Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 4526 – 4542
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Simulation-based sequential analysis of Markov switching stochastic volatility models

Carlos M. Carvalho[a, b, *], Hedibert F. Lopes[a, b]

[a] *ISDS-Duke University, NC, USA*
[b] *GSB-University of Chicago, IL, USA*

Received 13 March 2006; received in revised form 11 July 2006; accepted 18 July 2006
Available online 10 August 2006

Abstract

We propose a simulation-based algorithm for inference in stochastic volatility models with possible regime switching in which the regime state is governed by a first-order Markov process. Using auxiliary particle filters we developed a strategy to sequentially learn about states and parameters of the model. The methodology is tested against a synthetic time series and validated with a real financial time series: the IBOVESPA stock index (São Paulo Stock Exchange).

Keywords: Bayesian time series; Bayes factor; Markov chain Monte Carlo; Particle filters; Sequential analysis; Stochastic volatility models

1. Introduction

Over the years stochastic volatility (SV) models have been considered a useful tool for modeling time-varying variances, mainly in financial applications where economic agents are constantly facing decision problems dependent on measures of volatility and risk. See, for example, Engle (1982) and Bollerslev et al. (1994), for the family of autoregressive conditional heteroscedasticity (ARCH) models, and Jacquier et al. (1994), and Kim et al. (1998), for the Bayesian estimation of SV models.

One important discovery of the conditional variance and SV literature is the substantial persistence in high-frequency data, especially in financial markets (Chou, 1988). Lamoureux and Lastrapes (1990) argued that the persistence could be overestimated if structural changes in the volatility process were ignored in the model. In this paper we focus on a particular SV model that allows occasional discrete shifts in the parameter determining the level of the logarithm of volatility through a Markovian process. So et al. (1998) suggested that this model not only is a better way to explain volatility persistence but is also a tool to capture changes in volatility due to economic forces, as well as abrupt changes due to unusual market events.

* Corresponding author. ISDS-Duke University, NC, USA. Tel.: +1 919 6848753; fax: +1 919 6848594.
E-mail address: carlos@stat.duke.edu (C.M. Carvalho).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2006.07.019

Several attempts have already been made to fit these types of models (or similar ones). West and Harrison (1997) and Harvey (1989) introduced variants of the Kalman filter to deal with possibly time-varying parameters in a dynamic linear model. One of the main drawbacks of those Kalman-filter-like methods is their overparametrization. West and Harrison (1997) also introduced a multiprocess model that entertains several alternative models at once. Their idea was proved to be of limited practical use since the number of possible scenarios explodes even with a small number of observations.

More recently, Hamilton and Susmel (1994) proposed an ARCH model with regime switching for volatility level (SWARCH), while So et al. (1998) developed a methodology for SV models with regime switching. The first article uses an iterative method to find the maximum likelihood estimate for the model's parameters and states. So et al. (1998), on the other hand, use the forward filtering–backward sampling algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994) and the smoother algorithm of Shephard (1994) to sample from the exact posterior distribution of parameters and states of the model. Unfortunately, both methods can be considered static since inferences are made based upon the whole dataset. In real time applications, where decisions need to be made instantly, such methods are bound to fail and methods that allow for sequential, on-line updates are preferred.

Our main contribution is to combine the auxiliary particle filter developed by Pitt and Shephard (1999) with Liu and West's (2001) ideas on how to update the parameters to propose an algorithm that sequentially estimates a Markov switch SV model. In Section 2 we introduce the Markov switching SV (MSSV) model, while the proposed methodology to sequentially update parameters and states is presented in Section 3. The methodology is subsequently applied to simulated datasets and a financial time series, the Brazilian Ibovespa stock index. This is done in Section 4.

2. Markov switching stochastic volatility

Let y_t denote the observed value of a quantity of interest at time t, which is, in our application, the daily returns of the Ibovespa index. In the MSSV model the observations $y_1, \ldots, y_T$ are conditionally independent and normally distributed with time-varying log-volatilities $\lambda_1, \ldots, \lambda_T$, i.e., $y_t|\lambda_t \sim N(0, \exp(\lambda_t))$, or

$$p(y_t|\lambda_t) = (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}\left[\lambda_t + y_t^2/\mathrm{e}^{\lambda_t}\right]\right\} \tag{1}$$

with $(\lambda_t|\lambda_{t-1}, \xi, s_t) \sim N\left(\alpha_{s_t} + \phi\lambda_{t-1}, \sigma^2\right)$, or

$$p(\lambda_t|\lambda_{t-1}, \xi, s_t) = \left(2\pi\sigma^2\right)^{-1/2} \exp\left\{-\frac{\left(\lambda_t - \alpha_{s_t} - \phi\lambda_{t-1}\right)^2}{2\sigma^2}\right\} \tag{2}$$

and regime variables s_t following a k-state first order Markov process,

$$p_{ij} = Pr(s_t = j|s_{t-1} = i) \quad \text{for } i, j = 1, \ldots, k, \tag{3}$$

$\alpha = (\alpha_1, \ldots, \alpha_k)$, $\xi = \left(\alpha, \phi, \sigma^2\right)$ and $P = \left(p_{11}, \ldots, p_{1k-1}, \ldots, p_{k1}, \ldots, p_{k,k-1}\right)$. We will denote by θ the $\left(k^2+2\right)$-dimensional vector of parameters, i.e., $\theta = (\xi, P)$. For instance, in a 2-state model, there will be six parameters, while in a 3-state θ will be of dimension 11. It is common in the dynamic model literature to refer to $s = (s_1, \ldots, s_T)$ and $\lambda = (\lambda_1, \ldots, \lambda_T)$ as the states of the model. In order to avoid identification problems, we adopt the following reparametrization for α_{s_t}:

$$\alpha_{s_t} = \gamma_1 + \sum_{j=1}^{k} \gamma_j I_{jt}, \tag{4}$$

where $I_{jt} = 1$ if $s_t \geqslant j$ and zero otherwise, $\gamma_1 \in \Re$ and $\gamma_i > 0$ for $i > 1$. In this model, α corresponds to the level of the log-volatility and in order to allow occasional discrete changes the model introduces different α's following a first-order Markovian process. It would be straightforward to allow other parameters to change according to the same Markovian process, however, the goal of this model is to isolate clusters of high and low volatility, captured in the different α's, and therefore more precisely estimate the persistence parameter ϕ (Hamilton and Susmel, 1994).

The model specification is completed with the prior distribution for the vector θ: $\gamma_1 \sim N\left(\gamma_{10}, C_{\gamma_1}\right)$, $\phi \sim TN_{(-1,1)}\left(\phi_0, C_\phi\right)$, $\sigma^2 \sim IG(a,b)$, $\lambda_0 \sim N\left(\lambda_{00}, C_{\lambda_0}\right)$, $\gamma_i \sim TN_{(0,\infty)}\left(\gamma_{i0}, C_{\gamma_i}\right)$ for $i=2,\ldots,k$ and $p_i \sim \text{Dir}\,(u_{i0})$, for $p_i=(p_{i1},\ldots,p_{ik})$ and $i=1,\ldots,k$. The hyperparameters are chosen to represent fairly non-informative priors, so we set $\phi_0=\lambda_{00}=0$, $C_\phi=C_{\lambda_0}=100$, $a=2.001$, $b=1$, $\gamma_{i0}=0$, $C_{\gamma_i}=100$, and $u_{i0}=(0.5,\ldots,0.5)$ for $i=1,\ldots,k$.

A Markov chain Monte Carlo (MCMC) algorithm for the MSSV model has already been developed by So et al. (1998). The algorithm iterates through a forward filtering–backward sampling step for λ and s (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994), and a traditional linear regression (Gibbs) step for θ (Gelfand and Smith, 1990). The main problem with the MCMC algorithm is that when a new observation arrives, at time $T+1$, the whole algorithm must be re-run in order to generate draws from $p\,(\theta,\lambda,s|y_1,\ldots,y_{T+1})$, the posterior distribution. Computation becomes practically infeasible when, for instance, observations arrive every minute or second, such as in certain financial applications and target tracking problems (Doucet et al., 2001).

In the next section, a customized particle filter algorithm (Pitt and Shephard, 1999; Doucet et al., 2001) will be tailored to allow on-line (sequential) update of the joint posterior distribution of (θ,λ,s) of the MSSV model as observations arrive in time.

3. Sequential Monte Carlo filter

Eqs. (1)–(3) can be seen, respectively, as the observation and system equations of a non-normal and non-linear dynamic model (West and Harrison, 1997; Jacquier et al., 1994). More specifically,

$$p\,(y_t|x_t,\theta) \sim N\left(0, \mathrm{e}^{\lambda_t/2}\right),$$

$$p\,(\lambda_t|x_{t-1},\theta) \sim N\left(\alpha_{s_t}+\phi\lambda_{t-1},\sigma^2\right),$$

$$Pr\,(s_t|x_{t-1},\theta) = p_{s_{t-1}s_t},$$

where $x_t=(\lambda_t,s_t)$ plays the role of the (hidden) state vector. Following West and Harrison's (1997) standard notation, $D_t=\{y_1,\ldots,y_t\}$ and $p\,(x_0,\theta|D_0)$ is the posterior distribution of x_0 and θ at time $t=0$, or simply the prior distributions introduced in Section 2. The sequential learning process starts the cycle at time t with the posterior distribution of x_t and θ, denoted by $p\,(x_t,\theta|D_t)$, and finishes the cycle at time $t+1$ with the posterior distribution of x_{t+1} and θ, i.e., $p\,(x_{t+1},\theta|D_{t+1})$. First, we combine the system equation, $p\,(x_{t+1},\theta|x_t)$ (Eqs. (2)–(3)), with the time t posterior of x_t, to produce the time t prior distribution for x_{t+1} and θ,

$$p\,(x_{t+1},\theta|D_t) = \int p\,(x_{t+1},\theta|x_t)\,p\,(x_t|D_t)\,\mathrm{d}x_t. \tag{5}$$

Second, Bayes' theorem combines the prior $p\,(x_{t+1},\theta|D_t)$ and the likelihood (Eq. (1)), to produce the time $t+1$ posterior distribution of x_{t+1} and θ,

$$p\,(x_{t+1},\theta|D_{t+1}) \propto p\,(y_{t+1}|x_{t+1},\theta)\,p\,(x_{t+1},\theta|D_t)\,. \tag{6}$$

It is well known that closed form solutions for the evolution step (integral in Eq. (5)) and updating step (posterior in Eq. (6)) are only available under very restrictive conditions that usually lead to oversimplified models, such as Gaussianity and linearity (see West and Harrison, 1997, for further details about the normal dynamic linear model). Our MSSV model creates non-linearity in both the system and observation equations and non-normality in the observation equation, making close form on-line estimation practically and computationally infeasible.

