

Chapter 2

The Extended Liu and West Filter: Parameter Learning in Markov Switching Stochastic Volatility Models

Maria Paula Rios and Hedibert Freitas Lopes

2.1 Introduction

Since the seminal chapter by [Gordon, Salmond and Smith \(1993\)](#) with its *Bootstrap Filter (BF)*, simulation-based sequential estimation tools, commonly known as sequential Monte Carlo (SMC) methods or particle filters (PF), have been receiving increasing attention in its application to nonlinear and non-Gaussian state-space models. There has been a particular emphasis on the application of such methods in state estimation problems in target tracking, signal processing, communications, molecular biology, macroeconomics, and financial time series (see compendium edited by [Doucet, De Freitas and Gordon \(2001\)](#)).

Nonetheless, only recently sequential parameter estimation started to gain more formal attention, with [Liu and West \(2001\)](#) (LW, hereafter) being one of the first contributions to the area. Their main contribution was the generalization of the SMC filter of [Pitt and Shephard \(1999\)](#), namely the *Auxiliary Particle Filter (APF)*. LW incorporated sequential parameter learning in the estimation. Amongst other recent contributions in this direction are the *Practical Filter* of [Polson, Stroud and Muller \(2008\)](#) and the *Particle Learning* scheme of [Carvalho, Johannes, Lopes and Polson \(2010\)](#). The former relies on sequential batches of short MCMC runs while the latter relies on a recursive data augmentation argument, both of which aimed at replenishing the particles for both states and parameters. They also rely on the idea of sequential sufficient statistics for sequential parameter estimation ([Storvik \(2002\)](#) and [Fearnhead, \(2002\)](#)).

Implementation of the LW filter in various disciplines has shown that this methodology produces degenerate parameter estimates as discussed in [Carvalho et al. \(2010\)](#). Here we use volatility models to evidence the latter. One appreciates that the LW parameter estimates collapse to a point as further discussed

M.P. Rios • H.F. Lopes (✉)

Booth School of Business, University of Chicago 5807 S Woodlawn Avenue,
Chicago, IL 60637, USA

e-mail: maria@chicagobooth.edu; hlopes@chicagobooth.edu

(see Figs. 2.4 and 2.5). Parameter degeneracy limits the applicability of the LW methodology. In particular, without proper parameter estimates one cannot make accurate forecasts, which are desired in many of the applications where filters are implemented.

To overcome the limitations of the LW filter, we explore three more filters of similar nature. Using the APF and BF as starting points for the propagation and re-sampling of the latent state, we incorporate sequential parameter learning techniques to extend these two filters to accommodate for parameter estimation. The first algorithm relies on the kernel smoothing idea that LW present when introducing their filter (see [Liu and West \(2001\)](#)). The second methodology relies on parameter estimation via recursive computation of conditionally sufficient statistics. In short, we construct four filters¹ that are hybrids between the BF, APF, kernel smoothing, and sufficient statistics.

Throughout the chapter we emphasize our analysis on two filters of particular interest, the LW filter and the so-called APF + SS filter. The latter is the extension of the APF filter that incorporates conditional sufficient statistics (SS) in the fixed parameter estimation.

To highlight the shortcomings of the LW filter and the applicability and improvements the APF + SS filter and the other two filters introduced, we focus only on one of the many applications where this technique is relevant. In this chapter we revisit the work of [Carvalho and Lopes \(2007\)](#). They used the LW filter SMC for state filtering and sequential parameter estimation in Markov switching stochastic volatility (MSSV) models. Using [Carvalho and Lopes \(2007\)](#) as reference, we implement the filters to the estimation of MSSV models. We empirically show, using simulated and real data, that LW filter degenerates, has larger Monte Carlo error, and in general terms underperforms when compared to the other filters of interest.

2.1.1 Volatility Models

Bayesian filters are a general technique that have a broad application scope. As shown in [Carvalho *et al.* \(2010\)](#), particle learning techniques can be implemented in Gaussian dynamic linear models (GDLM) and conditional dynamic linear models (CDLM). In this chapter, however, we focus only on one of the possible applications of the filters of interest. In particular we estimate fixed parameters and latent states in MSSV models.

Over the years, stochastic volatility models have been considered a useful tool for modeling time-varying variances, mainly in financial applications where agents are constantly facing decisions dependent on measures of volatility and risk. Bayesian estimation of stochastic volatility models can be found [Jacquier *et al.* \(1994\)](#) and [Kim *et al.* \(1998\)](#). Comprehensive reviews of stochastic volatility models can be found in [Ghysels *et al.* \(1996\)](#).

¹ Two of the filters we construct have been previously described by [Liu and West \(2001\)](#) and [Storvik \(2002\)](#).

2.1.1.1 Log-Stochastic Volatility

The building block for the MSSV models is the standard univariate log-stochastic volatility model, SV hereon, (see, for example, [Jacquier *et al.* \(1994\)](#), or [Ghysels *et al.* \(1996\)](#)), where (log) returns r_t and log-volatility states λ_t follow a state-space model of the form,

$$r_t = \exp\{\lambda_t/2\}\varepsilon_t \quad (2.1)$$

$$\lambda_t = \alpha + \eta\lambda_{t-1} + \tau\eta_t \quad (2.2)$$

where the errors, ε_t and η_t , are independent standard normal sequences. We also assume the initial log-volatility follows $\lambda_0 \sim N(m_0, C_0)$. The parameter vector, θ_{sv} , consists of the volatility mean reversion parameters $\psi = (\alpha, \eta)$ and the volatility of volatility τ . It is worth mentioning that the model assumes conditional independence of the $r_t, t = 1, \dots, T$ variables.

2.1.1.2 Markov Switching Stochastic Volatility

Jumps have been a broadly studied characteristic of financial data (see, for example, [Eraker *et al.* \(2003\)](#)). [So *et al.* \(1998\)](#) suggest a model that allows for occasional discrete shifts in the parameter determining the level of the log-volatility through a Markovian process. They claim that this model not only is a better way to explain volatility persistence but is also a tool to capture changes in economic forces, as well as abrupt changes due to unusual market forces.

[So *et al.*'s \(1998\)](#) approach generalizes the SV model to include jumps by allowing the state space equation to follow a process that changes according to an underlying regime that determines the parameters for λ_t . For this, assume that s_t is an unobserved discrete random variable with domain $\{1, 2, \dots, k\}$. Assuming a k -state first-order Markov process, we define the transition probabilities as

$$p_{j,l} = P(s_t = l | s_{t-1} = j) \quad \text{for } j, l = 1, \dots, k \quad (2.3)$$

with $\sum_{j=1}^k p_{ij} = 1$ for $i = 1, \dots, k$. As suggested in [So *et al.* \(1998\)](#) (2.2) can be generalized to include such regime changes in the α parameter. [Carvalho and Lopes \(2007\)](#) suggest that α in this model corresponds to the level of the log-volatility and in order to allow occasional changes the model introduces different values α 's following the described first-order Markovian process.

Again, let r_t be the observed process, just like it was defined for the SV model, with observations r_1, \dots, r_T conditionally independent and identically distributed. To keep consistency with the previously defined SV model, same notation and the normality and independence assumptions on the error terms will also be used here. This means that the observation $r_t, t = 1, \dots, T$ is normal with time-varying log-volatilities $\lambda_1, \dots, \lambda_T$. More specifically,

$$r_t = \exp\{\lambda_t/2\}\varepsilon_t \quad (2.4)$$

$$\lambda_t = \alpha_{s_t} + \eta\lambda_{t-1} + \tau\eta_t \quad (2.5)$$

Let $\xi = (\alpha, \eta, \tau^2)$, $\alpha = (\alpha_1, \dots, \alpha_k)$, $p = (p_{11}, p_{1,2}, \dots, p_{1,k-1}, \dots, p_{k,1}, \dots, p_{k,k-1})$, then $\theta_{\text{MSSV}} = (\xi, p)$ is the set of $(k^2 + 2)$ parameters to estimate at each point in time. For instance, in a two-state model, six parameters must be estimated. It is common in the literature to refer to $S = (s_1, \dots, s_T)$ and $\lambda = (\lambda_1, \dots, \lambda_T)$ as the states of the model. The initial value of λ , λ_0 , is $N(m_0, C_0)$.

To avoid identification issues in α , [So et al. \(1998\)](#) suggest to re-parameterize it as

$$\alpha_{s_i} = \gamma_1 + \sum_{j=1}^k \gamma_j I_{ji} \quad (2.6)$$

where $I_{ji} = 1$ when $s_i \geq j$ and 0 otherwise, $\gamma_1 \in R$ and $\gamma_i > 0$ for all $i > 1$. The model described by (2.4)–(2.6) is known as an MSSV model. As previously discussed the case where $k = 1$ reduces to the SV model presented above.

In this chapter we explore to cases of the MSSV model, $k = 1$ (or log-stochastic volatility) and $k = 2$. We fit these two models to the simulated and real data that we explore in Sects. 2.3 and 2.4.

2.1.2 Particle Filters: A Brief Review

Particle filters are SMC methods that basically rely on a *sampling importance re-sampling* (SIR) argument in order to sequentially reweigh and/or resample particles as new observations arrive. More specifically, let the general state-space model be defined by

$$\text{Observation equation : } p(y_{t+1}|x_{t+1}) \quad (2.7)$$

$$\text{State equation : } p(x_{t+1}|x_t) \quad (2.8)$$

where, for now, all static parameters are kept known. The observed variables y_t and the latent state variables x_t can be univariate or multivariate, discrete or continuous. Nonetheless, for didactical reasons, we will assume both are continuous scalar quantities.

Particle filters aim at computing/sampling from the filtering density²

$$p(x_{t+1}|y^t) = \int p(x_{t+1}|x_t, y^t) p(x_t|y^t) dx_t \quad (2.9)$$

and computing/sampling the posterior density via Bayes' theorem

$$p(x_{t+1}|y^{t+1}) \propto p(y_{t+1}|x_{t+1}) p(x_{t+1}|y^t) \quad (2.10)$$

² To avoid confusion, y^t makes reference to all the data observed up to point t , while y_t refers to the data observation at time t .

Put simply, PFs are Monte Carlo schemes whose main objective is to obtain draws $\{x_{t+1}^{(i)}\}_{i=1}^N$ from the state posterior distribution at time $t + 1$, $p(x_{t+1}|y^{t+1})$, when the only draws available are $\{x_t^{(i)}\}_{i=1}^N$ from the state posterior distribution at time t , $p(x_t|y^t)$.

Recent reviews of PFs are [Lopes and Tsay \(2011\)](#), [Olsson et al. \(2008\)](#), [Doucet and Johansen \(2010\)](#), and [Lopes and Carvalho \(2011\)](#).

In what follows we briefly review two of the most popular filters for situations where static parameters are known. The BF or *sequential importance sampling with resampling* (SISR) filter and the APF. For these filters we assume that $\{x_0^{(i)}\}_{i=1}^N$ is a sample from $p(x_0|y^0)$.

2.1.2.1 The Bootstrap Filter

[Gordon, Salmond and Smith's \(1993\)](#) seminal filter basically uses the transition (2.8) in order to propagate particles, which then are resampled from the model (2.7). The BF can be summarized in the following two steps:

1. *Propagation.* Particles $\tilde{x}_{t+1}^{(i)}$ are drawn from $p(x_{t+1}|x_t^{(i)})$, for $i = 1, \dots, N$, so the particle set $\{\tilde{x}_{t+1}^{(i)}\}_{i=1}^N$ approximates the filtering density $p(x_{t+1}|y^t)$ from (2.9).
2. *Resampling.* The SIR argument converts prior draws into posterior draws by resampling from $\{\tilde{x}_{t+1}^{(i)}\}_{i=1}^N$ with weights proportional to the likelihood, $\omega_{t+1}^{(i)} \propto p(y_{t+1}|\tilde{x}_{t+1}^{(i)})$, for $i = 1, \dots, N$.

