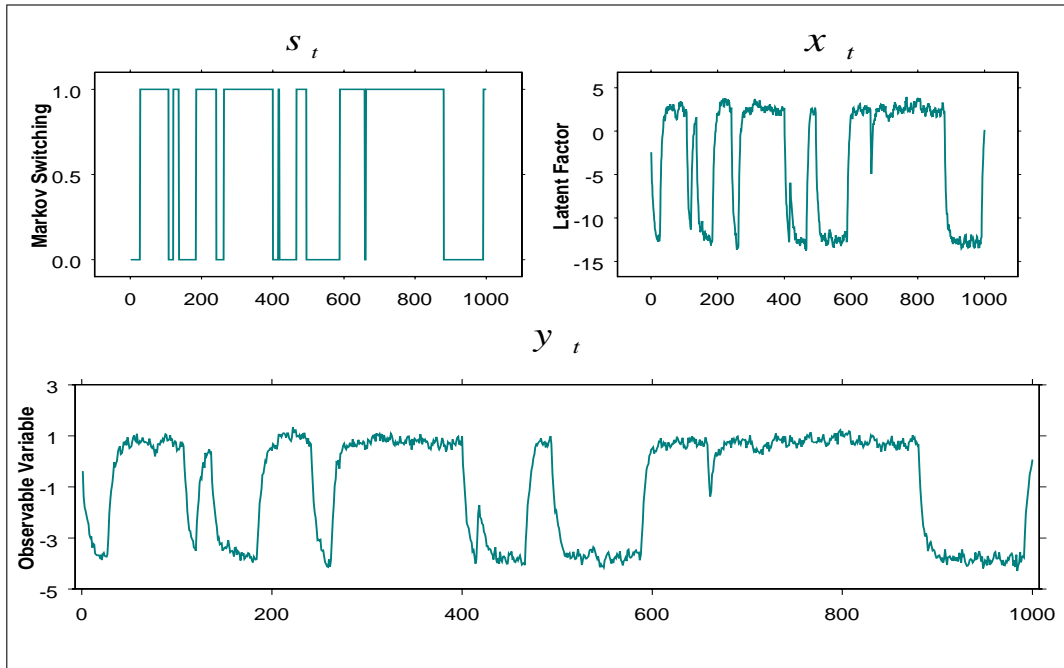
Figure 2: Simulation from the Markov switching stochastic trend model given in Example 2.2. We set parameters to be $\alpha = 0.3, \sigma_\varepsilon = 0.1, \rho = 0.8, \mu_0 = -2.5, \mu_1 = 0.5, \sigma_\eta = 0.1, p_{11} = 0.97, p_{22} = 0.99$.

basic sequential Monte Carlo simulation methods are introduced. Mainly, we refer to sequential importance sampling algorithm and to more advanced sequential Monte Carlo algorithms called Particle Filters. Finally we investigate the problem of estimation of the parameter vector both in a Bayesian MCMC based approach and in a sequential data-processing approach.

3.1 The Gibbs Sampler

In previous sections we examine some estimation algorithms for filtering, predicting and smoothing the state vector of a quite general probabilistic dynamic model. To show how Gibbs sampling applies, we consider the following dynamic model:

$$(\mathbf{y}_t|\mathbf{x}_t) \sim p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta) \tag{26}$$

$$(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta) \tag{27}$$

$$x_0 \sim p(\mathbf{x}_0|\theta) \tag{28}$$

$$\theta \sim p(\theta), \quad \text{with } t = 1, \ldots, T. \tag{29}$$

The estimation problem is solved in a Bayesian perspective by evaluating the mean of the joint posterior density of the state and parameter vectors $p(\mathbf{x}_{0:T}, \theta|\mathbf{y}_{1:T})$. Tanner and Wong (1987) motivates this solution by the *data augmentation principle*, which consists in considering the hidden state vectors as nuisance parameters.

If an analytical evaluation of the posterior mean is not possible, then simulation methods and in particular MCMC apply. The most simple solution is to implement a *single-move*

Gibbs sampler (see Carlin, Polson and Stoffer (1992) and Harrison and West (1997)). This method generates the states one at a time using the Markov property of the dynamic model and conditioning on the neighboring states.

The conditional posterior distributions of the parameter and state vectors are:

$$p(\theta|\mathbf{x}_{0:T}, \mathbf{y}_{1:T}) \propto p(\theta)p(\mathbf{x}_0|\theta) \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta) \qquad (30)$$

$$p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T}, \theta) \propto p(\mathbf{x}_0|\theta) \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta). \qquad (31)$$

The basic idea of the Gibbs sampler is to simulate sequentially from the parameter posterior (*parameter simulation step*) in equation (30), conditionally on the state vectors, and from the state posterior (*data augmentation step*) in equation (31), conditionally on the parameter vector simulated at the previous step. When conditional distributions cannot be directly simulated, the corresponding steps in the Gibbs algorithm can be replaced by Metropolis-Hastings steps. The resulting algorithms are called hybrid sampling algorithms and they are validated in Tierney (1994).

The single-move Gibbs sampler for the state vectors is then obtained by drawing each state vector conditionally on the other simulated state vectors and on the simulated parameter vector.

Algorithm 1 - *Single-Move Gibbs Sampler* -

(i) Simulate $\theta^{(i)}$ through a generic Gibbs sampler;

(ii) Given $\theta^{(i)}$ and $\mathbf{x}_{0:T}^{(i)}$, simulate state vectors as follows:

1. $\mathbf{x}_0^{(i+1)} \sim p(\mathbf{x}_0|\mathbf{x}_{2:T}^{(i)}, \mathbf{y}_{1:T}, \theta^{(i+1)})$

2. $\mathbf{x}_1^{(i+1)} \sim p(\mathbf{x}_1|\mathbf{x}_0^{(i+1)}, \mathbf{x}_{2:T}^{(i)}, \mathbf{y}_{1:T}, \theta^{(i+1)})$

3. ...