Approximations such as linear Bayes were extensively used during the seventies and eighties (see for example, West et al., 1985) to overcome these problems. In the early nineties, the MCMC revolution shifted the attention of most work on dynamic models from on-line sampling, i.e., $p\,(x_t|D_t)$, to smoothed sampling, i.e., $p\,(x_t|D_T)$ for $T>t$ (Carlin et al., 1992; Carter and Kohn, 1994; Carter and Kohn, 1994; Frühwirth-Schnatter, 1994; Frühwirth-Schnatter, 1994).

Sequential Monte Carlo algorithms are, however, abundant nowadays thanks mainly to the seminal work of Kitagawa (1996) and Gordon et al. (1993) and the last ten years have witnessed the appearance of several *particle filters*, as sequential Monte Carlo methods have become to be known. Roughly speaking, Kitagawa (1996) utilizes sequential sets of particles to approximate the prior–predictive–posterior distributions of the states in a general univariate non-linear dynamic model, while Gordon et al. (1993), propose the use of a sequential sampling importance resampling (SIR) scheme where importance weights are sequentially updated based on samples from the prior distributions. Their filter is widely used and became to be well known as the *bootstrap filter*. In this paper we adopt Pitt and Shephard's (1999) auxiliary particle filter that generalizes the bootstrap filter by alternative importance sampling distributions. For further sequential Monte Carlo algorithms, applications and recent developments, see the book edited by Doucet et al. (2001) and the scientific reports on the sequential Monte Carlo website http://www-sigproc.eng.cam.ac.uk/smc/.

3.1. Auxiliary particle filters with known θ

In order to simplify the technical details behind the auxiliary particle filter we adopt, we start with the assumption that θ is known and will be omitted from the following developments. In Section 3.2 we jointly update x_t and θ. Roughly speaking, particle filters define a class of simulation filters that approximate the posterior distribution of x_t, $p(x_t|D_t)$, by a set of *particles* $\left\{x_t^{(1)}, \ldots, x_t^{(M)}\right\}$ with discrete probabilities $w_t^{(1)}, \ldots, w_t^{(M)}$, a relation that will be denoted here by $\left\{x_t^{(j)}, w_t^{(j)}\right\}_{j=1}^{M} \sim p(x_t|D_t)$.

Therefore, quantities such as $E\{g(x_t)|D_t\}$, the posterior expectation $g(x_t)$, can be directly approximated by $\sum_{i=1}^{M} g\left(x_t^{(j)}\right) w_t^{(j)}$. Similarly, the time t prior distribution for x_{t+1} (Eq. (5) with θ fixed and omitted), can be approximated by

$$\widehat{p}(x_{t+1}|D_t) = \sum_{j=1}^{M} p\left(x_{t+1}|x_t^{(j)}\right) w_t^{(j)} \tag{7}$$

while the time $t+1$ posterior distribution of x_{t+1} (Eq. (6) with θ fixed and omitted), can be approximated by

$$\widehat{p}(x_{t+1}|D_{t+1}) \propto p(y_{t+1}|x_{t+1}) \sum_{j=1}^{M} p\left(x_{t+1}|x_t^{(j)}\right) w_t^{(j)} \tag{8}$$

with Eqs. (7) and (8), respectively, called the *empirical prediction density* and the *empirical filtering density* by Pitt and Shephard (1999).

In order to complete the filtering process, one needs a scheme that transforms w_t's into w_{t+1}'s such that $\left\{x_{t+1}^{(j)}, w_{t+1}^{(j)}\right\}_{j=1}^{M} \sim \widehat{p}(x_{t+1}|D_{t+1})$. Following an idea from Smith and Gelfand (1992), Gordon et al. (1993) suggested a SIR filter by using the approximate prior distribution (Eq. (7)) as the importance function. They named the scheme the *bootstrap filter*. Despite being an intuitive and simple strategy to implement that simply reweights the current particles according to their likelihood, the bootstrap filter is very sensitive to the amount of prior information relative to the likelihood. In other words, relatively diffuse priors or relatively informative likelihood will impoverish the filter by weighting up only a small subset of the particles, consequently leading to the degeneracy of the on-line algorithm. See Fig. 1 for an illustrative example where the overlap between relatively non-informative prior and relatively peaked likelihood produces poor samples.

Pitt and Shephard (1999) developed what became known as the *auxiliary particle filter*, since it looks at the empirical filtering density (Eq. (8)) as a mixture of M distributions and introduces a latent indicator for the mixture components, $p(x_{t+1}, k) \propto p(y_{t+1}|x_{t+1})\, p\left(x_{t+1}|x_t^{(k)}\right) w_t^{(k)}$, leading to a sequential scheme that first samples k^l from $p\left(y_{t+1}|\mu_{t+1}^{(j)}\right) w_t^{(j)}$, with $\mu_{t+1}^{(j)}$ representing a *guess*, such as the mean or mode, from $p\left(x_{t+1}|x_t^{(j)}\right)$, and then

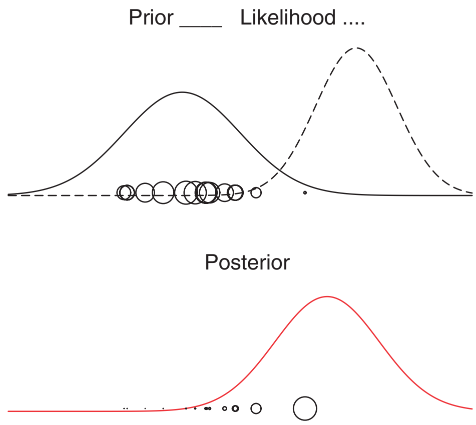

Fig. 1. Illustrative effect of the inefficiency of using the prior distribution as importance function in a sampling importance resampling.

Table 1
Parameter values for the four simulated examples

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
α_1	−2.500	−1.500	−0.500	−2.500
α_2	−1.000	−0.600	−0.200	−1.000
ϕ	0.500	0.700	0.900	0.500
p_{11}	0.990	0.990	0.990	0.500
p_{22}	0.985	0.985	0.985	0.500

samples $x_{t+1}^{(l)}$ from $p\left(x_{t+1}|x_t^{(k^l)}\right)$. Therefore, the importance function is $g\left(x_{t+1},k\right) \propto p\left(y_{t+1}|\mu_{t+1}\right) p\left(x_{t+1}|x_t^{(k)}\right) w_t^{(k)}$, which leads to weights $w_{t+1}^{(l)} \propto p\left(y_{t+1}|x_{t+1}^{(l)}\right)/p\left(y_{t+1}|\mu_{t+1}^{(k^l)}\right)$,and to $\left\{x_{t+1}^{(l)}, w_{t+1}^{(l)}\right\}_{l=1}^{M} \sim \widehat{p}\left(x_{t+1}|D_{t+1}\right)$.

With this procedure we are able to reduce the computational cost and improve the efficiency of the method by giving more importance to particles with larger predictive value. This procedure applies to any state-space model where both evolution and observational equations are known so, conditional on the parameters, it can be directly applied to our MSSV model, where λ_t and s_t are the components of x_t.

However, as mentioned at the beginning of this section, one should note that we need to update our information about θ, which in our context includes the volatility persistence ϕ, the volatility levels $\alpha_1, \ldots, \alpha_k$, the volatility variance σ^2 and the discrete state transition probabilities p_{ij}. In the next section we briefly describe Liu and West's (2001) ideas on sequential updating of fixed parameters, which will be incorporated in our MSSV filtering presented in Section 3.3.

3.2. Auxiliary particle filters with unknown θ

The problem of updating θ can be seen as a Bayesian sequential learning process where the goal is to update the following posterior density:

$$p\left(x_{t+1},\theta|D_{t+1}\right) \propto p\left(y_{t+1}|x_{t+1},\theta\right) p\left(x_{t+1}|\theta,D_t\right) p\left(\theta|D_t\right).$$

Gordon et al. (1993) suggest incorporating artificial evolution noise for θ and treat it as a state variable. The main drawback of their idea is simple: parameters are not states! Their approach imposes a loss of information in time as artificial uncertainties added to the parameters eventually resulting in a very diffuse posterior density for θ. Liu and West (2001) suggest a kernel smoothing to approximate $p\left(\theta|D_t\right)$ that relies upon West's (1992) mixture modeling ideas, which approximates the posterior density of θ (possibly transformed), $p(\theta|D_t)$, by a mixture of multivariate normals,

$$\widehat{p}(\theta|D_t) = \sum_{j=1}^{M} w_t^{(j)} N\left(a\theta_t^{(j)} + (1-a)\tilde{\theta}_t; b^2 V_t\right), \tag{9}$$

where $\tilde{\theta}_t = \sum_{j=1}^{M} w_t^{(j)} \theta_t^{(j)}$ and $V_t = \sum_{j=1}^{M} w_t^{(j)} \left(\theta_t^{(j)} - \tilde{\theta}_t\right)\left(\theta_t^{(j)} - \tilde{\theta}_t\right)'$ are approximations for $E(\theta|D_t)$ and $V(\theta|D_t)$, respectively. The constants a and b measure, respectively, the extent of the shrinkage and the degree of over dispersion of the mixture. As in Liu and West (2001), the choice of a and b will depend on a discount factor δ in (0, 1], typically around 0.95–0.99, so that $b^2 = 1 - ((3\delta - 1)/2\delta)^2$ and $a = \sqrt{1-b^2}$. This is a way to specify the controlling smoothing parameter b as a function of the amount of information preserved in each step of the filter.