If the resampling step is replaced simply by a reweighting step, then the weights are replaced by $\omega_{t+1}^{(i)} \propto \omega_t^{(i)} p(y_{t+1}|\tilde{x}_{t+1}^{(i)})$, where $\omega_0^{(i)}$ is usually set at $1/N$. The SIR scheme samples from the prior and avoids the potentially expensive and/or practically intractable task of point-wise evaluation of $p(x_{t+1}|y^t)$. The flexibility and generality that comes with this *blind* scheme is the usually unbearable price of high Monte Carlo errors. More importantly, it leads to *particle degeneracy*, a Monte Carlo phenomenon where, after a few recursions of steps 1 and 2 above, all particles collapse into a few points and eventually to one single point.

2.1.2.2 The Auxiliary Particle Filter

One of the first *unblinded* filters was proposed by [Pitt and Shephard \(1999\)](#), who resample old draws with weights proportional to the proposal or candidate density $p(y_{t+1}|g(x_t))$, for some function g , such as the mean or the mode of the evolution density, and then propagate such draws via the evolution equation in (2.9). Finally, propagated draws are resampled with weights given by step 2 below. Their argument is based on a Monte Carlo approximation to

$$p(x_{t+1}|y^{t+1}) = \int p(x_{t+1}|x_t, y^{t+1}) p(x_t|y^{t+1}) dx_t \quad (2.11)$$

which is based on the one-step smoothing density $p(x_t|y^{t+1})$. Pitt and Shephard's (1999) APF can be summarized in the following three steps:

1. *Resampling.* The set $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ is sampled from $\{x_t^{(i)}\}_{i=1}^N$ with weights $\{\pi_{t+1}^{(i)}\}_{i=1}^N$, where $\pi_{t+1}^{(i)} \propto p(y_{t+1}|g(x_t^{(i)}))$.
2. *Propagation.* The transition equation $p(x_{t+1}|\tilde{x}_t^{(i)})$ is used to draw $\tilde{x}_{t+1}^{(i)}$, whose corresponding weight is $\omega_{t+1}^{(i)} \propto p(y_{t+1}|\tilde{x}_{t+1}^{(i)})/p(y_{t+1}|g(\tilde{x}_t^{(i)}))$, for $i = 1, \dots, N$.
3. *Posterior draws.* The set $\{x_{t+1}^{(i)}\}_{i=1}^N$ is sample from $\{\tilde{x}_{t+1}^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.

Our main contributions are twofold. Firstly, by comparing the four filters of interest we highlight the limitations of the LW-type filters for two cases of MSSV models. Secondly, we introduce an extension of the APF filter to overcome such limitations and produce more accurate estimates. The remainder of the chapter is organized as follows. In the next section we introduce the sequential parameter learning strategies that we then incorporate in the two filters previously discussed to extend them to allow for parameter estimation (the LW filter is one of such extensions). Results are analyzed in two sections. Section 2.3 presents and analyzes all the simulated data study while Sect. 2.4 presents real data applications. Section 2.5 concludes.

2.2 Particle Filters with Parameter Learning

We extend the BF and APF filtering strategies to allow for fixed parameter learning. Incorporating two techniques to each of the filters we study the four resulting types of Bayesian filters, which will be compared and evaluated in order to determine which filter outperforms the rest.

2.2.1 Kernel Smoothing

The first strategy that we incorporate for fixed parameter estimation is kernel smoothing, KS hereon, that was introduced in Liu and West (2001).

Liu and West (2001) generalizes the Pitt and Shephard's (1999) APF to accommodate sequential parameter learning. They rely on West's (1993) mixture of normals argument, which assumes that, for a fixed parameter vector θ ,

$$p(\theta|y^t) \approx \sum_{i=1}^N f_N(\theta; m_t^{(i)}, h^2 V_t) \quad (2.12)$$

where $f_N(\theta; a, b)$ is the density of a multivariate normal with mean a and variance-covariance matrix b evaluated at θ . $\{\theta_t^{(i)}\}_{i=1}^N$ approximates $p(\theta|y^t)$, V_t approximates

the variance of θ given y^t , h^2 is a smoothing factor, and $m_t(\theta^{(i)}) = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$ for $\bar{\theta}_t$ an approximation to the mean of θ given y^t and a a shrinkage factor, usually associated with h through $h^2 = 1 - a^2$.

The performance of filters that implement a KS strategy depends on the choice of the tuning parameter a , which drives both the shrinkage and the smoothness of the normal approximation. It is common practice to use a around 0.98 or higher. Also, the normal approximation can be easily adapted to other distributions, such as the normal-inverse-gamma approximation for conditionally conjugate location-scale models.

2.2.1.1 APF + KS: LW Filter

The first filter we consider is the so-called LW filter. This filter incorporates the KS strategy to the APF filter (see [Liu and West \(2001\)](#)). This is the filter used by [Carvalho and Lopes \(2007\)](#) to sequentially learn about parameters and state in a MSSV model.

Let the particle set $\{(x_t, \theta_t)^{(i)}\}_{i=1}^N$ approximate $p(x_t, \theta | y^t)$, $\bar{\theta}_t$ and V_t estimates of the posterior mean and posterior variance of θ , respectively, with $g(x_t^{(i)}) = E(x_{t+1} | x_t^{(i)}, m(\theta_t^{(i)}))$ and $m_t(\theta^{(i)}) = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$, for $i = 1, \dots, N$. The LW filter can be summarized in the following three steps:

1. *Resampling*. The set $\{(\tilde{x}_t, \tilde{\theta}_t)^{(i)}\}_{i=1}^N$ is sampled from $\{(x_t, \theta_t)^{(i)}\}_{i=1}^N$ with weights $\{\pi_{t+1}^{(i)}\}_{i=1}^N$, where $\pi_{t+1}^{(i)} \propto p(y_{t+1} | g(x_t^{(i)}), m(\theta_t^{(i)}))$;
2. *Propagation*. For $i = 1, \dots, N$,
 - (a) *Propagating θ* . Sample $\tilde{\theta}_{t+1}^{(i)}$ from $N(m(\tilde{\theta}_t^{(i)}), h^2 V_t)$,
 - (b) *Propagating x_{t+1}* . Sample $\tilde{x}_{t+1}^{(i)}$ from $p(x_{t+1} | \tilde{x}_t^{(i)}, \tilde{\theta}_{t+1}^{(i)})$,
 - (c) *Computing weights*. $\omega_{t+1}^{(i)} \propto p(y_{t+1} | \tilde{x}_{t+1}^{(i)}, \tilde{\theta}_{t+1}^{(i)}) / p(y_{t+1} | g(\tilde{x}_t^{(i)}), m(\tilde{\theta}_t^{(i)}))$;
3. *Posterior draws*. The set $\{(x_{t+1}, \theta_{t+1})^{(i)}\}_{i=1}^N$ is sampled from $\{(\tilde{x}_{t+1}, \tilde{\theta}_{t+1})^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.

2.2.1.2 BF + KS

The second filter that we analyze in this chapter is the extension of the BF when we include the KS strategy in the fixed parameter estimation. The following algorithm summarizes the BF + KS filter.

Let the particle set $\{(x_t, \theta_t)^{(i)}\}_{i=1}^N$ approximate $p(x_t, \theta | y^t)$, $\bar{\theta}_t$ and V_t estimates of the posterior mean and posterior variance of θ , respectively, with $g(x_t^{(i)}) = E(x_{t+1} | x_t^{(i)}, m(\theta_t^{(i)}))$ and $m_t(\theta^{(i)}) = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$, for $i = 1, \dots, N$.

1. *Propagation.* For $i = 1, \dots, N$,

(a) *Propagating x_{t+1} .* Sample $\tilde{x}_{t+1}^{(i)}$ from $p(x_{t+1} | \tilde{x}_t^{(i)}, \tilde{\theta}_t^{(i)})$.

(b) *Propagating θ .* Sample $\tilde{\theta}_{t+1}^{(i)}$ from $N(m(\tilde{\theta}_t^{(i)}), h^2 V_t)$.

(c) *Computing weights.* $\omega_{t+1}^{(i)} \propto p(y_{t+1} | \tilde{x}_{t+1}^{(i)}, \tilde{\theta}_{t+1}^{(i)})$.

2. *Posterior draws.* The set $\{(x_{t+1}, \theta_{t+1})^{(i)}\}_{i=1}^N$ is sampled from $\{(\tilde{x}_{t+1}, \tilde{\theta}_{t+1})^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.

2.2.2 Sufficient Statistics

The second method that we consider for sequential parameter learning is the recursive sufficient statistics, SS hereon. This technique can be implemented in situations where the vector of fixed parameters θ admits recursive conditional sufficient statistics (Storvik, (2002), and Fearnhead, (2002)). That is the prior for θ is

$$p(\theta) = p(\theta | s_0) \quad (2.13)$$

One of the main advantages of this estimation technique is that Monte Carlo error is reduced by decreasing the number of parameters in Liu and West's kernel mixture approximation. In addition, tracking sufficient statistics can be seen as replacing the sequential estimation of fixed parameters by the sequential updating of a low-dimensional vector of deterministic states. This is particularly important when sequentially learning about variance parameters. See Carvalho *et al.* (2010) for further discussion. Furthermore, this methodology reduces the variance of the sampling weights, resulting in algorithms with increased efficiency and helps delaying the decay in the particle approximation often found in algorithms based on SIR.³

³ As an illustration, we present the SS for the MSSV $k = 2$ model.

The following two equations define the model.

$$\begin{aligned} r_t | \lambda &\sim N(0, e^{\lambda}) \\ \lambda_t | \lambda_{t-1}, \alpha, \eta, \gamma &\sim N(\alpha + \eta \lambda_{t-1} + \gamma s_t \sim N(\alpha + \eta \lambda_{t-1} + \gamma s_t, \tau^2) \end{aligned}$$

Priors and hyperparameter values are defined in Sect. 2.3. Let $x't = (\mathbf{1}, \lambda_{t-1}, s_t)$ and $\theta = (\alpha, \eta, \gamma)$. Therefore, conjugacy leads to

$$\begin{aligned} (\alpha, \eta, \gamma) | \tau^2, x)_{1:t} &\sim N(a_t, \tau A_t) \mathbf{1}_{\gamma > 0} \\ \tau^2 | x_{1:t} &\sim IG(v_t/2, v_t \tau^2/2) \\ p_{1,1} | s_{1:t} &\sim Beta(n_{11t}, n_{12t}) \\ p_{2,2} | s_{1:t} &\sim Beta(n_{21t}, n_{22t}) \end{aligned}$$

2.2.2.1 APF + SS

The main filter that we showcase in this chapter is the SS extension to the APF filter which we describe below. We call this filter the APF+SS filter and will show its ability to overcome many of the limitations present in other filtering strategies, like the LW filter.

Let the particle set $\{(x_t, \theta, s_t)^{(i)}\}_{i=1}^N$ approximate $p(x_t, \theta, s_t | y^t)$ with $g(x_t^{(i)}) = E(x_{t+1} | x_t^{(i)})$. The APF + SS can be summarized as follows:

1. *Resampling.* The set $\{(\tilde{x}_t, \tilde{\theta}, \tilde{s}_t)^{(i)}\}_{i=1}^N$ is sampled from $\{(x_t, \theta, s_t)^{(i)}\}_{i=1}^N$ with weights $\{\pi_t^{(i)}\}_{i=1}^N$, where $\pi_t^{(i)} \propto p(y_{t+1} | g(x_t^{(i)}))$.
2. *Propagation.* For $i = 1, \dots, N$,
 - (a) *Propagating x_{t+1} .* Sample $\tilde{x}_{t+1}^{(i)}$ from $p(x_{t+1} | \tilde{x}_t^{(i)}, \tilde{\theta}^{(i)})$.
 - (b) *Computing weights.* $\omega_{t+1}^{(i)} \propto p(y_{t+1} | \tilde{x}_{t+1}^{(i)}, \tilde{\theta}^{(i)}) / p(y_{t+1} | g(\tilde{x}_t^{(i)}), \tilde{\theta}^{(i)})$.
3. *Posterior draws.* The set $\{(x_{t+1}, \theta, s_t)^{(i)}\}_{i=1}^N$ is sampled from $\{(\tilde{x}_{t+1}, \tilde{\theta}, \tilde{s}_t)^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.
4. *Update sufficient statistics.* $s_{t+1}^{(i)} = \mathcal{S}(\tilde{s}_t^{(i)}, \tilde{x}_{t+1}^{(i)}, y_{t+1})$, for $i = 1, \dots, N$.
5. *Parameter learning.* Sample $\theta^{(i)} \sim p(\theta | s_{t+1}^{(i)})$, for $i = 1, \dots, N$.