4. $\mathbf{x}_T^{(i+1)} \sim p(\mathbf{x}_T|\mathbf{x}_{0:T-1}^{(i+1)}, \mathbf{y}_{1:T}, \theta^{(i+1)})$

The single-move algorithm can be implemented for general dynamic models. Moreover, note that the dynamic model given in equations (26)-(29) satisfies the Markov property. In this case the full posterior density of the state vector, given in the single-move Gibbs sampler (see the Algorithm 1), is simply:

$$p(\mathbf{x}_t|\mathbf{x}_{-t}, \mathbf{y}_{1:T}, \theta) \propto p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_{1:t}, \theta) \qquad (32)$$

and the implementation of the algorithm becomes easier. For a proof see Billio, Casarin and Sartore (2004).

Although the simplification due to the Markov property makes the single-move Gibbs sampler easier to implement, some problems arise. In particular, the Markovian dependence between neighboring states generates correlation between outputs of the Gibbs

sampler and origins slower convergence to the posterior distribution (see Carter and Köhn (1994)). As consequence, if an adaptive importance sampling is carried out by running parallel single-move Gibbs samplers, the number of replication before convergence of the parameter estimates could be very high.

A general method to solve this autocorrelation problem in the output of the Gibbs sampler is to group parameters (or states) and to simulate them simultaneously. This idea has been independently applied by Carter and Köhn (1994) and by Frühwirth-Schnatter (1994) to dynamic models and the resulting algorithm is the multi-move Gibbs sampler. The main idea of this method is to generate simultaneously all the state vectors using analytical filtering and smoothing relations. Their approach is less general than that of Carlin, Polson and Stoffer (1992), but for linear dynamic models with Gaussian mixture innovations in the observation equation, it is more efficient. In particular the multi-move Gibbs sampler has a faster convergence to the posterior distribution and the posterior moment estimates have smaller variance. These results are supported theoretically by Liu, Wong and Kong (1994, 1995) and Müller (1992), who show that generating variables simultaneously produces faster convergence[6]. Furthermore Frühwirth-Schnatter (1994) shows how the use of the multi-move Gibbs sampler improves the convergence rate of an adaptive importance sampling algorithm and makes a comparison with a set of parallel single-move Gibbs samplers.

The implementation of the multi-move Gibbs sampler depends on the availability of the analytical form of filtering and smoothing densities. We give here a general representation of the algorithm, but its implementation is strictly related to the specific dynamic model.

Algorithm 2 - *Multi-Move Gibbs Sampler* -

(i) Simulate $\theta^{(i)}$ through a generic Gibbs sampler;

(ii) Given $\theta^{(i)}$ and $\mathbf{x}_{0:T}^{(i)}$, run analytical filtering relations to estimate prediction and filtering densities for each $t = 0, \ldots, T$

   1. $\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta^{(i+1)})$

   2. $\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta^{(i+1)})$

(iii) Simulate state vectors by means of the recursive factorization of the smoothing density

   3. $\mathbf{x}_T^{(i+1)} \sim p(\mathbf{x}_T|\mathbf{y}_{1:T}, \theta^{(i+1)})$

   4. $\mathbf{x}_{T-1}^{(i+1)} \sim p(\mathbf{x}_{T-1}|\mathbf{x}_T^{(i+1)}, \mathbf{y}_{1:T-1}, \theta^{(i+1)})$

   5. $\ldots$

   6. $\mathbf{x}_1^{(i+1)} \sim p(\mathbf{x}_1|\mathbf{x}_2^{(i+1)}, \mathbf{y}_1, \theta^{(i+1)})$

---

[6]The idea of grouping parameters (or hidden states) when simulating is now commonly used in Bayesian inference on stochastic models with latent factors (see for example Shephard (1994) and Shephard and Pitt (1997), who discuss multi-move MCMC algorithms for non-Gaussian time series models).

The algorithm has been derived trough the recursive smoothing relation given in equation (20). Moreover, at each simulation step the posterior density is obtained by means of the estimated prediction and filtering densities. By applying the fundamental relation given in equation (21) we obtain

$$p(\mathbf{x}_t|\mathbf{x}_{t+1}^{(i+1)}, \mathbf{y}_{1:t}, \theta^{(i+1)}) = \frac{p(\mathbf{x}_{t+1}^{(i+1)}|\mathbf{x}_t, \theta^{(i+1)})\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta^{(i+1)})}{\hat{p}(\mathbf{x}_{t+1}^{(i+1)}|\mathbf{y}_{1:t}, \theta^{(i+1)})} \tag{33}$$

We stress once more that the multi-move Gibbs sampler does not easily apply to nonlinear and non-Gaussian models. Thus in a MCMC approach, the single-move Gibbs sampler remains the only numerical solution to the estimation problem.
A sequential sampling approach represents another possible solution to this problem and sequential Monte Carlo algorithms allow us to make inference on general dynamic models.

3.2   Adaptive Importance Sampling

The adaptive sequential importance sampling scheme is a sequential stochastic simulation method which adapts progressively to the posterior distribution. It also uses the information contained in the samples, which are simulated at the previous steps. The adaptation mechanism is based on the discrete posterior approximation and on the kernel density reconstruction of the prior and posterior densities. West (1992) proposed this technique to estimate parameters of static models. West (1993) and West and Harrison (1997) successively extended the method to estimate parameters and states of dynamic models.
The first key idea is to use importance sampling (see Casella and Robert (1999)) to obtain a weighted random grid of evaluation points in the state space. Let $\{\mathbf{x}_t^i, w_t^i\}_{t=1}^{n_t}$ be a sample drawn from the posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta)$ through an importance density g_t. The prediction density, given in equation (15), can be approximated as follows:

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) \approx \sum_{i=1}^{n_t} w_t^i p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta) \tag{34}$$