Analogously to Pitt and Shephard's (1999) algorithm, the sequential scheme proceeds as follows. First, k^l is sampled from $p\left(y_{t+1}|\mu_{t+1}^{(j)}, \theta_t^{(j)}\right) w_t^{(j)}$, with $\mu_t^{(j)}$ as before. Second, $\theta_{t+1}^{(l)}$ is sampled from $N\left(m_t^{(k^l)}; b^2 V_t\right)$ where $m_t^{(k^l)} = a\theta_t^{(k^l)} + (1-a)\tilde{\theta}_t$. Third, $x_{t+1}^{(l)}$ is sampled from $p\left(x_{t+1}|x_t^{(k^l)}, \theta_{t+1}^{(l)}\right)$, which leads to weights $w_{t+1}^{(l)}$ proportional to $p\left(y_{t+1}|x_{t+1}^{(l)}, \theta_{t+1}^{(l)}\right)/p\left(y_{t+1}|\mu_{t+1}^{(k^l)}, m_t^{(k^l)}\right)$ and samples from the posterior, $\left\{x_{t+1}^{(l)}, \theta_{t+1}^{(l)}, w_{t+1}^{(l)}\right\}_{l=1}^{M} \sim \widehat{p}(x_{t+1}, \theta|D_{t+1})$.

3.3. Sequentially updating the MSSV model

We combine the auxiliary particle filter from Section 3.1 with the kernel smoothing approximation from Section 3.2 in order to design a sequential Monte Carlo filter for our MSSV model. The algorithm will sequentially generate samples $\left\{\lambda_{t+1}^{(j)}, s_{t+1}^{(j)}, \theta_{t+1}^{(j)}\right\}_{j=1}^{M}$ from $p(\lambda_{t+1}, s_{t+1}, \theta|D_{t+1}) \propto p(y_{t+1}|\lambda_{t+1})\, p(\lambda_{t+1}|s_{t+1}, \theta)\, p(s_{t+1}|\theta)\, p(\theta|D_t)$. The guess for the state at time $t+1$, $\mu_{t+1}^{(j)}$, needs to be carefully chosen, since the regime state s_t is a discrete variable. It is important to note that the parameter vector θ now includes k volatility levels $\alpha_1, \ldots, \alpha_k$, and a $k \times (k-1)$ matrix of Markov switching probabilities.

We should mention that θ was transformed to allow its components to vary in the real line, as it is assumed by the mixture of multivariate normals. The transformations are: $\log\left(\sigma^2\right)$, $\log\left(\gamma_i\right)$ for $i = 2, \ldots, k$ and $\log\left(p_{ij}/\left(1-p_{ij}\right)\right)$. The following chart gives a step-by-step description of the sequential Monte Carlo we design for the MSSV model.

SMC filter for the MSSV model

Step 0: $\left\{\lambda_t^{(j)}, s_t^{(j)}, \theta_t^{(j)}, w_t^{(j)}\right\}_{j=1}^{M} \sim p(\lambda_t, s_t, \theta|D_t)$.

Step 1: For $j = 1, \ldots, M$,

$$\tilde{s}_{t+1}^{(j)} = \arg\max_{l \in 1,\ldots,k} Pr\left(s_{t+1} = l|s_t = s_t^{(j)}\right),$$

$$\mu_{t+1}^{(j)} = \alpha_{\tilde{s}_{t+1}^{(j)}}^{(j)} + \phi_t^{(j)} \lambda_t^{(j)}.$$

Step 2: For $l = 1, \ldots, M$:

1. Sample k^l from $\{1, \ldots, k\}$, with $Pr\left(k^l\right) \propto p\left(y_{t+1}|\mu_{t+1}^{(k^l)}\right) w_t^{(k^l)}$.
2. Sample $\theta_{t+1}^{(l)}$ from $N\left(m_t^{(k^l)}, b^2 V_t\right)$.
3. Sample $s_{t+1}^{(l)}$ from $1, \ldots, k$ with $Pr\left(s_{t+1}^{(l)}\right) = Pr\left(s_{t+1}^{(l)}|s_t^{(k^l)}\right)$.
4. Sample $\lambda_{t+1}^{(l)}$ from $p\left(\lambda_{t+1}|\lambda_t^{(k^l)}, s_{t+1}^{(l)}, \theta_{t+1}^{(l)}\right)$.

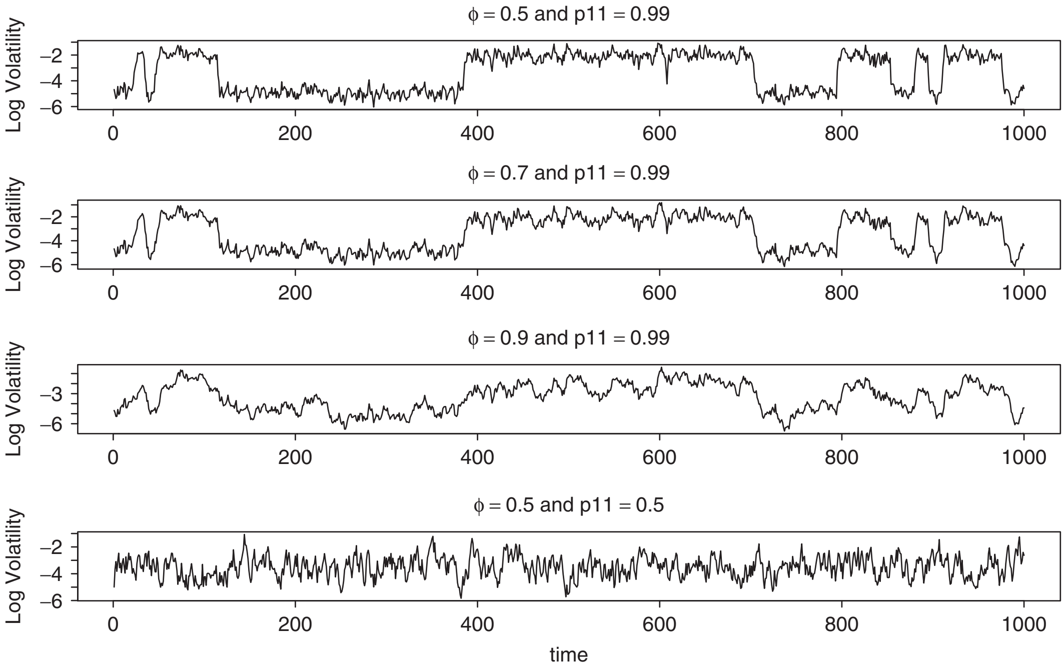

Fig. 2. Log-volatilities for all simulated datasets.

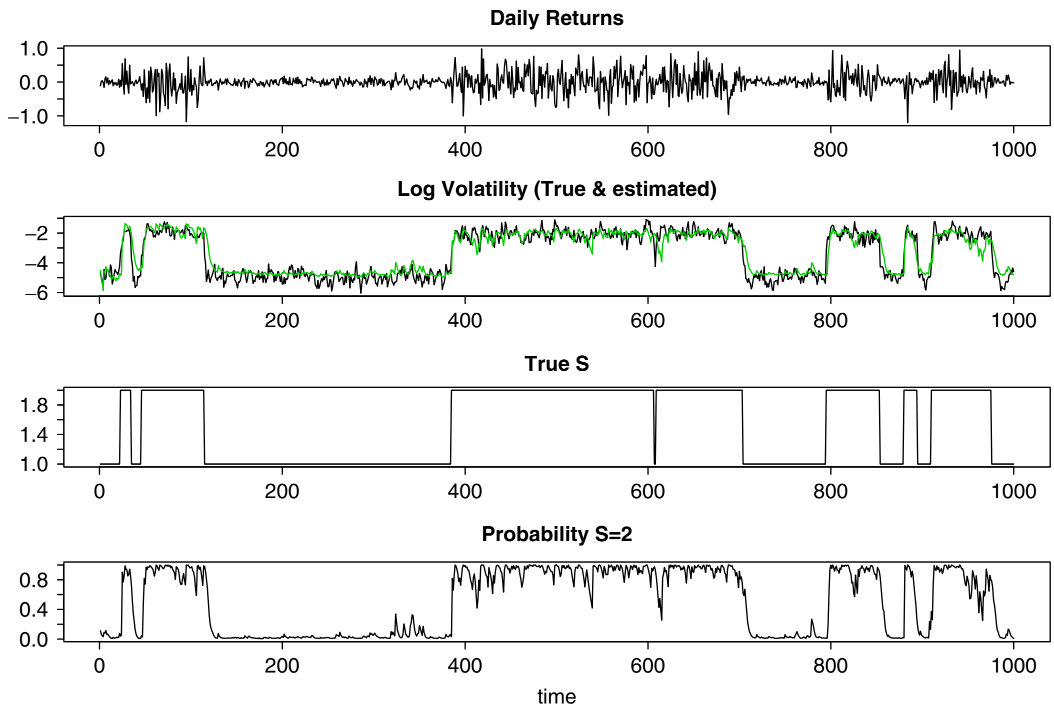

Fig. 3. Simulated data 1 (ϕ=0.5 and p_{11}=0.990): (top graph) simulated time series (y_t), (second graph) true (black line) and estimated log-volatilities ($\hat{E}\left(\lambda_t|D_t\right)$—green line), (third graph) true regime variables (s_t), and (bottom graph) estimated probability that $s_t = 2$, i.e., $\hat{Pr}\ (s_t = 2|D_t)$. In this example the rate of misclassification of s_t is equal to 4.2%.