Both APF + SS and particle learning algorithms (PL) presented in [Carvalho et al. \(2010\)](#) are particle filters that resample old particles first and then propagate them. While PL and APF + SS are quite similar when dealing with the parameter sufficient statistics, PL approximates the $\log \chi^2$ distribution of log squared returns by [Kim, Shephard and Chib's \(1998\)](#) mixture of seven normal densities while APF + SS uses [Pitt and Shephard's \(1999\)](#) APF that approximates the predictive density with the likelihood function. Further investigation comparing these algorithms for more general classes of stochastic volatility models is an open area beyond the scope of this chapter.

2.2.2.2 BF + SS

The last filter that we consider in this chapter is the SS extension to the BF, as suggested in [Storvik \(2002\)](#). What we will refer to as the BF + SS filter can be summarized with the following steps:

where $x_{1:t} = \{x_1, \dots, x_t\}$ and $v_t = v_{t-1} + 1$. $A_t, a_t, v_t \tau^2$ and $n_{ij,t}$ are the sufficient statistics defined recursively by

$$\begin{aligned}
 A_t^{-1} &= A_{t-1}^{-1} + x_t x_t' \\
 A_t^{-1} a_t &= A_{t-1}^{-1} a_t + x_t \lambda_t \\
 v_t \tau^2 &= v_{t-1} \tau_{t-1}^2 + \lambda_t^2 - x_t' a_t \lambda_t + a_{t-1}' A_{t-1}^{-1} a_{t-1} - a_t' A_{t-1}^{-1} a_{t-1} \\
 n_{ij,t} &= n_{ij,t-1} + \mathbf{1}_{(s_{t-1}=i, s_t=j)}
 \end{aligned}$$

Let the particle set $\{(x_t, \theta, s_t)^{(i)}\}_{i=1}^N$ approximate $p(x_t, \theta, s_t | y^t)$.

1. *Propagation.* For $i = 1, \dots, N$,

(a) *Propagating* x_{t+1} . Sample $\tilde{x}_{t+1}^{(i)}$ from $p(x_{t+1} | \tilde{x}_t^{(i)}, \tilde{\theta}^{(i)})$.

(b) *Computing weights.* $\omega_{t+1}^{(i)} \propto p(y_{t+1} | \tilde{x}_{t+1}^{(i)}, \tilde{\theta}^{(i)})$.

2. *Posterior draws.* The set $\{(x_{t+1}, \theta, s_t)^{(i)}\}_{i=1}^N$ is sampled from $\{(\tilde{x}_{t+1}, \tilde{\theta}, \tilde{s}_t)^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.

3. *Update sufficient statistics.* $s_{t+1}^{(i)} = \mathcal{S}(\tilde{s}_t^{(i)}, x_{t+1}^{(i)}, \theta_{t+1}^{(i)}, y_{t+1})$, for $i = 1, \dots, N$.

4. *Parameter learning.* Sample $\eta^{(i)} \sim p(\eta | s_{t+1}^{(i)}, \theta_{t+1}^{(i)})$, for $i = 1, \dots, N$.

2.3 Analysis and Results: Simulation Study

The first part of the analysis presented is a simulation study that provides insight into the behavior of the four filters discussed in this chapter. We are able to identify limitations and benefits of using each approach. The particle filters are compared in four ways: (1) degree of particle degeneracy and estimation accuracy; (2) accuracy in estimating regime-switching parameters; (3) size of the Monte Carlo error, and (4) computational cost. Completing the study we discuss the economic insight that can be inferred from the Bayesian estimates and end with a robustness analysis to control for data set-specific effects.

Our simulation analysis is based on 50, 5,000 particle runs. For each of the 50 iterations, we drew a new set of priors that was used to initiate each one of the four filters of interest. In the two filters that use a KS technique we use a shrinkage/smoothness factor of $a = 0.9$. For both the volatility process and the parameters, the median particle is used as the estimate and the 97.5 % and 2.5 % percentile particles are used as the upper and lower bounds of the 95 % confidence band, respectively.

Robustness results are based on ten different data sets and ten runs of the filters, each with a different starting set of prior draws.⁴

In the $k = 1$ case, all filters' prior distribution for τ^2 is inverse gamma, i.e. $\tau^2 \sim IG(v_0/2, v_0 \tau_0^2/2)$, with prior mean $v_0 \tau_0^2 / (v_0 - 2)$. For the filter that use sufficient statistics in the estimation (APF + SS and BF + SS), the prior distributions for α and η are conditionally conjugate, i.e. $\eta | \tau^2 \sim TN_{(-1,1)}(b_0, \tau^2 B_0)$ and $\alpha | \tau^2 \sim N(a_0, \tau^2 A_0)$, where $TN_A(a, b)$ is the normal distribution with mean a and variance b and truncated at A . For the filters with kernel smoothing (LW and BF + KS), the prior distributions are $\eta \sim TN_{(-1,1)}(b_0, B_0)$ and $\alpha \sim N(a_0, A_0)$. The difference between these priors has negligible effect on our empirical study.

⁴ The same prior draws are used in one run of all four filters, thus ensuring that results in this run are comparable across filter.

Hyperparameter values are set up to ensure uninformative priors. In all application scenarios presented in this chapter we used the following setup: $m_0 = 0$, $C_0 = 1$, $a_0 = b_0 = 0$, $A_0 = B_0 = 3$, $v_0 = 4.01$ and $\tau_0^2 = 0.01$. Changes in hyperparameter values were made and no significant change was observed in the results.

In the $k = 2$ case we use the same priors and hyperparameter values for τ^2 and η as the ones just described for the log-stochastic volatility model. Additionally for all filters we have that $p_i \sim \text{Dir}(u_{i0})$ ⁵ for $p_i = (p_{i1}, \dots, p_{ik})$, $i = 1, \dots, k$. For the filter kernel smoothing filters' implementation in this scenario we set $\gamma_1 \sim N(a_0, A_0)$ and $\gamma_i \sim TN_{(0, \infty)}(g_0, G_0)$ $i = 2, \dots, k$. Once more, for the sufficient statistic-based filters priors, we condition on τ^2 for γ_i , $i = 1, \dots, k$. That is, we have $\gamma_1 | \tau^2 \sim N(a_0, \tau^2 A_0)$ and $\gamma_i | \tau^2 \sim TN_{(0, \infty)}(g_0, \tau^2 G_0)$ $i = 2, \dots, k$. All of the hyperparameter values that were already defined for the SV remained unchanged, and the following new values were added: $u_{i0} = (0.5, \dots, 0.5)$ for $i = 1, \dots, k$, $g_0 = 0$ and $G_0 = 3$.

2.3.1 Simulated Data

As mentioned before, we focus our investigation on MSSV models, one of the many possible applications in which to implement the filters presented in Sect. 2.2. As mentioned before, we consider two cases of the number of states, k , in the model: (1) the *log-stochastic volatility* or $k = 1$ and (2) the *two-state MSSV* ($k = 2$).

These two models are simulated for a time frame of 1,000 time periods. For the *log-stochastic volatility* case we use $\alpha = -1$, $\eta = 0.9$, and $\tau^2 = 1$. Time series plots for the return y_t , latent state x_t , and volatility processes are presented in Fig. 2.1. In the $k = 2$ parameter values were chosen to match the values of the first data set used in Carvalho and Lopes (2007). The parameter vector, Θ_2 , is determined by $\alpha_1 = -2.5$, $\alpha_2 = -1$, $\eta = 0.5$, $\tau^2 = 1$, $p_{11} = 0.99$, and $p_{22} = 0.985$. A graphical summary of the processes of interest, y_t , x_t , volatility and the state in which the process is on, s_t , are presented in Fig. 2.2.

2.3.2 Exact Estimation Path

Given the Bayesian nature of the filters analyzed in this chapter, we use an *exact estimation path* as a reference for which the *best* estimation path should be. Likewise, we use the confidence bands obtained in the *exact path estimation* as reference for what sensible confidence bands are for the estimates of interest. The path and bands are obtained by running one of the filters with a large enough number of particles what ensures that both the path and the bands will be replicated by the filter regardless of the prior draws used to initiate the filter. In a non-time constrained world this

⁵ $\text{Dir}(u_{i0})$ means that the prior distribution is Dirichlet with parameter u_{i0} .

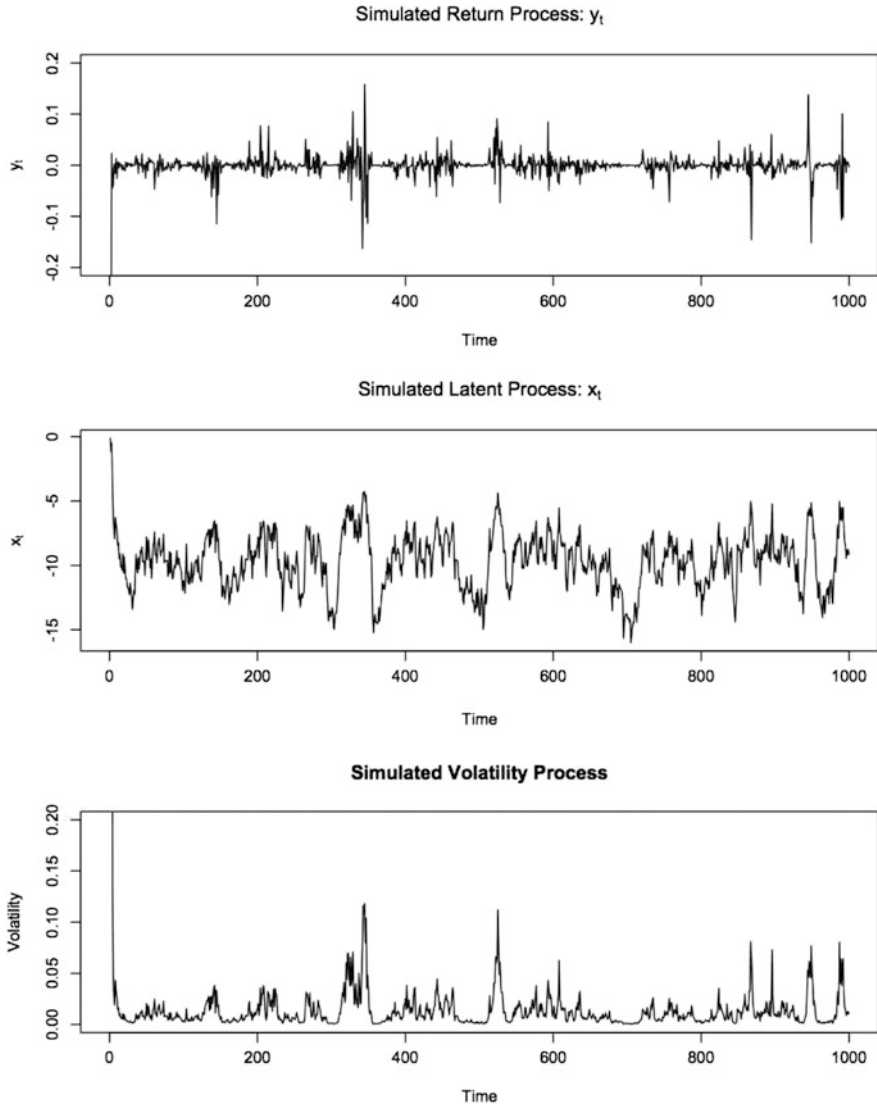


Fig. 2.1 Time series of the simulated return, latent state, and volatility processes for the MSSV model with $k = 1$. Details of the choice of model parameter values can be found in Sect. 2.3.1.

would be the ideal path to use; however, given the current computational capacity, running the filters for sufficiently large number of particles is not time efficient.