Algorithm 3   - *Adaptive Sequential Importance Sampling -*

Given a weighted random sample $\{\mathbf{x}_t^i, w_t^i\}_{t=1}^{n_t}$, for $i = 1, \ldots, n_t$,

1. *Simulate $\tilde{\mathbf{x}}_{t+1}^i \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta)$*

2. *Calculate $m_t = \sum_{i=1}^{n_t} w_t^i \tilde{\mathbf{x}}_{t+1}^i$, $V_t = \sum_{i=1}^{n_t} w_t^i (\tilde{\mathbf{x}}_{t+1}^i - m_t)'(\tilde{\mathbf{x}}_{t+1}^i - m_t)$*

3. *Generate from the Gaussian kernel $\mathbf{x}_{t+1}^i \sim \sum_{i=1}^{n_t} w_t^i N(\mathbf{x}_{t+1}|(m_t a + \mathbf{x}_t^i(1-a)), h^2 V_t)$*

4. *Update the weights $w_{t+1}^i \propto w_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i)p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i)}{N(\mathbf{x}_{t+1}^i|(m_t a + (1-a)\mathbf{x}_t^i), h^2 V_t)}$*

The second key idea is to propagate points of the stochastic grid by means of the transition density and to build a smoothed approximation of the prior density. This approximation is obtained through a kernel density estimation. West (1993) suggested to use Gaussian or Student-t kernels due to their flexibility in approximating other densities. For example, the Gaussian kernel reconstruction is:

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) \approx \sum_{i=1}^{n_t} w_t^i N(\mathbf{x}_{t+1}|m_t a + \mathbf{x}_t^i(1-a), h^2 V_t) \tag{35}$$

The final step of the algorithm consists in updating the prior density and in producing a random sample, $\{\mathbf{x}_{t+1}^i, w_{t+1}^i\}_{i=1}^{n_{t+1}}$, from the resulting posterior density. The sample is obtained by using the kernel density estimate as importance density.

The main advantage of this algorithm relies in the smoothed reconstruction of the prior density. This kernel density estimate of the prior allows to obtain adaptive importance densities and to avoid the information loss, which comes from cumulating numerical approximation over time. However, adaptive importance sampling requires the calibration of parameters a and h, which determine the behavior of the kernel density estimate. The choice of these shrinkage parameters influences the convergence of the algorithm and heavily depends on the complexity of the model.

Adaptive importance sampling belongs to a more general class of sequential simulation algorithms, called particle filters, which will be reviewed in the next section.

3.3   Particle Filters

In the following we focus on *Particle filters*, also referred in the literature as *Bootstrap filters*, *Interacting particle filters*, *Condensation algorithms* or *Monte Carlo filters* and on the estimation of the states[7].

Assume that the parameter vector, θ is known. At each step $t+1$, as a new observation $\mathbf{y}_{t+1}$ becomes available, we are interested in predicting and filtering the hidden variables and the parameters. In particular, we want to approximate the prediction and filtering densities given in equations (15) and (16) by means of sequential Monte Carlo methods. Assume that a weighted sample $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$ has been drawn from the filtering density at time t,

$$\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta) = \sum_{i=1}^N w_t^i \delta_{\{\mathbf{x}_t^i\}}(d\mathbf{x}_t) \tag{36}$$

Each simulated value $\mathbf{x}_t^i$ is called *particle* and the particles set, $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$, can be viewed as a random discretization of the state space $\mathcal{X}$, with associated weights w_t^i. It is possible to approximate, by means of this particle set, the prediction density given in equation (15) as follows:

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \theta)p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta)d\mathbf{x}_t \simeq \sum_{i=1}^N w_t^i p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta) \tag{37}$$

---

[7]See also Doucet, Freitas and Gordon (2001) for an updated review on particle filter techniques, their applications and the main convergence results.
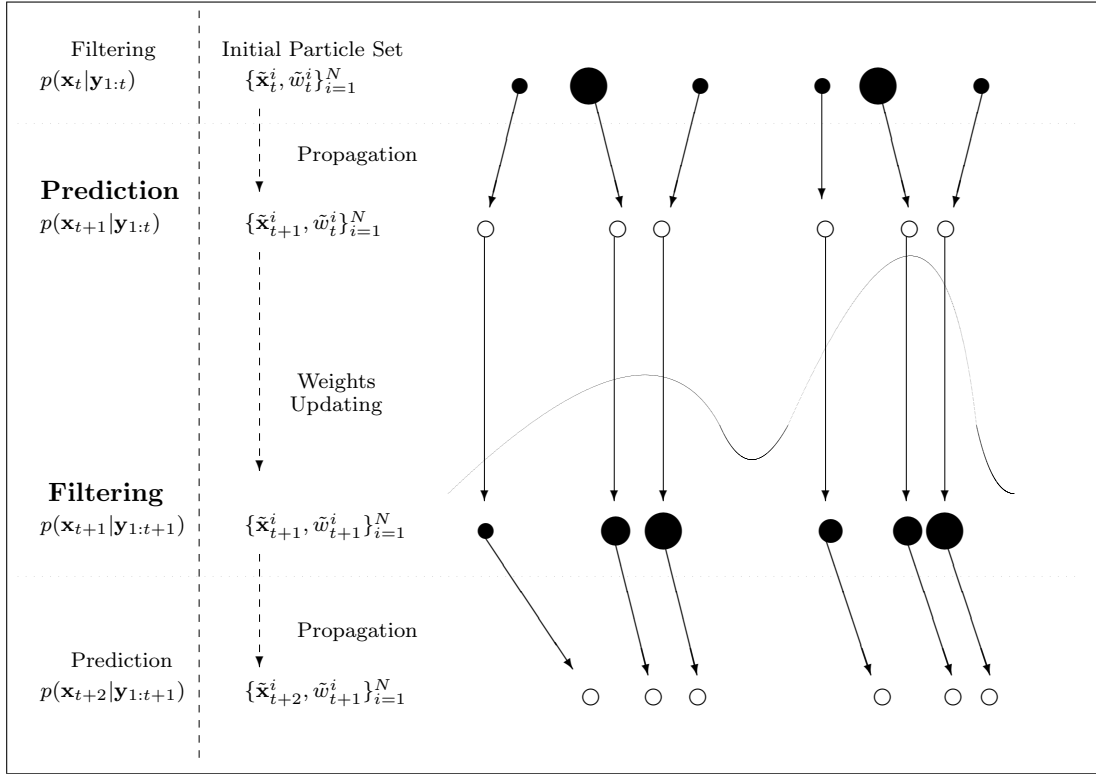
Figure 3: Evolution of the particle set in the Sequential Importance Sampling algorithm.