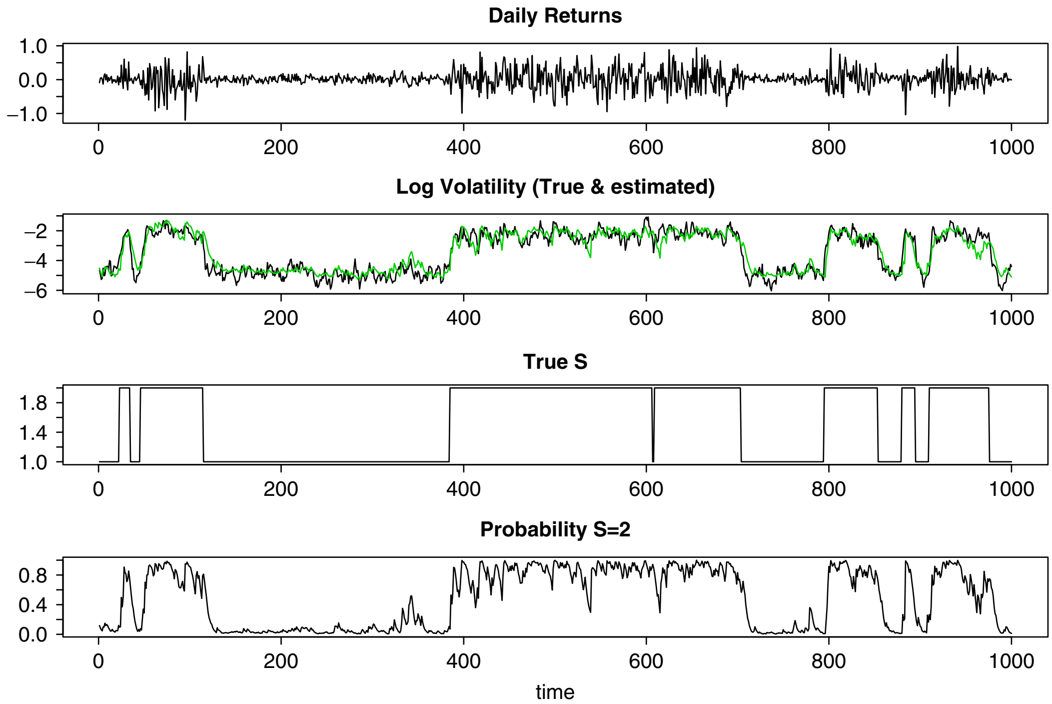

Fig. 4. Simulated data 2 (ϕ=0.7 and p_{11}=0.990): (top graph) simulated time series (y_t), (second graph) true (black line) and estimated log-volatilities ($\hat{E}\left(\lambda_t|D_t\right)$—green line), (third graph) true regime variables (s_t), and (bottom graph) estimated probability that $s_t = 2$, i.e., $\hat{Pr}\ (s_t = 2|D_t)$. In this example the rate of misclassification of s_t is equal to 6.5%.

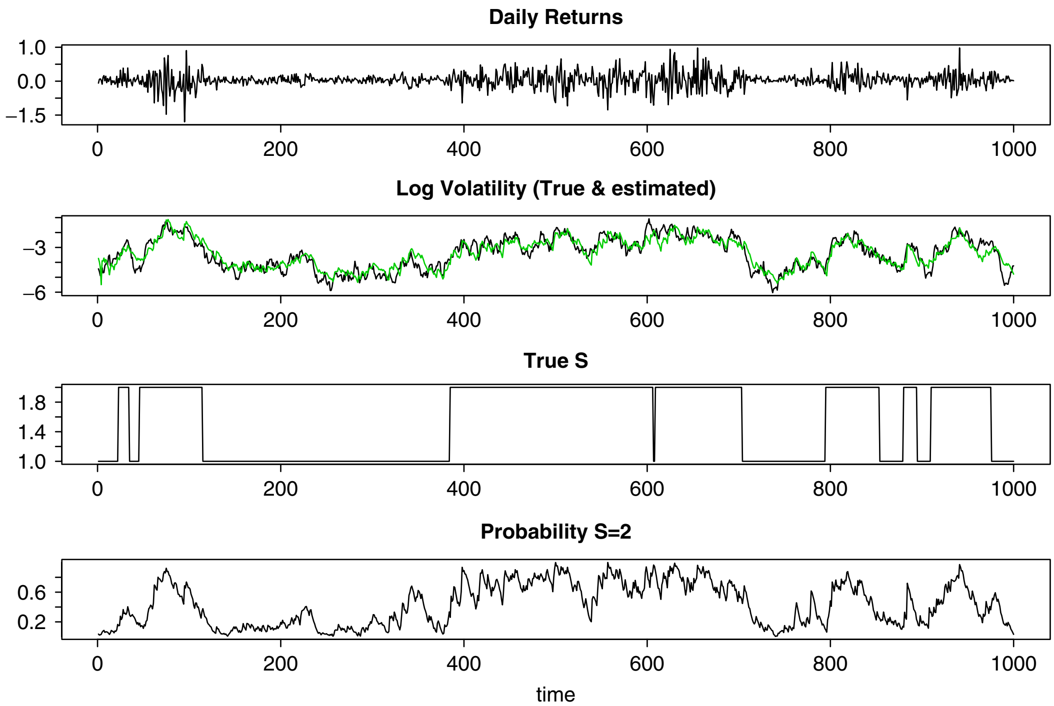

Fig. 5. Simulated data 3 (ϕ=0.9 and p_{11}=0.990): (top graph) simulated time series (y_t), (second graph) true (black line) and estimated log-volatilities ($\hat{E}\left(\lambda_t|D_t\right)$—green line), (third graph) true regime variables (s_t), and (bottom graph) estimated probability that $s_t = 2$, i.e., $\hat{Pr}\ (s_t = 2|D_t)$. In this example the rate of misclassification of s_t is equal to 16.6%.

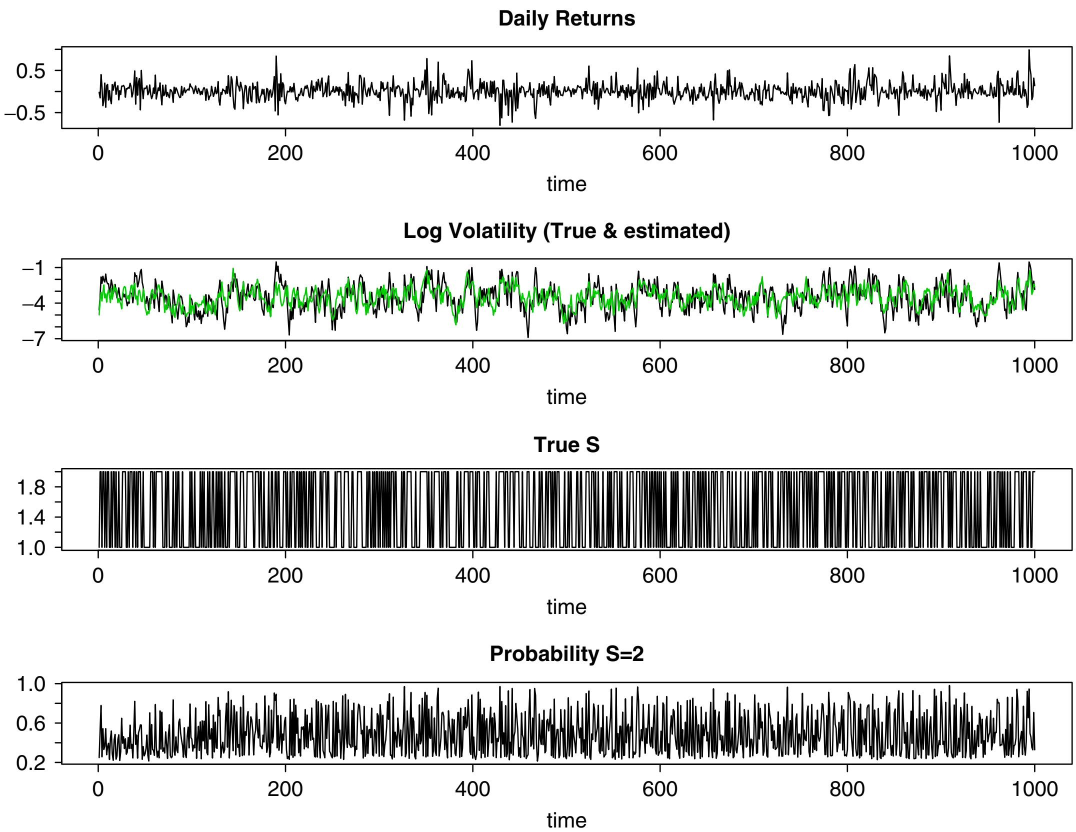

Fig. 6. Simulated data 4 (ϕ=0.9 and p_{11}=0.500): (top graph) simulated time series (y_t), (second graph) true (black line) and estimated log-volatilities ($\hat{E}\ (\lambda_t|D_t)$—green line), (third graph) true regime variables (s_t), and (bottom graph) estimated probability that $s_t = 2$, i.e., $\hat{Pr}\ (s_t = 2|D_t)$. In this example the rate of misclassification of s_t is equal to 39.8%.

Step 3: For $l = 1, \ldots, M$, compute new weights

$$w_{t+1}^{(l)} \propto p\left(y_{t+1}|\lambda_{t+1}^{(l)}\right) / p\left(y_{t+1}|\mu_{t+1}^{(k^l)}\right).$$

Step 4: $\left\{\lambda_{t+1}^{(j)}, s_{t+1}^{(j)}, \theta_{t+1}^{(j)}, w_{t+1}^{(j)}\right\}_{j=1}^{M} \sim p\ (\lambda_{t+1}, s_{t+1}, \theta|D_{t+1})$.

Step 5: Resample.

4. Applications

We apply the sequential Monte Carlo filter to the MSSV model to four synthetic time series and one real dataset. The simulated examples are based on examples from So et al. (1998). As for the real data we analyze the Brazilian stock market index, the Ibovespa.