Under the premise that these path and bands are perceived as *the true path*, the choice of filter should be irrelevant. Here, the *exact estimation path* is calculated by running a 100,000 particle APF + SS filter. For both the $k = 1, 2$ simulated data

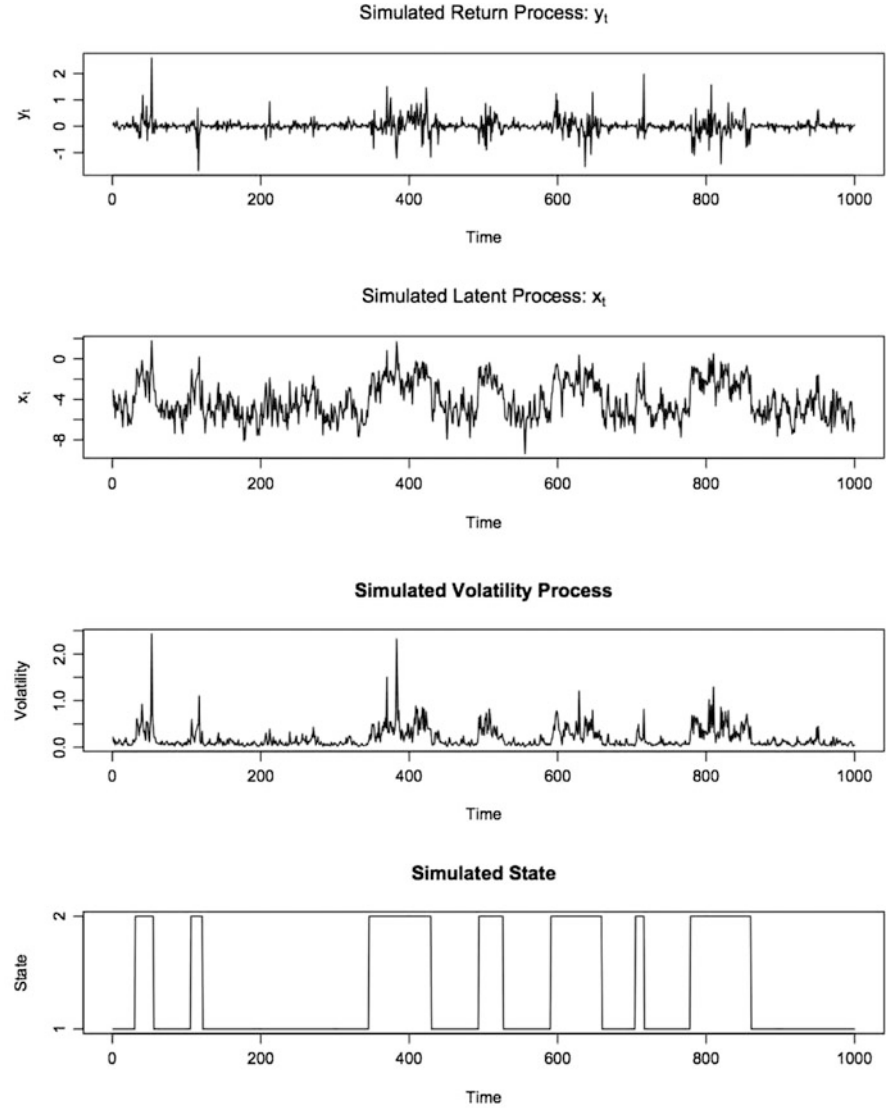


Fig. 2.2 Time series of the simulated return, latent state, and volatility processes for the MSSV model with $k = 2$. Details of the choice of model parameter values can be found in Sect. 2.3.1.

sets, the *exact* parameters and volatility paths and their 95 % confidence bands are estimated. In this chapter setting the *exact* estimates and confidence bands paths obtained will be regarded as the *true paths and bands* and as a result they will be the benchmark used to compare the filters.

2.3.3 Estimate Evaluation

2.3.3.1 Parameter Degeneration and Estimate Accuracy

The behavior of the filter estimates is first analyzed in terms of parameter degeneration and estimate accuracy. Determining how well the filters are able to correctly replicate the volatility processes and estimate the parameter values is paramount to the filters performance.

Correct latent state estimation is the ultimate goal of any filter. In order to evaluate how well the filters presented in this chapter are replicating the volatility process we compare the true simulated process with the filtered estimates. We use a *mean squared error* (MSE) to measure the deviation between the real and estimated processes. The MSE is defined by

$$MSE = \frac{1}{T} \sum_{t=1}^T (V_t - \hat{V}_t)^2 \quad (2.14)$$

where V_t is the real volatility process and \hat{V}_t is the filtered volatility process.

Table 2.1 presents the mean MSE, averaged across the 50 filter repetitions,⁶ for all filters and both volatility models. Divergence from the real volatility process is small and similar in all four filters for the two MSSV cases, showing that the filters are able to accurately replicate the latent state x_t , and thus produce good volatility estimates.

Closer inspection of the behavior of the estimated paths reveals that the discrepancies between the real and the estimated volatilities happen when there are sudden increases in volatility. None of the four filters are able to completely capture these peaks. The problem magnifies in the $k = 2$ case.

Table 2.1 Mean MSE between the real and the filtered volatility processes, averaged across the 50 repetitions of each filter.

k	APF + SS	BF + SS	BF + KS	LW
1	0.000520	0.000527	0.000578	0.000552
2	0.019394	0.019613	0.019155	0.020136

Another element that is worth exploring is the variability that exists within the MSE of the 50 repetitions. From Fig. 2.3 one appreciates that the LW filter results have a significantly wider spread. Moreover, we see that the two strategies that

⁶ The mean MSE presented in Table 2.1 is averaged across repetitions for each one of the filters. That is:

$$\text{Mean } MSE = \frac{1}{50} \sum_{i=1}^{50} MSE_i = \frac{1}{50} \sum_{i=1}^{50} \frac{1}{T} \sum_{t=1}^T (V_t - \hat{V}_t)^2 \quad (2.15)$$

where MSE_i is the MSE for repetition i of a given filter.

involve an APF in the propagation and sampling of the underlying process are less stable than the two filters that implement a BF strategy. As expected, the lower the variability, the more powerful the claim we can make on the accuracy of the filters, as any run will likely have the same deviations from the real process.

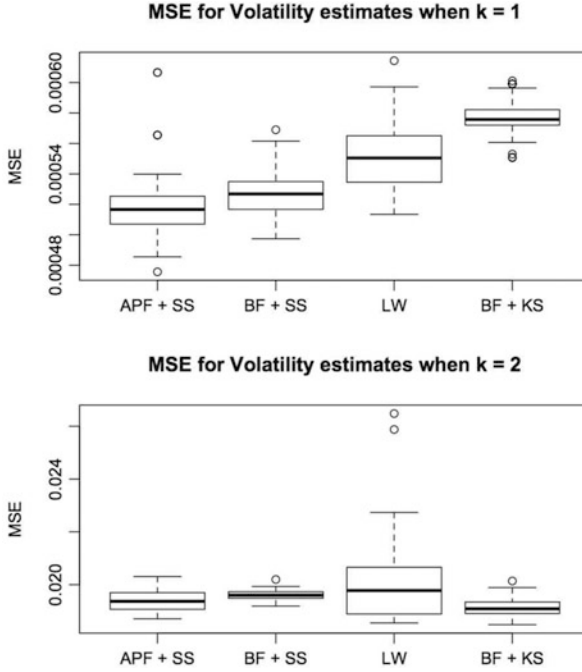


Fig. 2.3 Box plots of the MSE of the estimated volatility process compared to the real simulated volatility process for each filter. The *left plot* presents the results for the MSSV with $k = 1$ and the *right plot* presents the results for the MSSV with $k = 2$.

Switching to parameter accuracy we focus on parameter degeneracy, a phenomenon that appears when the resampling weights concentrate on one or a few mass points and make the parameter estimates and their confidence bands collapse to very narrow ranges and sometimes even to a single point. Our filter comparison uses the *exact path's* 95 % band as a benchmark for reasonable values of estimate's confidence credibility bandwidth. Furthermore, we use the *effective sample size* (ESS) to complement this part of the study. ESS is defined as:

$$ESS = \left(\sum_{t=1}^T w_t^2 \right)^{-1} \quad (2.16)$$

where w_t is the resampling weight at time t (see [Lopes and Carvalho \(2011\)](#) and [Kong, Liu and Wong \(1994\)](#)). This measure is a good proxy for the number of particles where the weights have mass, thus making them the most likely candidates in the resampling step. As we will further explore, there arguably is a relationship between parameter collapsing and ESS value.

The first component of the parameter degeneration analysis is to understand which filters, and in what proportion, present cases of the latter phenomenon. To this end, we look at how many filter runs have parameter 95 % confidence bands' width narrower than two threshold percentages of the *parameter exact 95 % confidence band's width*. In particular, we are interested in confidence credibility bandwidths narrower than 10 % and 20 % of the benchmark 95 % confidence band's width.

Table 2.2 presents a summary of the results for the two volatility models discussed. In the $k = 1$ case, only the LW filter presents collapsing parameters, with at least 20 % of the filter runs presenting this anomaly. In the $k = 2$ case all four filters have at least one parameter for which the estimates degenerate. Yet, it is again the LW filter the one that presents a more delicate situation with the most collapsing cases. At least 20 % of the filter's runs appear to produce defective parameter estimates. The high proportions of cases with parameter collapses issues found in the LW filter raise a flag on the accuracy and applicability of this widespread filter.

Table 2.2 The left side presents the number of runs that reveals parameter collapses in the 50 runs. A collapsing case is a filter repetition in which the width of the 95 % credibility bands is narrower than 10 % or 20 % of the width of the exact parameter path. The right side presents the 25 %, 50 %, and 75 % percentiles for the effective sample size of the non-collapsing filters.

k	Filter	Parameter								Collapse	ESS		
		α	α_1	α_2	η	τ^2	$p_{1,1}$	$p_{2,2}$	25 %		50 %	75 %	
		0.1 0.2	0.1 0.2	0.1 0.2	0.1 0.2	0.1 0.2	0.1 0.2	0.1 0.2					
1	APF + SS	0 0			0 0	0 0			No	3,614.798	4,226.624	4,484.099	
	BF + SS	0 0			0 0	0 0			No	3,352.356	3,915.876	4,322.197	
	BF + KS	0 0			0 0	0 0			No	3,321.399	3,882.107	4,303.528	
	LW	13 19			13 20	16 28			Yes				
2	APF + SS		0 0	0 5	0 0	0 0	0 1	2 44	No	3,616.132	4,151.599	4,422.509	
	BF + SS		0 0	0 0	0 0	0 0	0 0	4 45	No	3,352.903	3,937.380	4,293.923	
	BF + KS		0 0	0 0	0 0	0 0	2 16	25 32	No	3,317.990	3,919.167	4,301.583	
	LW		11 22	12 32	12 25	22 22	11 23	32 48	Yes				

Graphical examples of the particle degeneration phenomenon for one run of the LW filter for both $k = 1, 2$ are presented in Figs. 2.4 and 2.5, respectively. The dotted lines highlight the points where the minimum ESS happens in the analyzed run. In the two showcased examples the minimum ESS obtained are 1.05 and 1.37 for the $k = 1, 2$ cases, respectively. In other words, it appears that for both cases the LW filter reaches a point where it will give weight to only a very small set of particles, thus only resampling from this limited set. The reader can see in the plots how this clearly translates. To the right of the dotted lines, the estimates collapse to almost a single point and the confidence bands become extremely narrow. Furthermore, for both MSSV models the values where the estimates collapse to are different to the true parameter values. Additionally, due to the band's narrowness, the true value does not lie within the estimates of 95 % confidence band, hence leading to erroneous conclusions about the parameter values. This is a critical estimation accuracy shortcoming of the LW filter.