which is called *empirical prediction density* and is denoted by $\hat{p}(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta)$. By applying the Chapman-Kolmogorov equation, it is also possible to obtain an approximation of the filtering density given in equation (16),

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1}, \theta) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) \simeq \sum_{i=1}^{N} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta)w_t^i \quad (38)$$

which is called *empirical filtering density* and is denoted by $\hat{p}(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1}, \theta)$.

Assume now that the quantity $\mathbb{E}(f(\mathbf{x}_{t+1})|\mathbf{y}_{1:t+1})$ is of interest. It can be evaluated numerically by a Monte Carlo sample $\{\mathbf{x}_{t+1}^i, w_{t+1}^i\}_{i=1}^N$, drawn from the filtering distribution

$$\mathbb{E}(f(\mathbf{x}_{t+1})|\mathbf{y}_{1:t+1}) \simeq \frac{\frac{1}{N}\sum_{i=1}^{N} f(x_{t+1}^i)w_{t+1}^i}{\frac{1}{N}\sum_{i=1}^{N} w_{t+1}^i}. \quad (39)$$

A simple way to obtain a weighted sample from the filtering density at time $t + 1$ is to apply importance sampling to the empirical filtering density given in equation (38). This step corresponds to propagate the initial particle set (see figure 3) through the importance density $q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta)$. Moreover, if we propagate each particle through the transition density $p(\mathbf{x}_t|\mathbf{x}_{t-1}^i, \theta)$, then the particle weights update as follows:

$$w_{t+1}^i \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta)w_t^i}{q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta)} = w_t^i \, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta) \quad (40)$$

This is the natural choice for the importance density, because the transition density represents a sort of prior at time t for the state x_{t+1}. However, as underlined in Pitt and

Shephard (1999), this strategy is sensitive to outliers[8]. The basic particle filter developed through the previous equations is called *Sequential Importance Sampling* (SIS). In Algorithm 4, we give a pseudo-code representation of this method.

Sequential importance sampling permits to obtain recursive updating of the particle weights and is based on the sequential decomposition of the joint filtering density and on a particular choice of the importance density. To highlight these aspects, we consider the smoothing density $p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1}, \theta)$ of the state vectors and approximate it as follows:

$$p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1}, \theta) \simeq \sum_{i=1}^{N} \tilde{w}_{t+1}^i \delta_{\{\mathbf{x}_{0:t+1}^i\}}(d\mathbf{x}_{0:t+1}) \tag{41}$$

by simulating $\{\mathbf{x}_{0:t+1}^i\}_{i=1}^N$ from a proposal distribution $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}, \theta)$ and by correcting the weights of the resulting empirical density. The correction step comes from an importance sampling argument, thus the *unnormalized particle weights*[9] are defined as follows:

$$w_{t+1}^i \triangleq \frac{p(\mathbf{x}_{0:t+1}^i|\mathbf{y}_{1:t+1}, \theta)}{q(\mathbf{x}_{0:t+1}^i|\mathbf{y}_{1:t+1}, \theta)}. \tag{42}$$

The key idea used in the SIS algorithm consists in obtaining a recursive relation for the weights updating. This property makes them particulary appealing for on-line applications. Assume that the dynamic model of interest is the one described in equations (12), (13) and (14) and choose the importance density to factorize as follows: $q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1}, \theta) = q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}, \theta)q(\mathbf{x}_{t+1}|\mathbf{x}_{0:t}, \mathbf{y}_{1:t+1}, \theta)$, then the weights can be rewritten in the following recursive form:

$$w_{t+1}^i = w_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta)p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \theta)}{q(\mathbf{x}_{t+1}^i|\mathbf{x}_{t+1}^i, \mathbf{y}_{t+1}, \theta)} \tag{43}$$

This relation is a direct consequence of the Bayes rule and the Markov property of the system.

Algorithm 4 - *SIS Particle Filter* - *Given the initial set of particles $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$, for $i = 1, \ldots, N$:*

1. *Simulate $\mathbf{x}_{t+1}^i \sim q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta)$*

2. *Update the weights: $w_{t+1}^i \propto w_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta)\, p(\mathbf{x}_{t+1}|\mathbf{x}_t^i; \theta)}{q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta)}$*

---

[8]See also Crisan and Doucet (2000), for a discussion on the choice of the importance densities. They focused on the properties of the importance density, which are necessary conditions for the a.s. convergence of the sequential Monte Carlo algorithm.

[9]Note that importance sampling requires to know the importance and the target distributions up to a proportionality constant, thus the unnormalized weights may not sum to one. However *normalized importance sampling weights* can be easily obtained as follows

$$\tilde{w}_t^i = \frac{w_t^i}{\sum_{j=1}^N w_t^j} \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T.$$

The normalization procedure causes the loss of the unbiasness property.