4.1. Simulation study

Four datasets of size 1000 were simulated from the MSSV model (Eqs. (1) and (2)) with $k = 2$ and parameters as in Table 1. These datasets are based on the examples presented by So et al. (1998). In the three initial examples

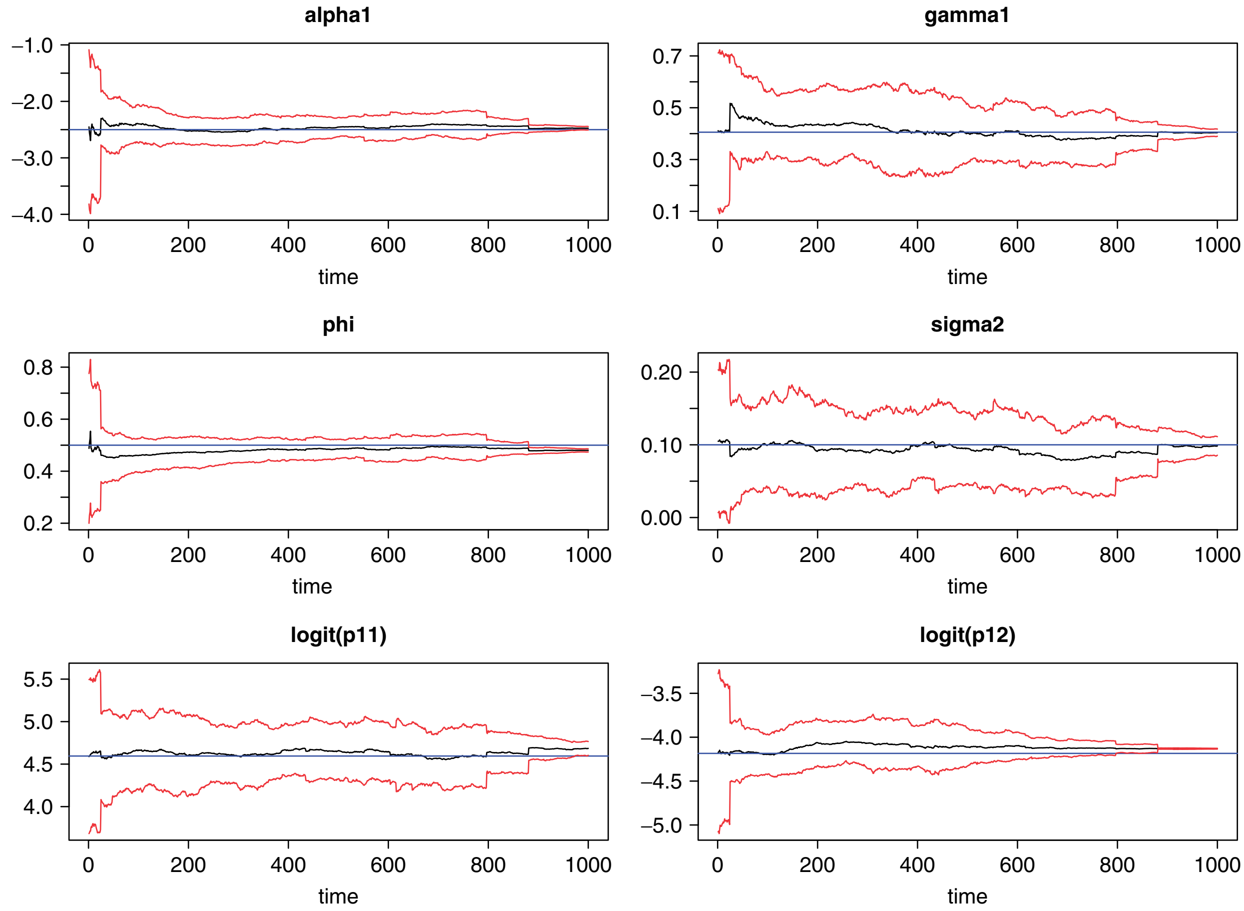

Fig. 7. Posterior mean, 5 and 95% quantiles of θ for the first simulated data ($\phi = 0.5$): α_1, γ_1, ϕ, σ^2, $\log(p_{11}/(1-p_{11}))$, and $\log(p_{21}/(1-p_{21}))$.

most of the masses in the transition probability matrix are concentrated in the diagonal, implying high persistence in each regime representing the occasional changes in the volatility process. In fact, the choice of parameters in these examples coincide with the phenomena observed by Hamilton and Susmel (1994) when exploring the returns of NYSE index by a SWARCH model. The fourth example is presented as a way to highlight the limitations of our estimation procedure as the frequent changes ($p_{11} = 0.5$ and $p_{22} = 0.5$) in the volatility state will be very hard to detect with a sequential estimation procedure. Also, in all examples, the unconditional mean for the volatility process is maintained at the same value: $(\alpha_1, \alpha_2)/(1-\phi) = (-5.0, -2.0)$. By varying ϕ and α_{s_t} accordingly, we are able to study the performance of our SMC algorithm in situations where volatility regimes are well separated or not (Fig. 2).

We initialize the filtering strategy with very diffuse independent draws from a multivariate normal centered at the parameters' and states' true values, summarizing the joint posterior density at time $t = 0$. Using the current notation, we set $M = 3000$. All simulations were performed using a 1.6 GHz Pentium 4 CPU and each filtering iteration took less than 1 s. Pseudo random number were generated using standard Fortran 90 libraries.

Figs. 3–6 show the sequentially predicted log-volatility for all simulated time series along with the true value of log-volatility. They also display the true regimes at time t, s_t, and the predicted probability for the high volatility state, i.e., $\hat{Pr}(s_t = 2|D_t)$. Also, in the caption of each figure, the misclassification rate of s_t estimated based on $\hat{s_t} = \arg\max[Pr(s_t = i)]$ is presented. One can see that in the first three datasets the proposed filtering strategy is handling well the task of sequentially estimating the state vectors λ_t and s_t of the model. Overall, the algorithm is able to correctly identify clusters of high and low volatility and, as one would expect, the further apart the volatility regimes are the easier it is for the filter to flag the regimes. This is made clear by comparing the misclassification rates

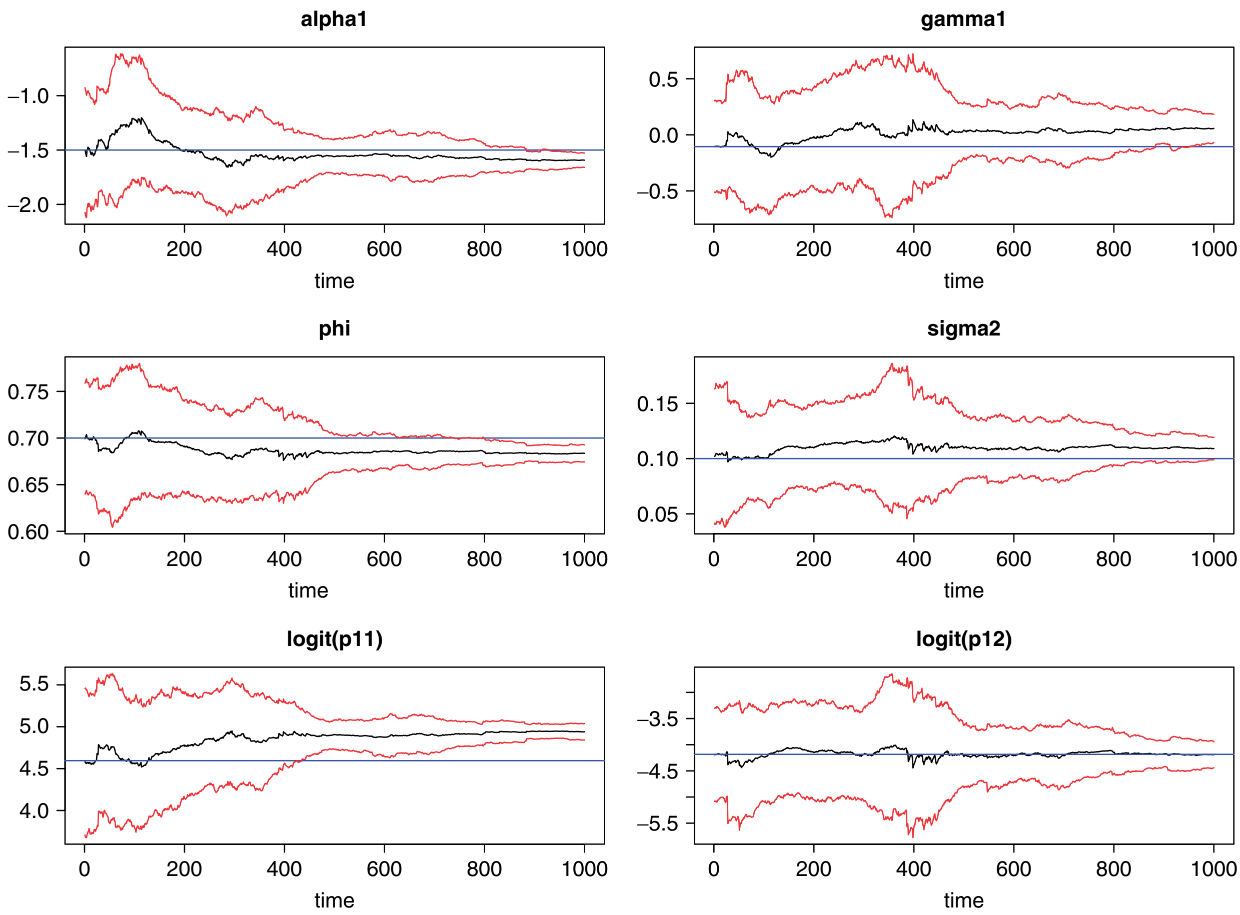

Fig. 8. Posterior mean, 5 and 95% quantiles of θ for the second simulated data (ϕ=0.7): α_1, γ_1, ϕ, σ^2, $\log(p_{11}/(1-p_{11}))$, and $\log(p_{21}/(1-p_{21}))$.

in Dataset 1 and Dataset 3. In the less stable example (Fig. 6) where switches occur all the time, the filter is not as accurate as in the other examples, however, it still does a reasonable job estimating the volatility state λ.