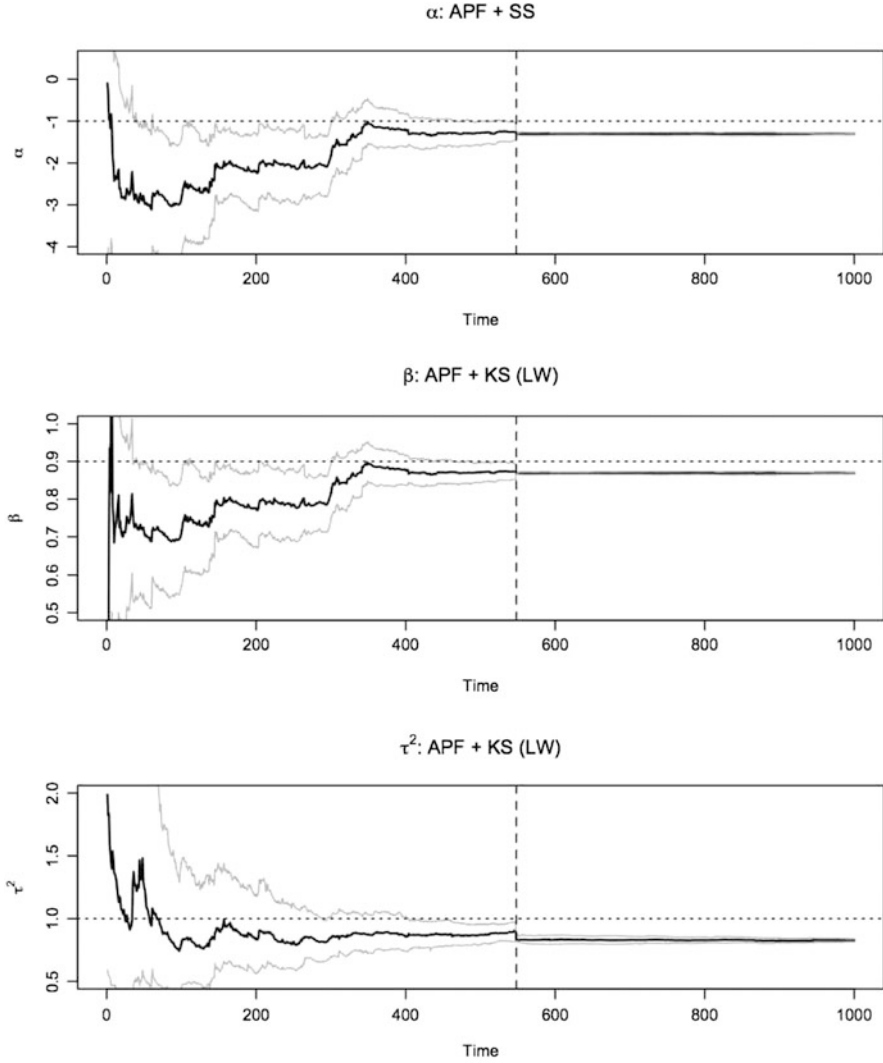


Fig. 2.4 Parameter estimates for LW Filter, in an MSSV $k = 1$ repetition where parameters collapse. *Black lines* represent parameter estimates (median particle for each time period); *gray lines* represent the 97.5 % and 2.5 % quantiles for the particle estimates and the *dotted line* is the true parameter value. The *dashed line* highlights the time period where the min ESS happens in that particular run.

Exploring the ESS behavior for the non-collapsing parameters, Table 2.2 presents⁷ the 25 %, 50 %, and 75 % quantiles for the ESS values obtained in the

⁷ In spite of the fact that the APF + SS, BF + SS, and BF + KS filters show evidence of parameter collapses in the $p_{2,2}$ parameter of the $k = 2$ MSSV model, we still consider them as non-collapsing filters. This is due to the fact that this phenomenon is only present in one of the parameters.

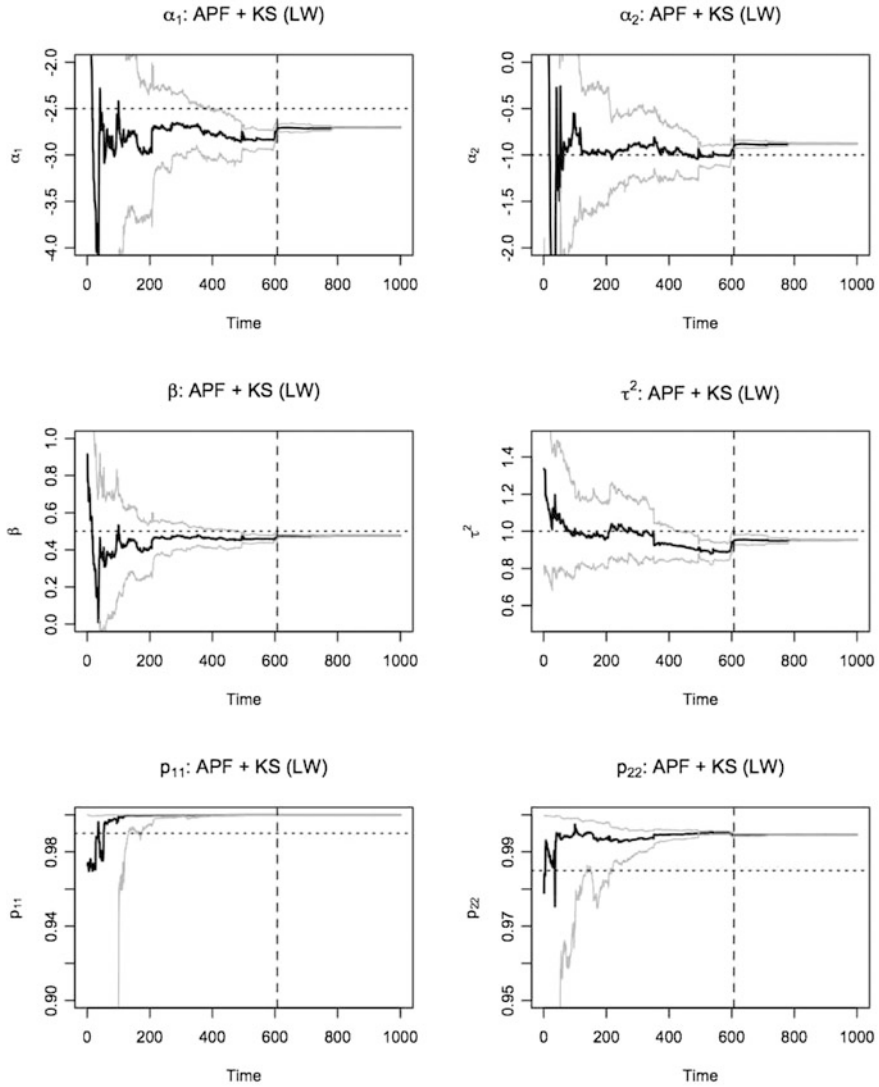


Fig. 2.5 Parameter estimates for LW Filter, in an MSSV $k = 2$ repetition where parameters collapse. *Black lines* represent parameter estimates (median particle for each time period); *gray lines* represent the 97.5 % and 2.5 % quantiles for the particle estimates and the *dotted line* is the true parameter value. The *dashed line* highlights the time period where the min ESS happens in that particular run.

50 replications of each filter. The latter results show that for the most part, the three filters of interest rely on *healthy* amounts of particles to resample from, ensuring variability in the resampling weights which is critical to accurate estimation of the parameters.

2.3.3.2 Regime-Switching Estimation

Particular to the type of volatility models we are estimating in cases where $k \geq 2$, it is important to understand how the filters are able to capture regime changes and track the states in which the model is at. For this we focus on analyzing the two-state MSSV.

Following Bruno and Otranto (2008), Otranto (2001), and Bodart et al. (2005) we use the *quadratic probability score (QPS)*, developed by Diebold and Rudebusch (1989), to evaluate the filters' abilities to correctly determine the state in which the economy is at. The QPS is defined by

$$QPS = \frac{100}{T} \sum_{t=1}^T [Pr(S_t = 2) - D_t]^2 \quad (2.17)$$

where d_t is an indicator variable equal to 1 when the true process is in state 2 and $Pr(S_t = 2)$ is the estimated probability that the process is in the second state. The index varies between 0 and 100. It is equal to 0 in the case of correct assignment of the state variable for all time periods and equals 100 in the opposite case.

Table 2.3 Mean QPS for the $k = 2$ model.

APF + SS	BF + SS	BF + KS	LW
8.96	8.10	7.94	17.95

Table 2.3 presents a summary of the results. For each one of the four filters, the QPS averaged across the 50 replications is presented.⁸ The APF + SS, BF + SS, and BF + KS filters all appear to have similar abilities of correctly estimating the correct state. The LW filter, however, produces considerably less accurate estimates of the regime where the economy is. This could be linked to the particle degeneracy found in the LW filter, as inaccurate parameter estimates lead to erroneous state estimates.

The SS-based methods are specially good at tracking regime changes. For certain scenarios and applications this is an extremely important feature. Thus this is another aspect in which we can claim that the LW filter has shortcomings while the APF + SS and BF + SS filters are outperforming.

⁸ The mean QPS presented in Table 2.3 is averaged across repetitions for each one of the filters. That is:

$$\text{Mean } QPS = \frac{1}{50} \sum_{i=1}^{50} QPS_i = \frac{1}{50} \sum_{i=1}^{50} \frac{100}{T} \sum_{t=1}^T [Pr(S_t = 2) - D_t]^2 \quad (2.18)$$

where QPS_i is the QPS for repetition i of a given filter.

2.3.3.3 Monte Carlo Error

Next, the four filters are evaluated in terms of stability of the produced estimates. In other words, how much Monte Carlo variability is found in the estimates. Under ideal conditions we would like to have parameter estimates that perfectly replicate the estimation paths regardless of the set of prior draws used to initialize the filters. However, this is not a realistic scenario, and all four filters of interest have some Monte Carlo variability, or as we like to call it Monte Carlo error.

In order to analyze the latter variability we once more use the *exact estimate paths* described in Sect. 2.3.2 as benchmark. Assuming that the *exact paths* are what the *true* estimate paths should look like, we analyze the deviations between the different estimate paths that the filters produce and the so-called *exact path*.

A preliminary graphical exploration of the estimates allows to get a first impression of the way that the estimates behave. Figures 2.6–2.8 present the estimate paths for the three parameters in the $k = 1$ MSSV model, while Figs. 2.9–2.14 show the paths for the estimates of the six parameters in the $k = 2$ MSSV model. In these panels, the solid black lines present the *exact path* and the gray lines are the estimate paths for each one of the 50 runs of each filter.

The plots reveal that the two filters that have the least Monte Carlo variability are the two using an SS approach in the fixed parameter estimation. On the other hand, the filter that consistently appears to have the largest variability is the so-called LW filter, which in the $k = 2$ model has considerably greater variation for η , $p_{1,1}$, and $p_{2,2}$.