It is well known in the literature (see for example Arulampalam *et al.* (2001)), that basic SIS algorithms have a degeneracy problem. After some iterations the empirical distribution degenerates into a single particle, because the variance of the importance weights is non-decreasing over time (see Doucet *et al.* (2000)). In order to solve this degeneracy problem, the *Sampling Importance Resampling* (SIR) algorithm has been introduced by Gordon *et al.* (1993). This algorithm belongs to a wider class of bootstrap filters, which use a re-sampling step to generate a new set of particles with uniform weights. This step introduces diversity in the particle set, avoiding degeneracy.

Algorithm 5 - *SIR Particle Filter* - *Given the initial set of particles* $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$:

1. *Simulate* $\mathbf{x}_{t+1}^i \sim q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta)$

2. *Update the weights:* $\tilde{w}_{t+1}^i \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta)$

3. *Normalize the weights:* $\bar{w}_{t+1}^i = \tilde{w}_{t+1}^i \left(\sum_{j=1}^N \tilde{w}_{t+1}^j\right)^{-1}$, *for* $i = 1, \ldots, N$.

4. *Simulate* $\{x_{t+1}^i\}_{i=1}^N$ *from the empirical density* $\{\mathbf{x}_t^i, \bar{w}_t^i\}_{i=1}^N$

5. *Assign* $w_{t+1}^i = 1/N$, *for* $i = 1, \ldots, N$.

Note that in the SIR particle filter, we assumed $q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta) = p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta)$. Moreover, due to the resampling step, the weights are uniformly distributed over the particle set: $w_t^i = 1/N$. Thus, the weights updating relation becomes: $\tilde{w}_{t+1}^i \propto w_t^i\, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta)$.

However, the basic SIR algorithm produces a progressive impoverishment of the information contained in the particle set, because of the resampling step and of the fact that particles do not change over the filter iterations. Many solutions have been proposed in literature. We recall the *Regularised Particle Filter* proposed by Musso *et al.* (2001), which is based on a discretisation of the continuous state space. Gilks and Berzuini (2001) propose the SIR-Move algorithm, which moves particles after the re-sampling step. Thus, particle value changes and the impoverishment is partially avoided. Finally, Pitt and Shephard (1999) introduce the *Auxiliary Particle Filter* (APF) and apply it to a Gaussian ARCH-type stochastic volatility model. They find that the auxiliary particle filter works well and that the sensibility to outliers is lower than in the basic filters.

In order to avoid re-sampling, the APF algorithm uses an auxiliary variable to select most representative particles and to mutate them through a simulation step. Then, weights of the regenerated particles are updated through an importance sampling argument. In this way particles with low probability do not survive to the selection and the information contained in the particle set is not wasted. In particular, the auxiliary variable μ_t^i contains and resumes the information on the previous particle set and it is used in the selection step to sample the random particle index. Note that the empirical filtering density given in equation (38) is a mixture of distributions, which can be reparameterised by introducing an auxiliary variable $i \in \{1, \ldots, N\}$, which indicates the component of the

mixture. The joint distribution of the hidden state and of the index i is then:

$$p(\mathbf{x}_{t+1}, i|\mathbf{y}_{1:t+1}, \theta) = \frac{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \mathbf{x}_{t+1}, i)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \theta)} p(\mathbf{x}_{t+1}, i|\mathbf{y}_{1:t}, \theta) = \tag{44}$$

$$= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \theta)} p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta) w_t^i.$$

The basic idea of the APF is to refresh the particle set while reducing the loss of information due to this operation. Thus, the algorithm generates a new set of particles by jointly simulating the particle index i (*selection step*) and the selected particle value $\mathbf{x}_{t+1}$ (*mutation step*) from the reparameterised empirical filtering density, according to the following importance density:

$$q(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1}, \theta) = q(\mathbf{x}_{t+1}^j|\mathbf{y}_{1:t+1}, \theta) q(i^j|\mathbf{y}_{1:t+1}, \theta)$$

$$= p(\mathbf{x}_{t+1}^j|\mathbf{x}^{i^j}, \theta)(p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta) w_t^{i^j}) \tag{45}$$

for $j = 1, \ldots, N$. Note that the index is sampled using weights which are proportional to the observation density conditional on a summary statistics of the initial particle set. In this way, less informative particles are discarded. The information contained in each particle is evaluated with respect to both the observable variable and the initial particle set. By following the usual importance sampling argument, the updating relation for the particle weights is:

$$w_{t+1}^j \triangleq \frac{p(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1}, \theta)}{q(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1}, \theta)} = \frac{p(\mathbf{x}_{t+1}^j|\mathbf{x}^{i^j}, \theta) p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta) w_t^{i^j}}{p(\mathbf{x}_{t+1}^j|\mathbf{x}^{i^j}, \theta) p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta) w_t^{i^j}} = \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta)} \tag{46}$$

Algorithm 6 - *Auxiliary Particle Filter* - *Given the initial set of particles* $\{\mathbf{x}_t^j, w_t^j\}_{j=1}^N$, *for* $j = 1, \ldots, N$,

1. *Calculate* $\mu_{t+1}^j = \mathbb{E}(\mathbf{x}_{t+1}|\mathbf{x}_t^j, \theta)$

2. *Simulate* $i^j \sim q(i|\mathbf{y}_{1:t+1}, \theta) \propto w_t^i p(\mathbf{y}_{t+1}|\mu_{t+1}^i, \theta)$ *with* $i \in \{1, \ldots, N\}$

3. *Simulate* $\mathbf{x}_{t+1}^j \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^{i^j}, \theta)$

4. *Update particles weights:* $\tilde{w}_{t+1}^j \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta)}$.

5. *Normalize the weights:* $w_{t+1}^i = \tilde{w}_{t+1}^i (\sum_{j=1}^N \tilde{w}_{t+1}^j)^{-1}$, *for* $i = 1, \ldots, N$.