The sequential learning process for the fixed parameters of the model can be seen in Figs. 7–10. These plots show the estimated posterior mean at time t for each parameter together with approximate credible intervals (two standard deviations around the posterior mean), along with the true value for each parameter. Here we can see that even while losing some efficiency along the way, a characteristic shared by all particle filters, the filter is able to correctly estimate the fixed parameters, presented in Table 1.

In all examples the values of a and b were determined by a discount factor $\delta = 0.85$ which implies $a = 0.824$ and $b = 0.566$. As a side note, we explored the impact of the choice of δ in the performance of the filter, and in these examples any δ in (0.50, 0.99) presented results similar to the ones presented here. West and Harrison (1997) argue that δ should be chosen between 0.8 and 0.99 as a function of the amount of information that the modeler is willing to preserve in the filtering process.

Finally, Fig. 11 shows the ratio between the cumulative predictive power of models estimated with $k = 2$ and 1. This is basically a Bayes factor plotted over time to check the predictive performance of a MSSV against a SV. This exercise was mainly intended to justify the use of the MSSV model by showing that if it is true that a time series presents different volatility levels, one should use a SV model that incorporates possible structural changes. This fact is emphasized by Fig. 12 where the cumulative difference in the mean square predictive error for the volatility is presented. In all examples the MSSV performs better than a simple SV model and as noted before the difference is more evident in situations where the volatility regimes are very distinct. It is important to point out that even in Dataset 4 where changes occur very frequently the MSSV model performs better than a simple SV.

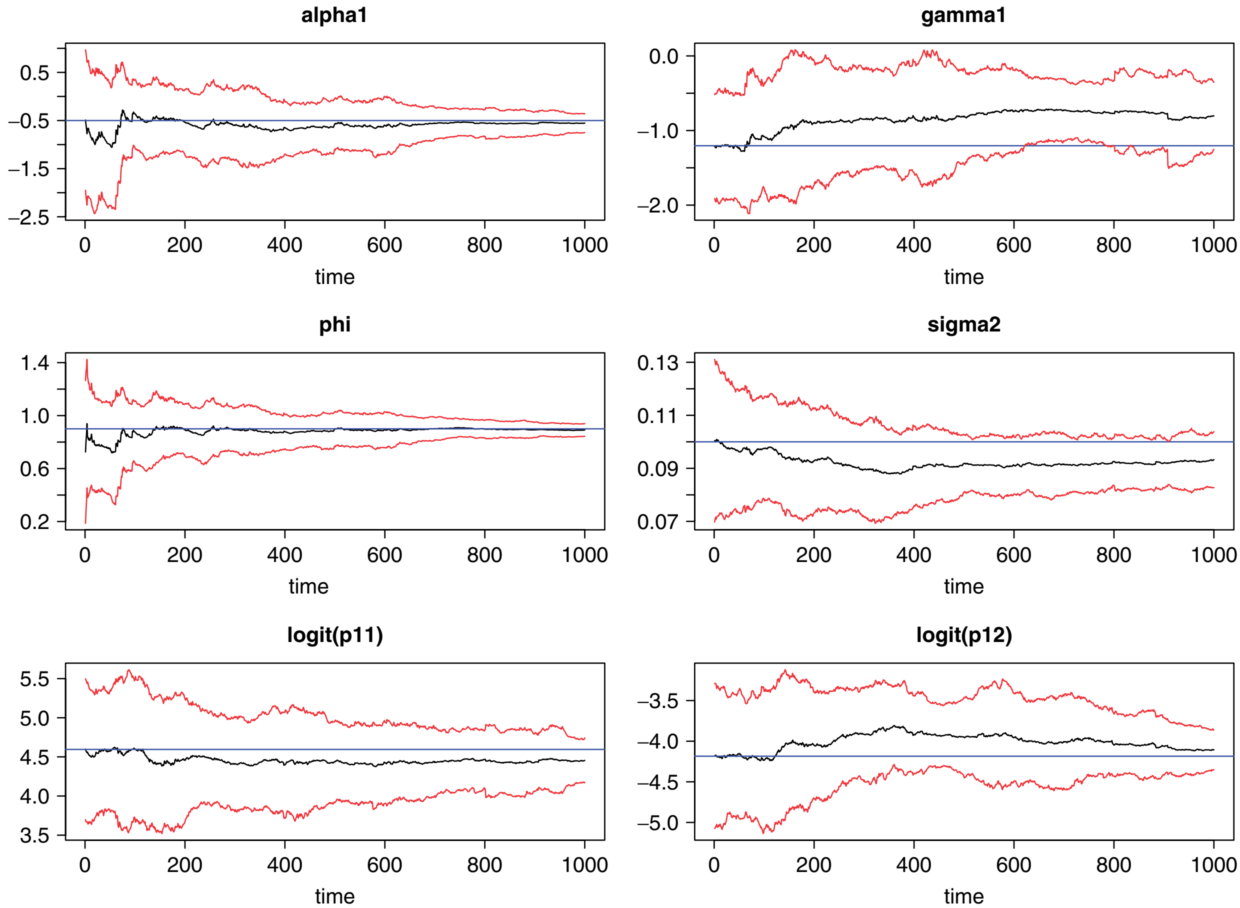

Fig. 9. Posterior mean, 5 and 95% quantiles of θ for the third simulated data ($\phi=0.9$): α_1, γ_1, ϕ, σ^2, $\log(p_{11}/(1-p_{11}))$, and $\log(p_{21}/(1-p_{21}))$.

4.2. Real data: Ibovespa

We now apply the proposed algorithm (with two regimes) to the IBOVESPA stock index (Sao Paulo Stock Exchange) from 01/02/1997 to 01/16/2001 (1000 observations). This period includes a set of currency crises, such as the Asian crisis in 1997, the Russian crisis in 1998 and the Brazilian crisis in 1999 all of which directly affected emerging countries, like Brazil, generating high levels of uncertainty in the markets and consequently high levels of volatility (see Fig. 13). For this reason we decided fit a two regime MSSV model.

To give a practical perspective to this applied example, we set aside a small fraction of the (initial part) of the data for prior elicitation. For this small fraction, we have run So et al.'s (1998) MCMC algorithm and used the draws for the last time period as a sample from θ, x_t for $t=0$ in our sequential analysis.

The Ibovespa appears in Fig. 13 along with the estimated (posterior mean) s_t and λ_t. The vertical lines indicate key market events that identify what agents in the markets would refer to as the beginning and end of the crisis (Table 2). It can be seen that our sequential estimation scheme is able to identify these structural changes in the Ibovespa index by accurately flagging moments of higher volatility through the discrete state prediction.

Fig. 14 shows the sequential estimation of the fixed parameters in the model. It is interesting to point out that in accordance with findings by So et al. (1998), the persistence parameter ϕ is no longer overestimated. By allowing discrete shifts in the volatility level the posterior mean for ϕ is no longer close to one (see Table 3). The diagonal elements of the transition probability matrix for the discrete states are estimated to be high with $E(p_{11}|D_T)=0.993$ and 0.964. This implies that the duration in each regime is quite long with a predominance of the low-volatility regime, which is a fact also encountered by So et al. (1998) when analyzing the US S&P500 series. Fig. 15 compares the