A more rigorous way to analyze the Monte Carlo error is to *measure* the deviation between the *exact path* and the estimate path. To avoid confusion with the MSE previously used, we here use the mean absolute error between the two paths, which we call the Monte Carlo mean absolute error (MCMAE), defined by

$$MCMAE = \frac{1}{T} \sum_{t=1}^T |p_t - \hat{p}_t| \quad (2.19)$$

where p_t is the *exact parameter path* and \hat{p}_t is the estimated path at time t . Given that we are looking at 50 runs of each one of the filters we will focus on analyzing the mean across runs of the MCMAE. A summary of the mean MCMAE results for all the parameters in the two MSSV models is shown in Table 2.4.⁹

The results in the latter table corroborate the graphical findings. The APF + SS and BF + SS filters have the smaller deviations. For most parameters the APF + SS has slightly lower Monte Carlo error than the BF + SS. Analyzing the behavior

⁹ The mean MCMAE presented in Table 2.4 is averaged across repetitions for each one of the filters. That is:

$$\text{Mean } MCMAE = \frac{1}{50} \sum_{i=1}^{50} MCMAE_i = \frac{1}{50} = \frac{1}{T} \sum_{t=1}^T |p_t - \hat{p}_t| \quad (2.20)$$

where $MCMAE_i$ is the MCMAE for repetition i of a given filter.

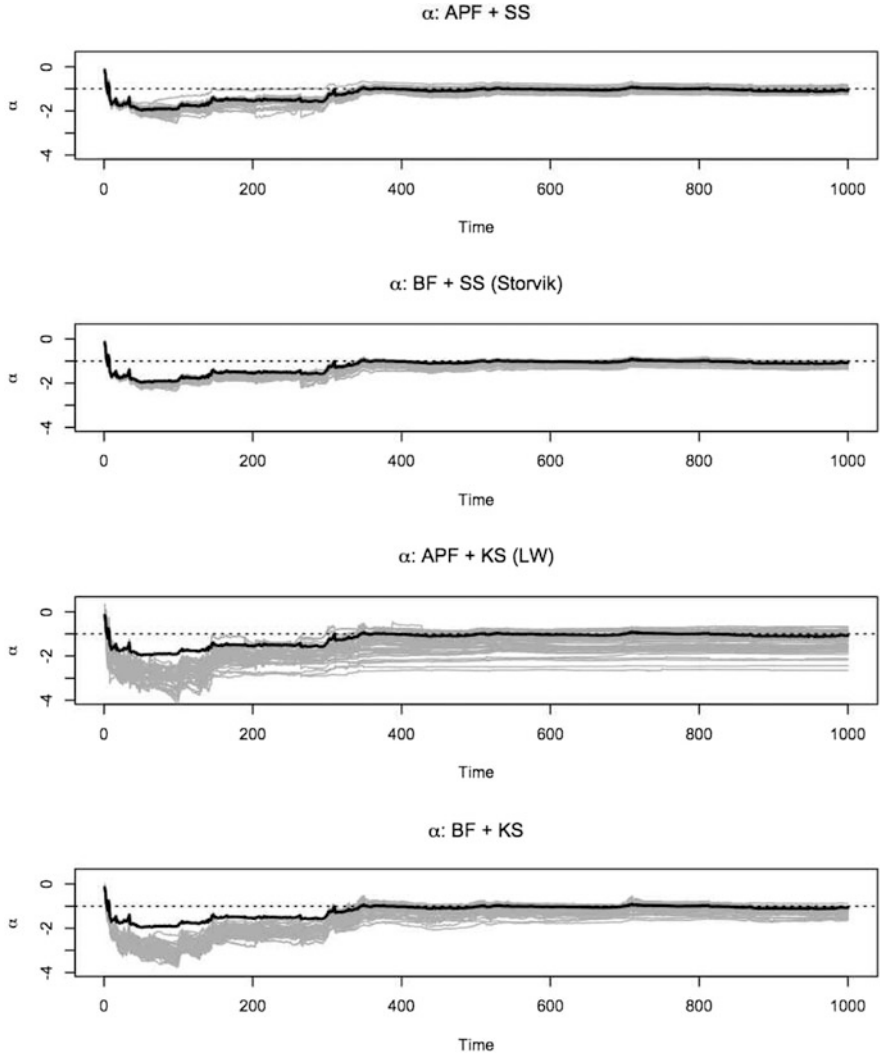


Fig. 2.6 Estimate paths for α in the MSSV $k = 1$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

of the kernel smoothing related filters, one appreciates that the LW filter is consistently more variable than the BF + KS filter. The largest deviations are significantly greater; in some cases the LW filter MCMAE is twice as much as the APF + SS' MCMAE. Once more, we see that on this dimension the LW filter appears to underperform the other filters.

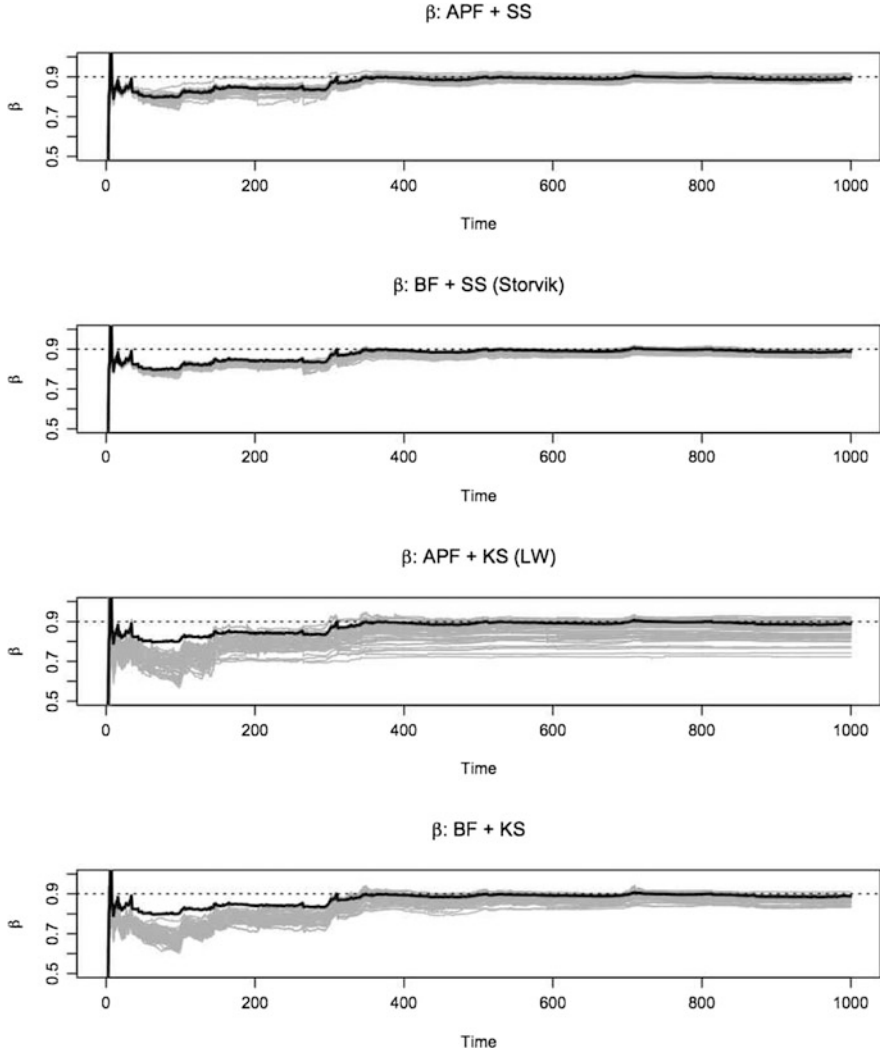


Fig. 2.7 Estimate paths for η in the MSSV $k = 1$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

2.3.3.4 Computational Time

The last dimension on which we compare the filters in the simulation study is the amount of time taken to complete one run. Table 2.5 presents the estimation times in seconds, averaged across runs¹⁰ for the four filters and the two models of interest.

¹⁰ The MSSV with $k = 1$ model runs and the MSSV with $k = 2$ model runs were implemented in different machines. In this chapter, the MSSV $k = 2$ was run in a more powerful computer.

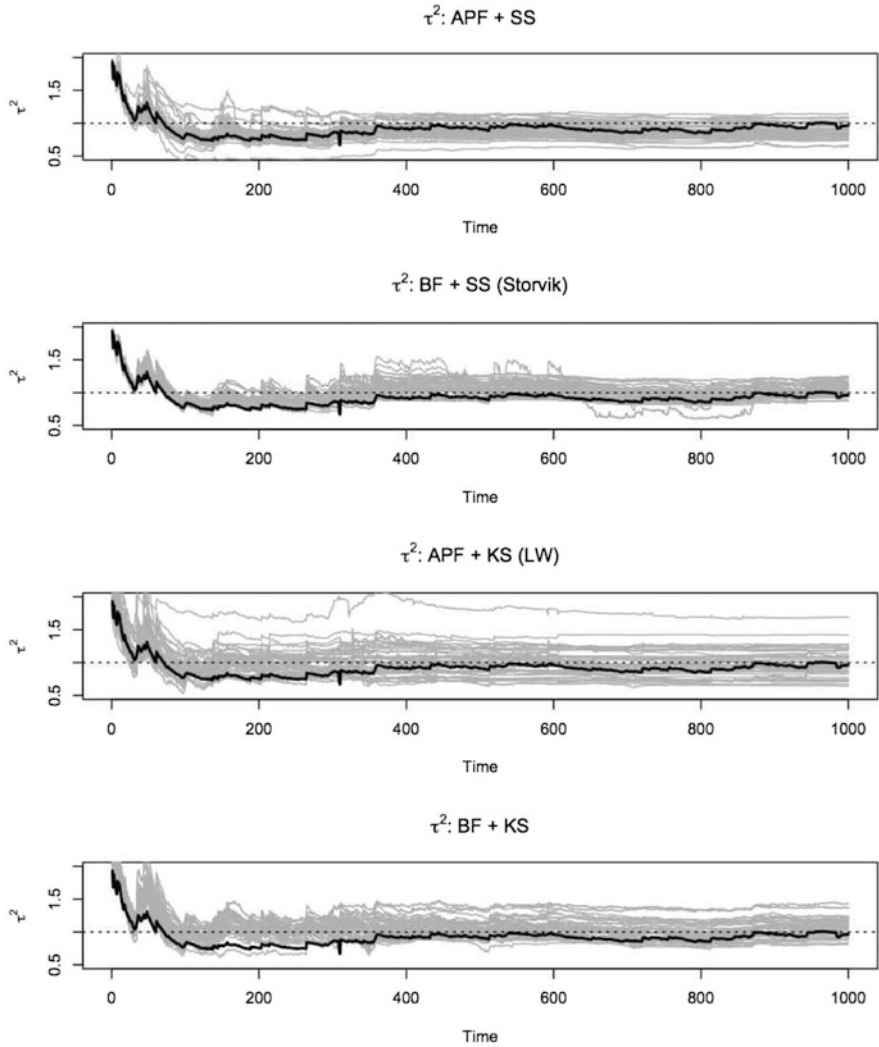


Fig. 2.8 Estimate paths for τ^2 in the MSSV $k = 1$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

The filters that implement an SS approach to parameter estimation take significantly more time. The reason for this is the complexity of the operations needed to update the sufficient statistics. Furthermore, we run all of our simulations in R which is known to struggle with loop calculations, which are, unfortunately, unavoidable in the SS updating. Operations that parameter estimation requires in the kernel smoothing technique are considerably simpler, making the LW and BF + KS