We conclude this section with a brief discussion of the problem of parameter estimation, for dynamic models with hidden variables, in a sequential data-processing approach. Note that following the engineering literature, a common way to solve the parameter estimation problem is to treat parameters θ as hidden state of the system (see Berzuini et al. (1997)). The model is restated assuming time dependent parameter vectors θ_t, and imposing the constraint: $\theta_t = \theta_{t-1}$ on its dynamic.

In principle, parameter estimate and state filtering can be treated separately (see Storvik (2002)). In many applications of particle filter techniques, parameters are treated as

known and MCMC parameter estimates are used instead of the true parameter values. But in this way parameter estimate are not continuously updated as the hidden states. MCMC is typically a off-line approach, then it does not allow the sequential updating of parameter estimates, as new observations arrive. Moreover, when applied sequentially, MCMC estimation method is more time consuming than particle filter algorithms.

One of the main issue in research on particle filter is the inclusion of the parameter estimation procedure in the state filtering algorithm. See for example Berzuini et al. (1997) and Storvik (2002) for a general discussion of the problem, Liu and West (2001) for the joint application of adaptive importance sampling for parameter estimation and APF for hidden state filtering. See also Casarin (2004a), (2004b) for an updated review and an application to heavy tail stochastic volatility models.

4   An Application to Business Cycle Models

The aim of this section is to show how particle filter algorithms apply to a widely used class of business cycle models: *Markov switching stochastic latent factor models*. We apply APF algorithm to synthetic data in order to verify the efficiency of the algorithm and to detect possible degeneracy of the APF algorithm.

We refer to the model of Example 2.2 and apply the algorithm of to Liu and West (2001). This algorithm combines adaptive importance sampling for sequential estimation of the parameter vector with the auxiliary particle filter for filtering and predicting the hidden state. Observe that the latent structure of the MS model in the example exhibits two levels. The first one is given by the stochastic latent factor x_t and the second one is given by the regime switching process s_t.

We adapt the algorithm of Liu and West (2001) and obtain the following particle filter algorithm.

Algorithm 7 *Given an initial set of particles* $\{\mathbf{x}_t^i, s_t^i, \theta_t^i, w_t^i\}_{i=1}^N$,

1. *Compute* $V_t = \sum_{i=1}^N (\theta_t^i - \bar{\theta}_t)(\theta_t^i - \bar{\theta}_t)' w_t^i$ *and* $\bar{\theta}_t = \sum_{i=1}^N \theta_t^i w_t^i$

2. *For* $i = 1, \ldots, N$ *and with a and b well chosen tuning parameters, calculate the following summarizing constant:*

   (a) $\tilde{S}_{t+1}^i = \underset{l \in 1,2}{\arg\max}\, \mathbb{P}(s_{t+1} = l | s_t = s_t^i)$

   (b) $\tilde{X}_{t+1}^i = \mu_t^i + \nu_t^i \tilde{S}_{t+1}^i + \rho_t^i x_t^i$

   (c) $\tilde{\theta}_t^i = a\theta_t^i + (1-a)\bar{\theta}_t$, *where* $\tilde{\theta} = (\tilde{\alpha}, \tilde{\sigma}_\varepsilon, \tilde{\rho}, \tilde{\mu}, \tilde{\nu}, \tilde{\sigma}_\eta, \tilde{p}_{11}, \tilde{p}_{22})$

3. *For* $i = 1, \ldots, N$:

   (a) *Simulate* $k^i \propto q(k|\mathbf{y}_{1:t+1}, \theta) = N(\mathbf{y}_{t+1}|\tilde{\alpha}_t^k \tilde{X}_{t+1}^k, \tilde{\sigma}_{\varepsilon t}^k) w_t^k$, *with* $k \in \{1, \ldots, N\}$

   (b) *Simulate* θ_{t+1}^i *from* $N(\tilde{\theta}_t^{k^i}, b^2 V_t)$

   (c) *Simulate* $s_{t+1}^i \in \{1, 2\}$ *from* $\mathbb{P}(s_{t+1}^i = i | s_t^{k^i})$

(d) Simulate x_{t+1}^i from $N(\mu_{t+1}^i + \nu_{t+1}^i s_{t+1}^i + \rho_{t+1}^i x_t^{k^i}, \sigma_{\eta\,t+1}^i)$

4. Update weights $\tilde{w}_{t+1}^i \propto N(\mathbf{y}_{t+1}|\alpha_{t+1}^i x_{t+1}^i, \sigma_{\varepsilon\,t+1}^i)/N(\mathbf{y}_{t+1}|\tilde{\alpha}_t^{k^i} \tilde{X}_{t+1}^{k^i}, \tilde{\sigma}_{\varepsilon\,t}^{k^i})$

5. Normalize weights $w_{t+1}^i = \tilde{w}_{t+1}^i \left(\sum_{i=1}^N \tilde{w}_{t+1}^i\right)^{-1}$, for $i = 1, \ldots, N$.

The tuning parameters a and b are equal to $\frac{3\delta-1}{2\delta}$ and $\sqrt{1-a^2}$ respectively, where we chose $\delta = 0.99$ as suggested in West (1993).

Figure 4 shows on-line estimation of parameters α, σ_ε, ρ, μ_0, μ_1, σ_η, p_{11}, p_{22} obtained by running APF algorithm on the synthetic dataset exhibited in figure 2. We use a set of $N = 1000$ particles to obtain the empirical filtering and prediction densities. All computations have been carried out on a Pentium IV 2.4 Ghz, and the APF algorithm has been implemented in GAUSS 4.0. Figure 5 shows on-line estimation of the latent factor x_t. To detect the absence of degeneracy in the output of the APF algorithm we evaluate at each step the *Survival Rate*: it is the number of particles survived to the selection step over the total number of particles. Particle set degenerates when persistently exhibiting a high number of dead particles from a generation to the subsequent one. Survival rate is calculated as follows:

$$SR_t = \left\{N - \sum_{i=1}^N \mathbb{I}_{\{0\}}(Card(I_{i,t}))\right\}N^{-1} \tag{47}$$

where $I_{i,t} = \{j \in \{1, \ldots, N\}|i_t^j = i\}$ is the set of all random index values, which select at time t the i-*th* particle. If at time t the particle k does not survive then the set $I_{k,t}$ becomes empty. Figure 5 shows the survival rate at each time step. The rate does not decrease thus we conclude that the APF algorithm does not degenerate in our simulation study.