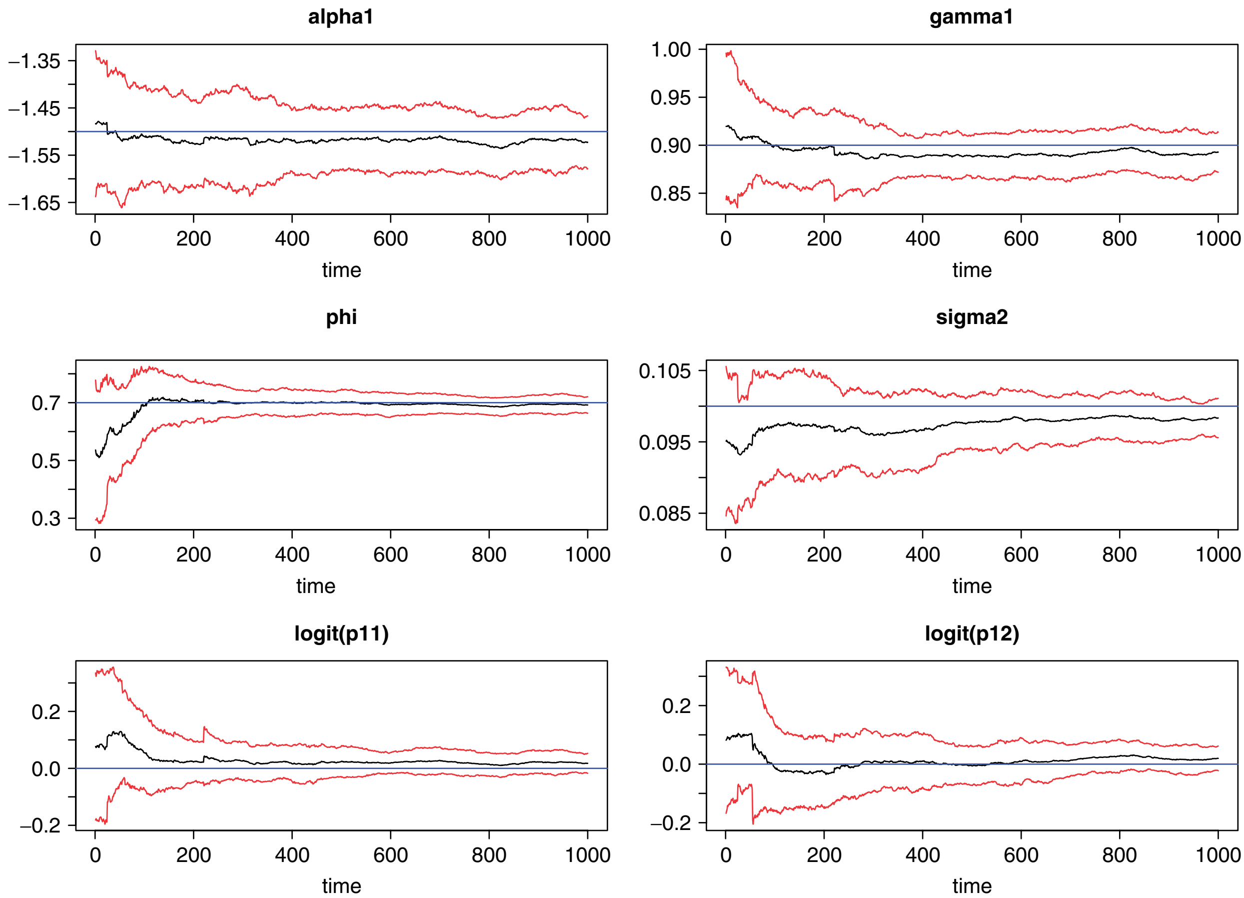

Fig. 10. Posterior mean, 5 and 95% quantiles of θ for the fourth simulated data ($\phi = 0.5$ and $p_{11} = 0.5$): $\alpha_1, \gamma_1, \phi, \sigma^2$, $\log(p_{11}/(1-p_{11}))$, and $\log(p_{21}/(1-p_{21}))$.

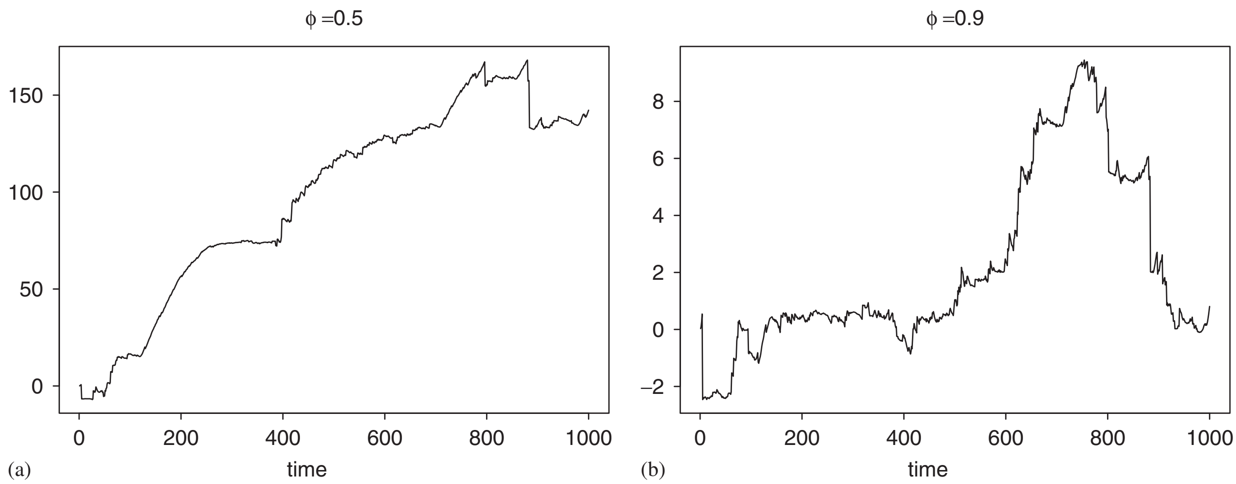

Fig. 11. Sequential Bayes factor: $\log\left(\prod_{\tau=1}^{t} p_1\left(y_\tau|D_{\tau-1}\right)/\prod_{\tau=1}^{t} p_2\left(y_\tau|D_{\tau-1}\right)\right)$ for $t = 1, 2, \ldots, T$. (a) $\phi = 0.5$, (b) $\phi = 0.9$.

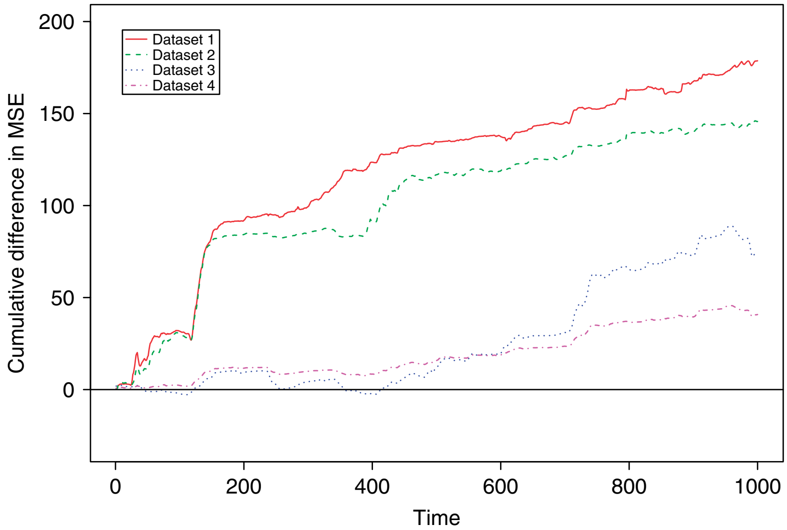

Fig. 12. Cumulative difference in mean square predictive error between the MSSV and SV model in all simulated datasets.

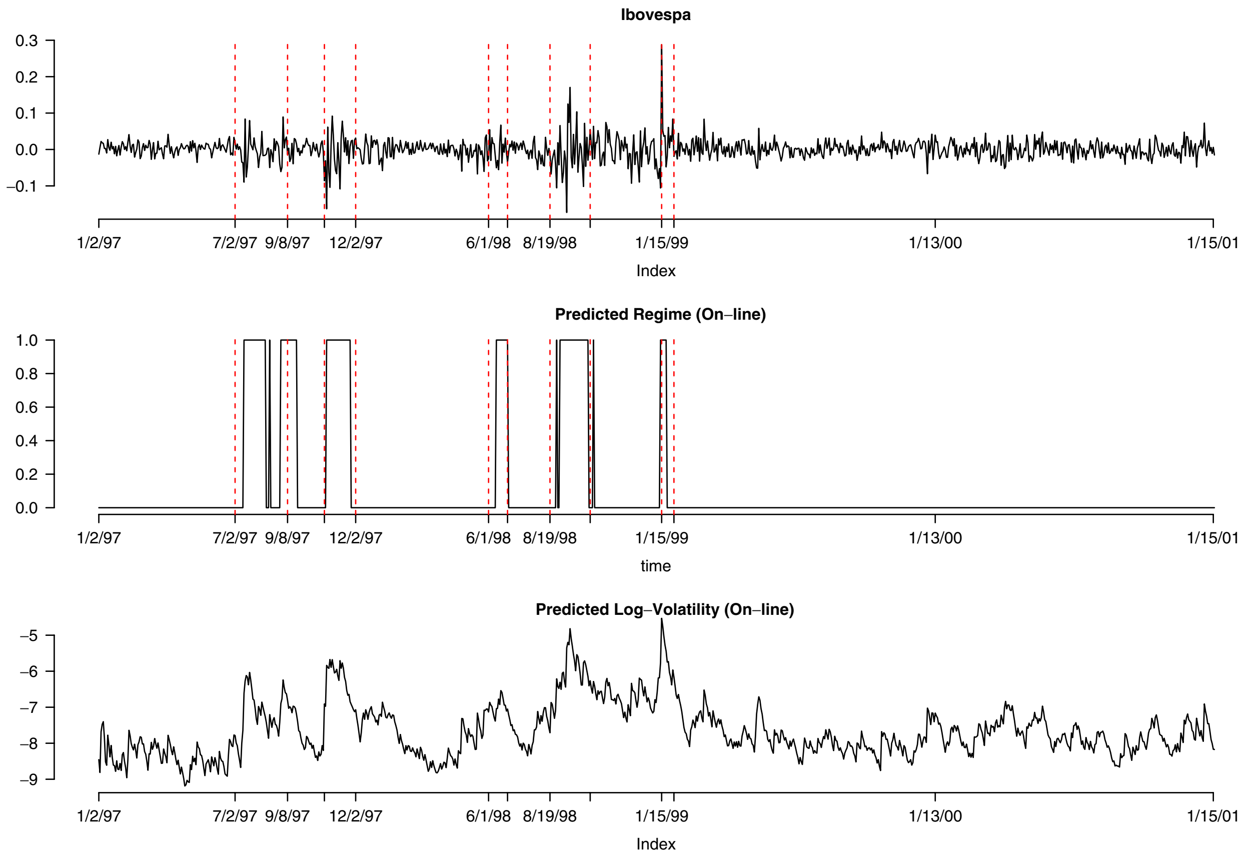

Fig. 13. Ibovespa index (top), estimated (posterior mean) s_t (center), and λ_t (bottom). The vertical lines indicate key market events that identify what agents in the markets would refer to as the beginning and end of the crisis (Table 2).