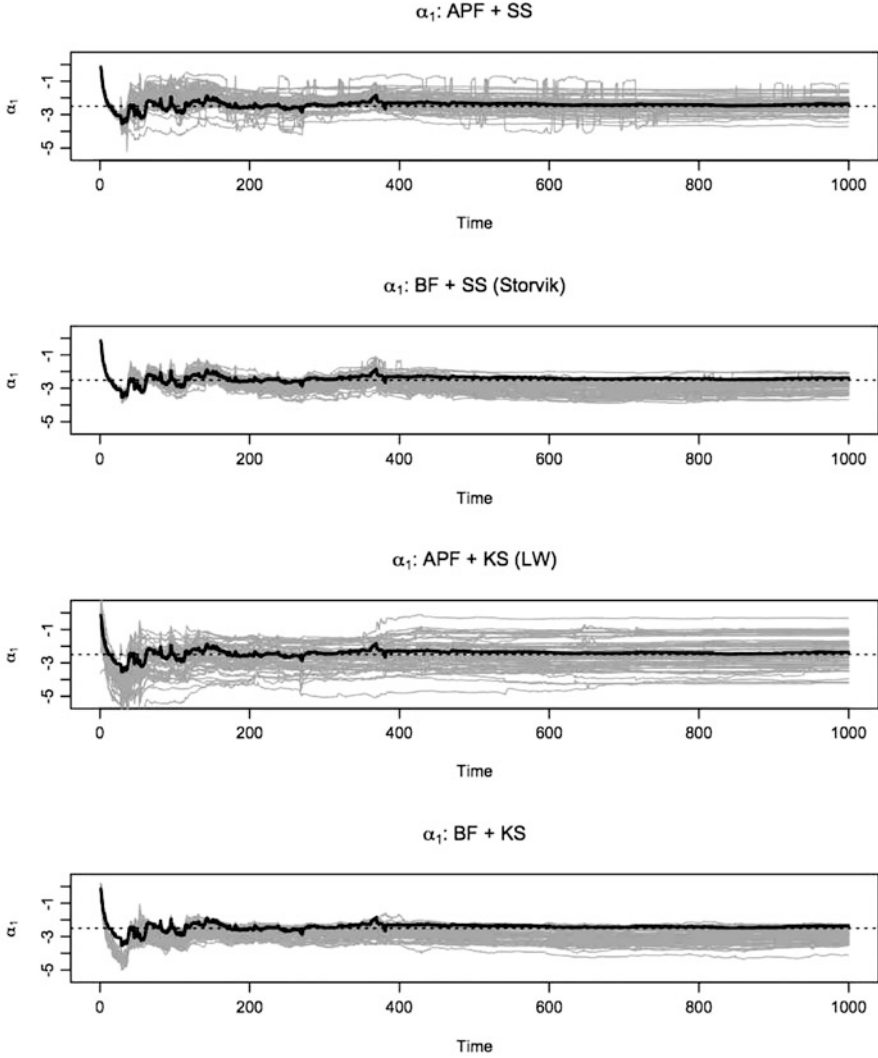


Fig. 2.9 Estimate paths for α_1 in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

filters much more efficient in computation time terms. Unlike the other dimensions that we have explored so far, the LW filter is one of the filters that outperforms.

There appears to be an interesting trade-off between accuracy and computation time. The more accurate filters appear take longer to estimate. Therefore the question is how much accuracy you are willing to give up for a faster estimation. Another perspective from which this issue can be analyzed is how many particles I will implement. Accuracy and time are closely related to the amount of particles used.

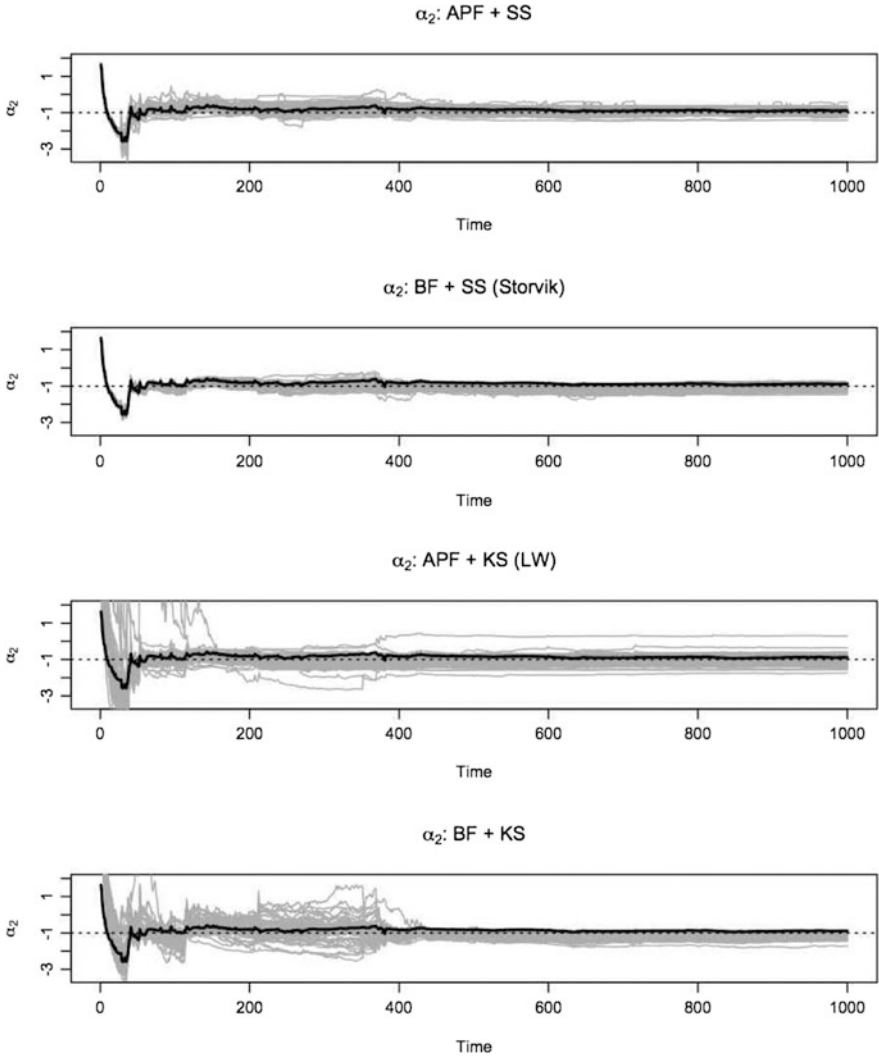


Fig. 2.10 Estimate paths for α_2 in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

A good compromise to getting faster more accurate estimation could be to increase the number of particles in the LW filter estimation. Another option is to use less particles in an APF + SS filter. Preliminary analysis shows that this filter produces accurate estimates with a smaller number of particles. Details on how the increases in particles would affect accuracy and estimation time is beyond the scope of this chapter.

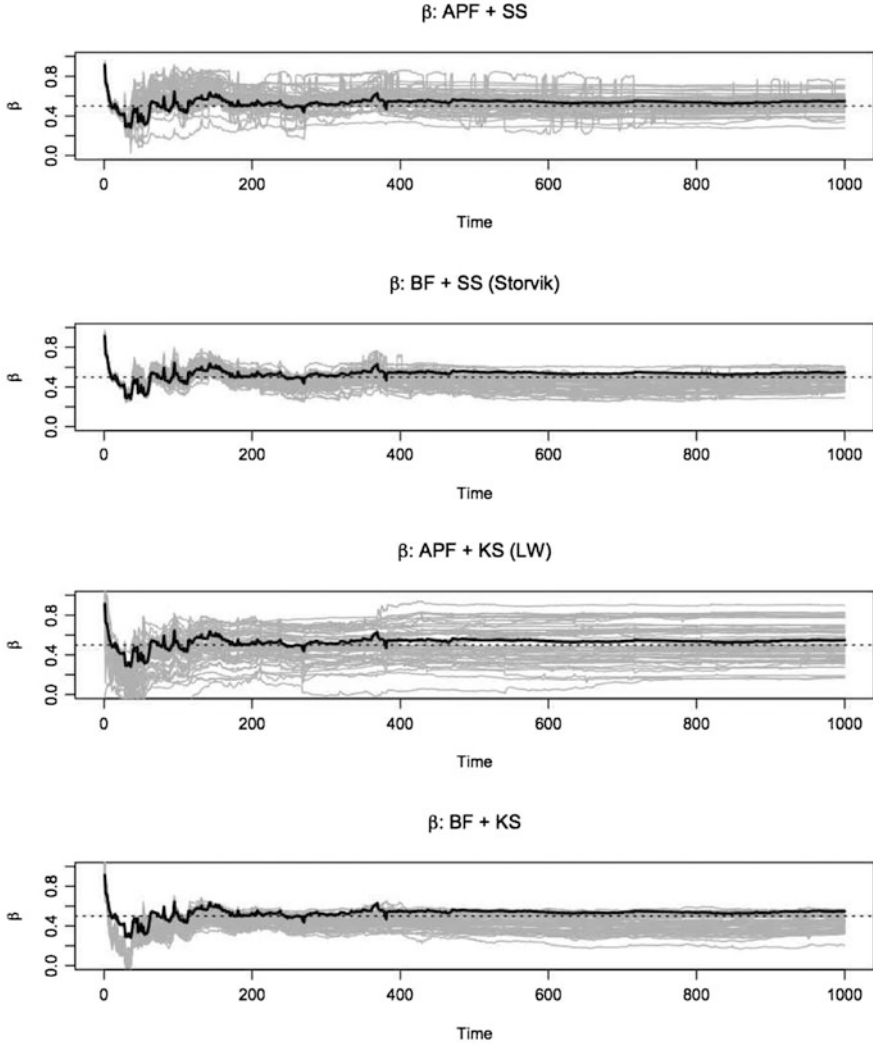


Fig. 2.11 Estimate paths for η in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

2.3.4 Economic Insight

Exploring in more detail the estimates we observe that they bring interesting economic insight. At every point in time, the Bayesian nature of the results allows to infer information about the distribution of the estimates.

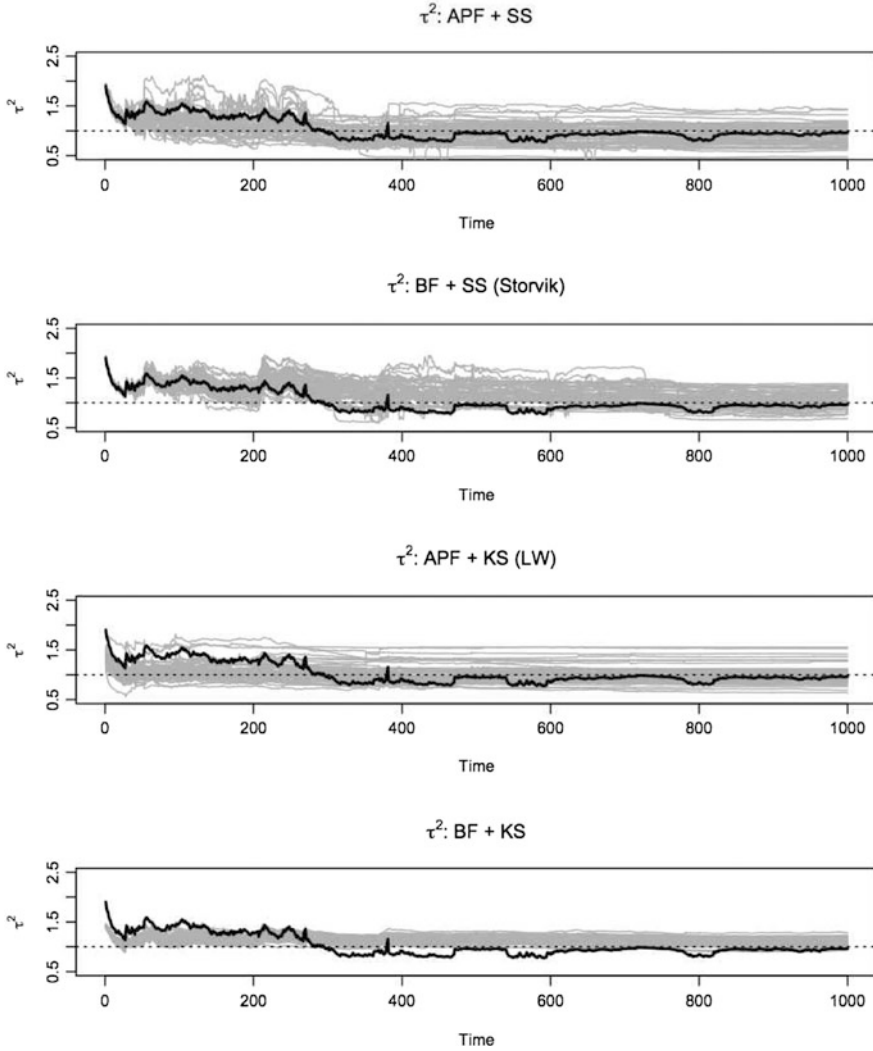


Fig. 2.12 Estimate paths for τ^2 in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

Using the exact path estimates we can provide the economic interpretation based on the posterior distribution of the parameter. In particular, τ^2 provides interesting information about the volatility process in the two-state MSSV. Figure 2.15 shows that τ^2 moves along the volatility, i.e. when there are economy shifts between regimes. From the two panels in Fig. 2.15 one appreciates that when the volatility process shifts to a higher level state, the τ^2 estimates arguably also switch to a higher volatility of volatility level.