5 Conclusion

We describe the Bayesian approach to general dynamic models analysis. We briefly review the literature on the business cycle modelling, focusing on the Bayesian approach and recognizing the importance of simulation based methods. To better understand the usefulness of simulation based methods in business cycle analysis, we analyze the problems of state filtering and parameter estimation for a quite general class of dynamic models used in time series analysis and exhibit general filtering, predicting and smoothing relations. Furthermore, we focus on Bayesian simulation based inference, suggest the use of sequential simulation techniques to make inference on business cycle models and provide an application on synthetic data.

References

ANAS, J., AND L. FERRARA (2002): "Detecting cyclical turning points: the ABCD approach and two probabilistic indicators," presented at the 26th Ciret Conference in Taiwan (www.coe.ccip.fr).
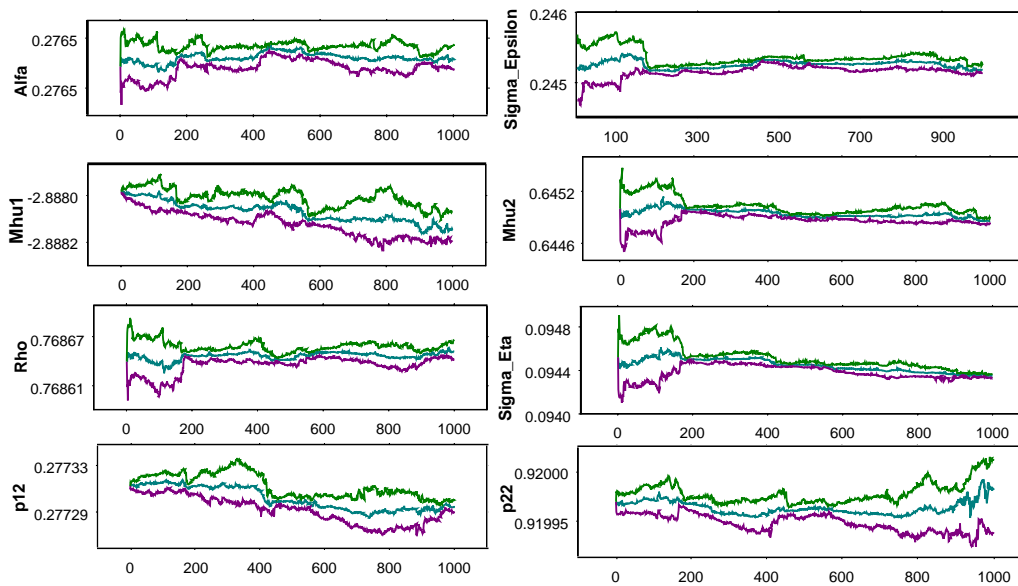
Figure 4: On-line parameter estimates. Graphs exhibit at each date the empirical mean and quantiles at 0.275 and 0.925 of the parameters.
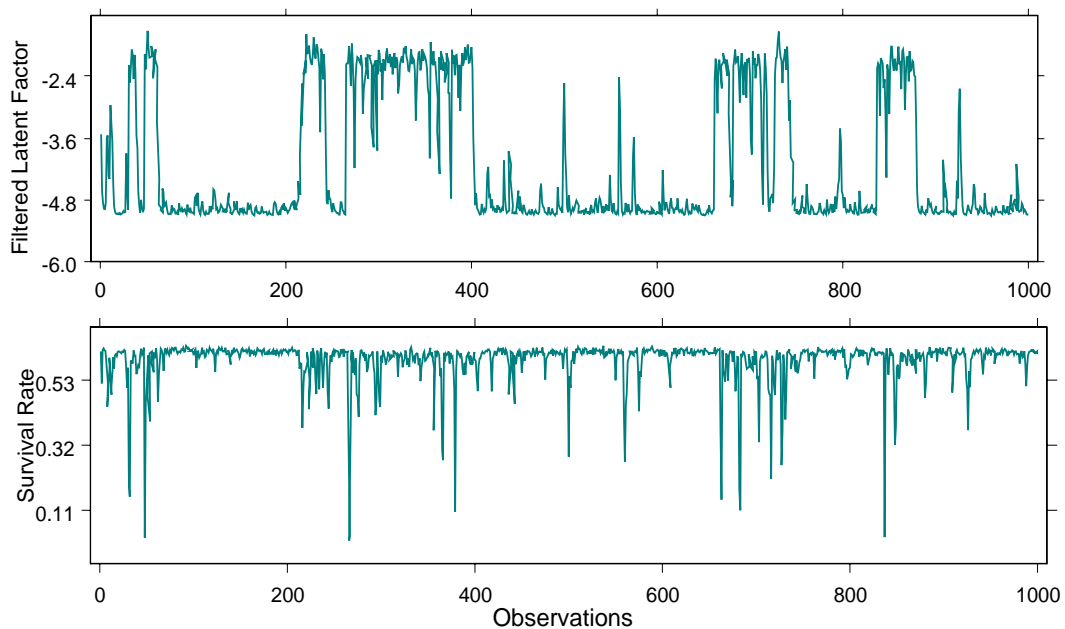

Figure 5: Sequentially filtered latent factor and survival rate of the particle set, over $T = 1,000$ observations.

ARULAMPALAM, S., S. MASKELL, N. GORDON, AND T. CLAPP (2001): "A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking," Technical Report, QinetiQ Ltd., DSTO, Cambridge.

BAUWENS, L., M. LUBRANO, AND J. F. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models.* Oxford University Press, New York.