Table 2
Currency crisis—some key dates

07/02/1997	Thailand devalues the Baht by as much as 20%
08/11/1997	IMF and Thailand set a rescue agreement
10/23/1997	Hong Kong's stock index falls 10.4%. South Korea won starts to weaken
12/02/1997	IMF and South Korea set a bailout agreement
06/01/1998	Russia's stock market crashes
06/20/1998	IMF gives final approval to a loan package to Russia
08/19/1998	Russia officially falls into default
10/09/1998	IMF and World Bank joint meeting to discuss the global economic crisis
	The Fed cuts interest rates
01/15/1999	The Brazilian government allows its currency, the Real, to float freely by lifting exchange controls
02/02/1999	Arminio Fraga is named President of Brazil's Central Bank

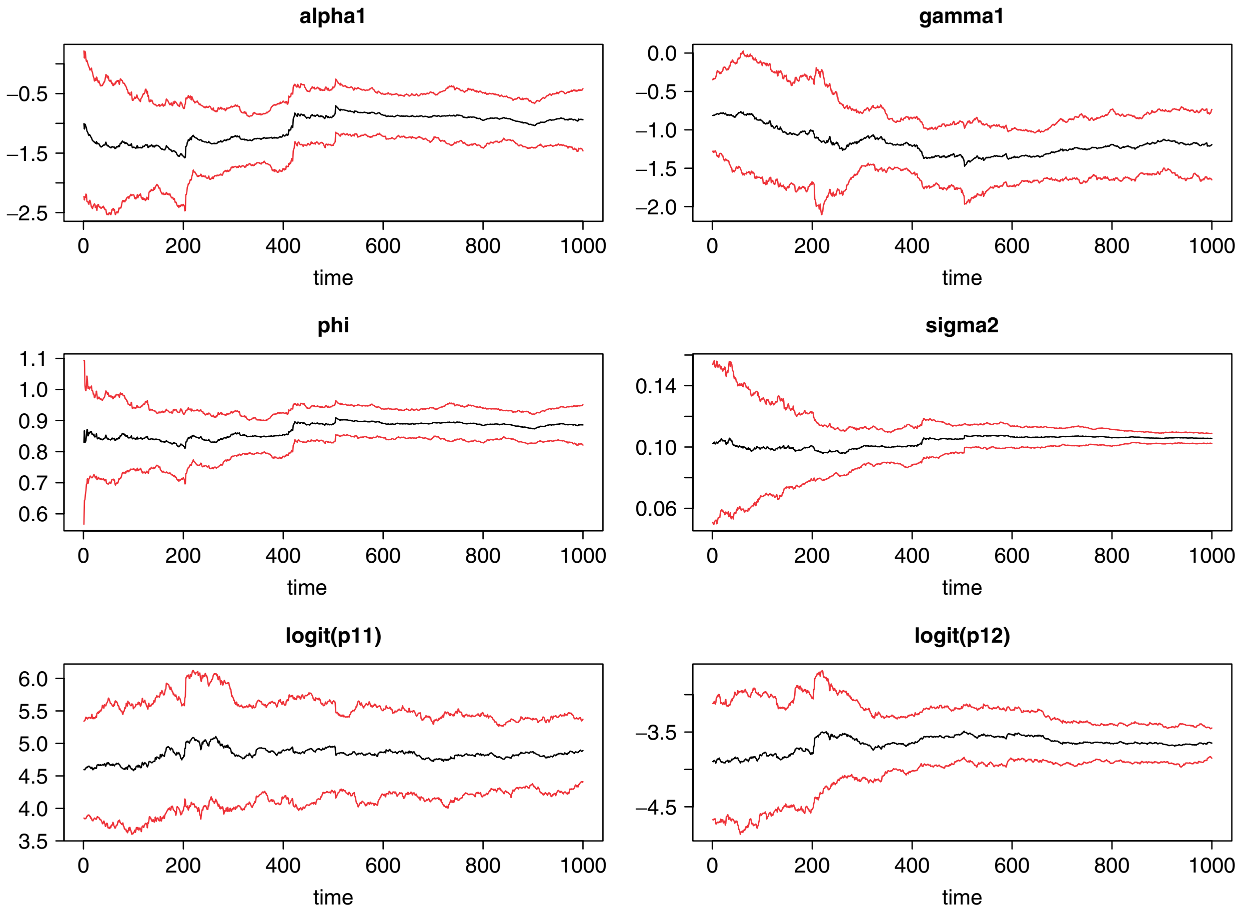

Fig. 14. Posterior mean, 5 and 95% quantiles of θ for the Ibovespa data α_1, $\log(\gamma_1)$, ϕ, σ^2, $\log(p_{11}/(1-p_{11}))$, and $\log(p_{21}/(1-p_{21}))$.

predictive power of the MSSV with $k=2$ against a SV model on the Ibovespa index. The better performance of the MSSV reinforces the theory that financial time series present blocks of high and low volatility and ignoring this fact yields misspecified volatility persistence.

Table 3
Posterior summary for ϕ

Model	95% credible interval	$E(\phi\|D_T)$
SV	(0.9325;0.9873)	0.9525
MSSV	(0.8481;0.8903)	0.8707

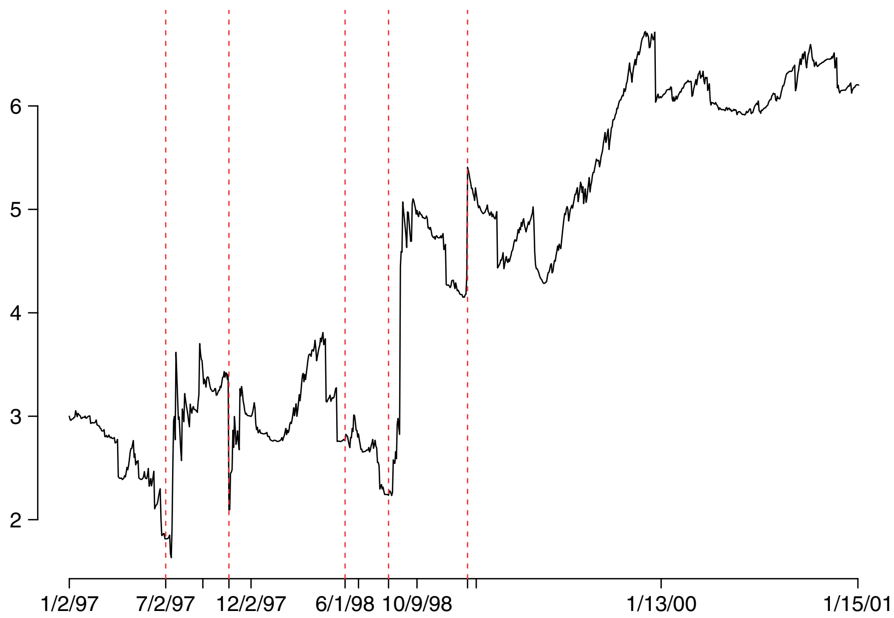

Fig. 15. Ibovespa: Bayes factor—MSSV vs. SV.

5. Final remarks

In this article we developed and implemented a simulation-based sequential algorithm to estimate a univariate Markov switching stochastic volatility (MSSV) model as described in So et al. (1998). The use of simulated examples was intended to show the performance of the proposed method while the Ibovespa example shows its applicability to real market problems.

The simulated examples were able to show the ability of the sequential algorithm to perform accurate sequential inferences. The more distinct the volatility regimes are the better is the performance of the sequential filter presented. The presentation of Dataset 4 with highly unstable volatility regimes aimed to touch on the limitation of the method—this is a situation where not a lot of information is carried from one time point to the next and therefore, a sequential estimation procedure does not perform very well. One should be aware of this problem despite of the fact that this example is not representative of the applications that the MSSV model and the particle filter presented here hope to address. It is important to highlight that we were also able to show, by comparing predictive power, that in all simulated examples the MSSV outperforms a simple stochastic volatility (SV) model ($k = 1$).

In the Ibovespa example we were able to successfully separate moments of high risk from moments that presented a calm trade pattern. We were also able to link the volatility regimes to well defined emerging market crisis. By including structural changes the model no longer overestimates the volatility persistence and the idea of non-stationarity is eliminated. In this context the MSSV ($k = 2$) outperformed a simple SV model in predictive power by allowing sequential predictions to react faster to market events.

References

Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models, in: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4, pp. 2959–3038.

Carlin, B.P., Polson, N.G., Stoffer, D.S., 1992. A Monte Carlo approach to nonnormal and nonlinear state-space modeling. J. Amer. Statist. Assoc. 87, 493–500.

Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.

Chou, R.Y., 1988. Volatility persistence and stock valuations: some empirical evidence using Garch. J. Appl. Econom. 3, 279–294.

Doucet, A., deFreitas, N., Gordon, N., 2001. Sequential Monte Carlo Methods in Practice. Springer, Berlin.

Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. Econometrica 50, 987–1007.

Frühwirth-Schnatter, S., 1994. Data augmentation and dynamic linear models. J. Time Ser. Anal. 15, 183–202.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling based approaches to calculating marginal densities. J. Amer. Statist. Assoc. 85, 398–409.

Gordon, N.J., Salmond, D.J., Smith, A.F.M., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. Comm. Radar Signal Process. 140, 107–113.

Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. J. Econometrics 64, 307–333.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge, MA.

Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models—with comments. J. Business Econom. Statist. 12, 371–417.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH model. Rev. Econom. Stud. 65, 361–393.

Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space model. J. Comput. Graph. Statist. 5, 1–25.

Lamoureux, C., Lastrapes, W., 1990. Persistence in variance, structural change and the Garch model. J. Business Econom. Statist. 8, 225–234.

Liu, J., West, M., 2001. Combined parameter and state estimation in simulation based filtering. In: de Freitas, N., Gordon, N., Doucet, A. (Eds.), Sequential Monte Carlo Methods in Practice. Springer, Berlin.

Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filters. J. Amer. Statist. Assoc. 94, 590–599.

Shephard, N., 1994. Partial non-Gaussian state space. Biometrika 81, 115–131.

Smith, A.F.M., Gelfand, A., 1992. Bayesian statistics without tears: a sampling-resampling perspective. Amer. Statist. 46, 84–88.

So, M.K.P., Lam, K., Li, W.K., 1998. A stochastic volatility model with Markov switching. J. Business Econom. Statist. 16, 244–253.

West, M., 1992. Modelling with mixtures, in: Berger, J.O., Bernardo, J.M., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 4, Oxford University Press, Oxford.

West, M., Harrison, P.J., 1997. Bayesian Forecasting and Dynamic Model. Springer, New York.

West, M., Harrison, P.J., Migon, H.S., 1985. Dynamic generalized linear model and Bayesian forecasting. J. Amer. Statist. Assoc. 80, 73–97.