BERZUINI, C., N. G. BEST, W. R. GILKS, AND C. LARIZZA (1997): "Dynamic conditional independence models and Markov chain Monte Carlo methods," *Journal of the American Statistical Association*, 92, 1403–1441.

BERZUINI, C., AND W. R. GILKS (2001): "Following a moving average-Monte Carlo inference for dynamic Bayesian models," *J.R. Statist. Soc. B*, 63, 127–146.

BILLIO, M., R. CASARIN, AND D. SARTORE (2004): "Bayesian Inference on Dynamic Models with Latent Factors," Working Paper GRETA, Venice.

CARLIN, B. P., N. G. POLSON, AND D. S. STOFFER (1992): "A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modelling," *Journal of American Statistical Association*, 87, n. 418, 493–500.

CARTER, C. K., AND R. KÖHN (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81, n. 3, 541–553.

CASARIN, R. (2004a): "Bayesian Inference for Markov Switching Stochastic Volatility Models," Cahier du CEREMADE, N. 0414.

——— (2004b): "Bayesian Monte Carlo Filtering for Stochastic Volatility Models," Cahier du CEREMADE, N. 0415.

CASELLA, G., AND C. P. ROBERT (1999): *Monte Carlo Statistical Methods.* Springer Verlag, New York.

CRISAN, D., AND A. DOUCET (2000): "Convergence of sequential Monte Carlo methods," Technical Report N. 381, CUED-F-INFENG.

DIEBOLD, F. X., AND G. D. RUDEBUSCH (1996): "Measuring Business Cycles: A Modern Perspective," *The Review of Economics and Statistics*, 78, 67–77.

DOUCET, A., J. G. FREITAS, AND J. GORDON (2001): *Sequential Monte Carlo Methods in Practice.* Springer Verlag, New York.

DOUCET, A., S. GODSILL, AND C. ANDRIEU (2000): "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, 10, 197–208.

DURBIN, J., AND S. J. KOOPMAN (2001): *Time Series Analysis by State Space Methods.* Oxford University Press, New York.

FRÜHWIRTH-SCHNATTER, S. (1994): "Data augmentation and dynamic linear models," *Journal of Time Series Analysis*, 15, N .2, 183–202.

GOLDFELD, S. M., AND R. E. QUANDT (1973): "A Markov Model for Switching Regression," *Journal of Econometrics*, 1, 3–16.

GORDON, N., D. SALMOND, AND A. F. M. SMITH (1993): "Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation," *IEE Proceedings-F*, 140, 107–113.

HAMILTON, J. D. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–384.

HARRISON, J., AND M. WEST (1997): *Bayesian Forecasting and Dynamic Models, 2nd Ed.* Springer Verlag, New York.

HARVEY, A. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.

KALMAN, R. (1960): "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering, Transaction ASME, Series D*, 82, 35–45.

KALMAN, R. E., AND R. S. BUCY (1960): "New results in linear filtering and prediction problems," *Transaction of the ASME-Journal of Basic Engineering, Series D*, 83, 95–108.

KIM, C. J. (1994): "Dynamic linear models with Markov switching," *Journal of Econometrics*, 60, 1–22.

KIM, C. J., AND C. J. MURRAY (2002): "Permanent and Transitory Components of Recessions," *Empirical Economics*, 27/2, 163–183.

KIM, C. J., AND C. R. NELSON (1999): *State-Space Models with Regime Switching.* MIT press, Cambridge.

KROLZIG, H. (1997): *Markov-switching vector autoregressions. Modelling, statistical inference and applications to business cycle analysis.* Springer, Berlin.

LIU, J., AND M. WEST (2001): "Combined Parameter and State Estimation in Simulation Based Filtering," in *Sequential Monte Carlo Methods in Practice*, ed. by F. J. Doucet A., and G. J. Springer-Verlag, New York.

LIU, J. S., AND R. CHEN (1998): "Sequential Monte Carlo Methods for Dynamical System," *Journal of the American Statistical Association*, 93, 1032–1044.

LIU, J. S., W. H. WONG, AND A. KONG (1994): "Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes," *Biometrika*, 81, 27–40.

——— (1995): "Correlation structure and convergence rate of the Gibbs sampler with various scans," *Journal of the Royal Statistical Society B*, 57, 157–169.

MÜLLER, P. (1992): "Alternatives to the Gibbs sampling scheme," Tech. report, Institute of Statistics and Decision Sciences, Duke University.

MUSSO, C., N. OUDJANE, AND F. LEGLAND (2001): "Improving Regularised Particle Filters," in *Sequential Monte Carlo in Practice*, ed. by F. J. Doucet A., and G. J. Springer Verlag, New York.

PITT, M., AND N. SHEPHARD (1999): "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94(446), 590–599.

SHEPHARD, N. (1994): "Partial non-Gaussian state space," *Biometrika*, 81, 115–131.

SHEPHARD, N., AND M. K. PITT (1997): "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, 84, Issue 3, 653–667.

SICHEL, D. E. (1991): "Business cycle duration dependence: A parametric approach," *Review of Economics and Statistics*, 73, 254–256.

STORVIK, G. (2002): "Particle filters for state space models with the presence of unknown static parameters," *IEEE Trans. on Signal Processing*, 50, 281–289.

TANNER, M., AND W. WONG (1987): "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, 82, 528–550.

TIERNEY, L. (1994): "Markov chains for exploring posterior distributions," *Ann. of Statist.*, 22, 1701–1786.

WATSON, J. (1994): "Business cycle durations and postwar stabilization of the u.s. economy," *American Economic Review*, 84, 24–46.

WEST, M. (1992): "Mixture models, Monte Carlo, Bayesian updating and dynamic models," *Computer Science and Statistics*, 24, 325–333.

——— (1993): "Approximating posterior distribution by mixtures," *Journal of Royal Statistical Society, B*, 55, 409–442.