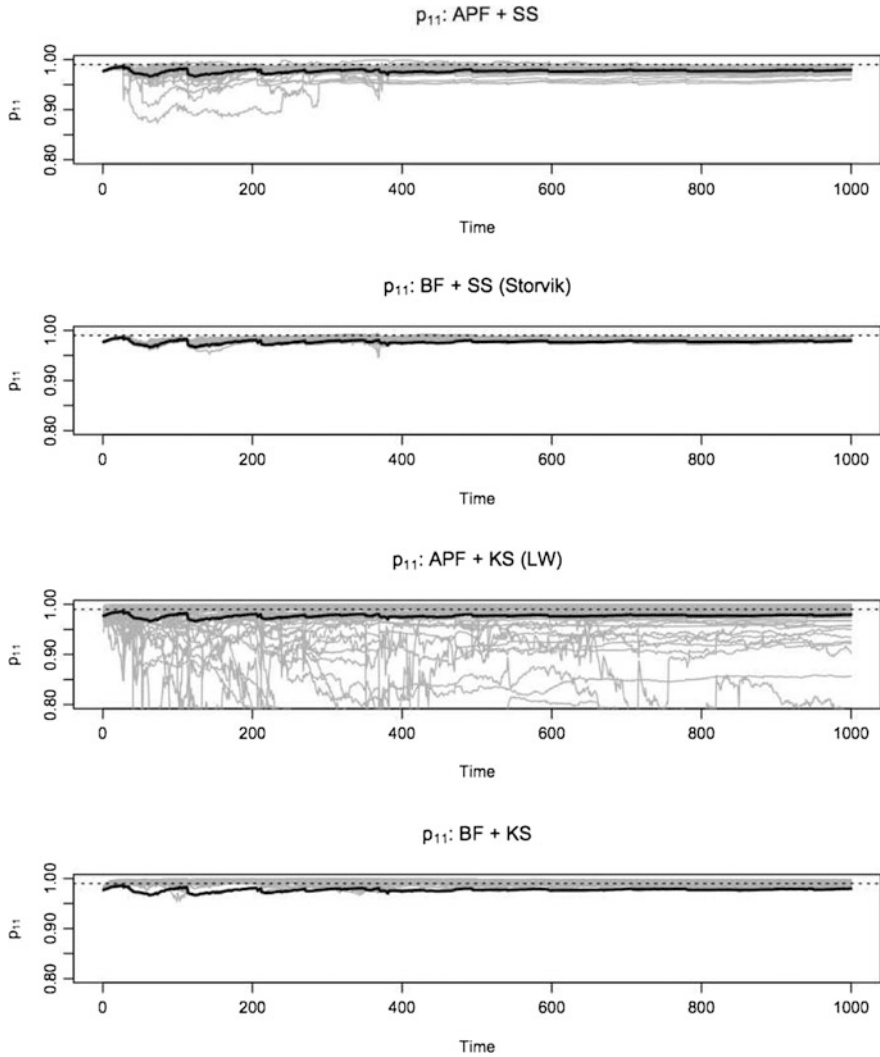


Fig. 2.13 Estimate paths for $p_{1,1}$ in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

2.3.5 Robustness

Using the same parameter values as the ones discussed in Sect. 2.3.1, ten new data sets were simulated for the $k = 1$ and $k = 2$ cases. Each new data sets was estimated for ten runs. Results were then analyzed on the four dimensions presented above.

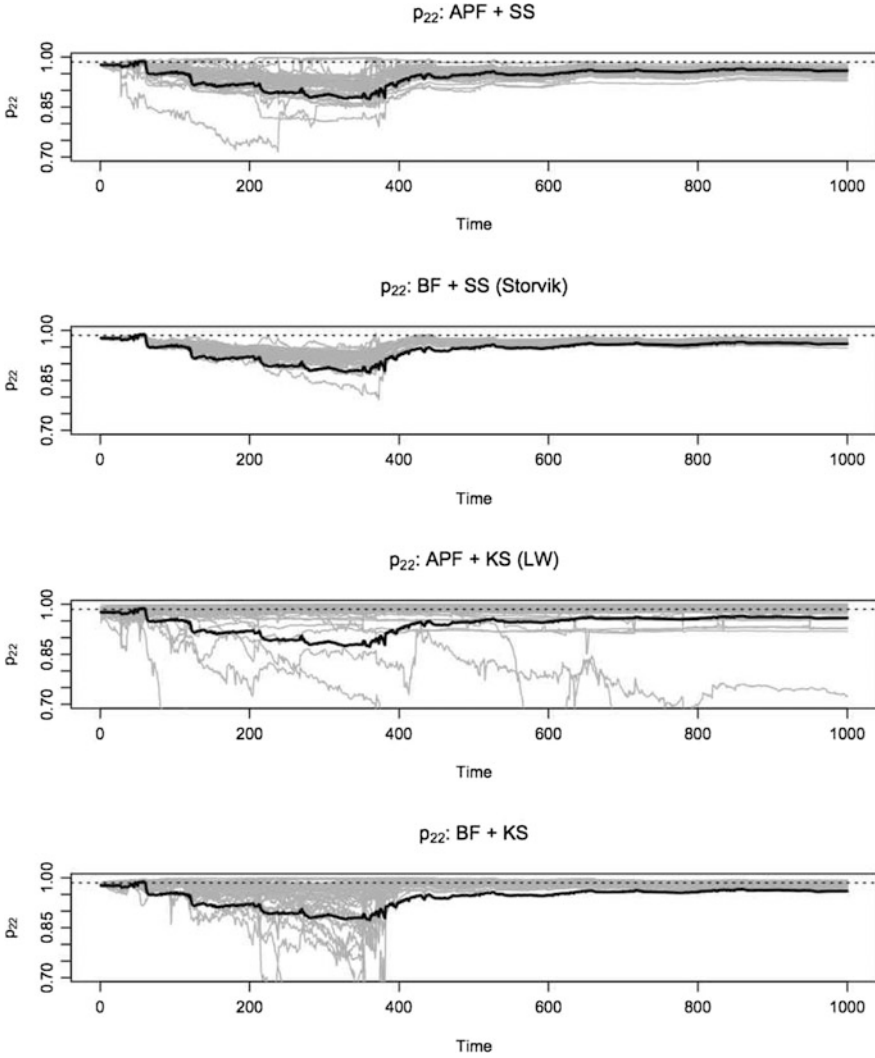


Fig. 2.14 Estimate paths for $p_{2,2}$ in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

A detailed exploration of the ten data sets and ten runs for each one of the models revealed findings consistent with the ones previously discussed. All the results presented in Sects. 2.3.3.1–2.3.3.4 are robust to the data set chosen.

We see that in the additional tested cases, the APF + SS filter is the filter that appears to outperform the other filters. Likewise, we observed that the LW filter continues to have the same shortcomings. It has collapsing parameters, has the largest

Table 2.4 Mean MCMAE between the *exact path* and the estimated path for the parameters of interest in the MSSV models where $k = 1, 2$. The MCMAE are averaged across the 50 repetitions of each filter.

k	Parameter	APF + SS	BF + SS	BF + KS	LW
1	α	0.102	0.135	0.381	0.508
	η	0.0109	0.0136	0.0396	0.0544
	τ^2	0.0959	0.123	0.845	0.921
2	α_1	0.391	0.472	0.567	0.731
	α_2	0.172	0.253	0.392	0.568
	η	0.0806	0.0928	0.1170	0.1550
	τ^2	0.185	0.225	0.207	0.199
	p_{11}	0.00627	0.00572	0.207	0.08460
	p_{22}	0.0185	0.0177	0.0364	0.0603

Table 2.5 Mean time in seconds taken to estimate the MSSV models with $K = 1$ and 2 by the four different filtering strategies. Computational times are averaged across the 50 filter repetitions.

k	APF + SS	BF + SS	BF + KS	LW
1	5429.02	5447.22	31.58	39.15
2	4432.43	4244.95	159.62	220.55

Monte Carlo error, and has the biggest discrepancies when capturing the regime switches. The latter evidence reassures our previous findings on the applicability and accuracy of the four filters of interest.

2.4 Analysis and Results: Real Data Applications

In the second part of the analysis we use two equity indices and analyze their volatility processes using the outperforming filter, that is the APF + SS filter. The first is the IBOVESPA index¹¹ which is presented in order to replicate the results presented in [Carvalho and Lopes \(2007\)](#). The second series that we use is the S&P 500 index, where we explore a short and a long series allowing us to explore and highlight more properties of the Bayesian filtering estimation techniques, and the APF + SS filter in particular.

All data analyzed here was obtained from *Bloomberg* using the last price as a proxy for the day's trading price and only including data from days where trading took place. Table 2.6 presents summary statistics of the three analyzed series.

¹¹ IBOVESPA is an index of about 50 stocks that are traded on the Sao Paulo Stock, Mercantile and Futures Exchange (BOVESPA).

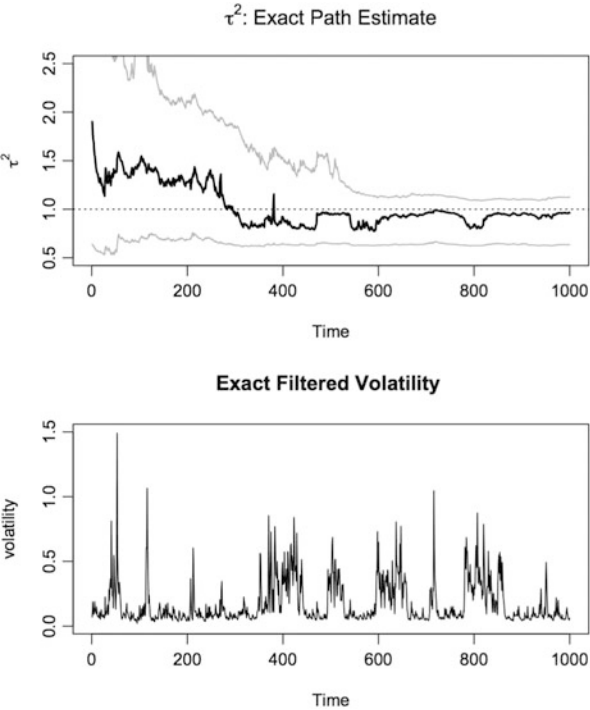


Fig. 2.15 Top panel shows the exact parameter estimate for τ^2 . Bottom panel shows the exact volatility estimate for the MSSV $k = 2$ model.

Table 2.6 Summary statistics for the three real data sets used in the applications in Sect. 2.4.

Series	Start date	End date	Obs	Mean	SD	Min	Max	Kurtosis	Skewness
IBOVESPA	01/02/1997	01/16/2001	1000	0.000878	0.0294	−0.172	0.288	15.920	0.597
S&P 500 (short)	9/1/2006	8/31/2011	1258	−5.791e−05	0.0163	−0.0947	0.110	10.665	−0.280
S&P 500 (long)	9/1/1971	8/31/2011	10094	0.00024	0.011	−0.229	0.110	29.272	−1.075

2.4.1 IBOVESPA

We implement the proposed filter on the IBOVESPA stock index from 01/02/1997 to 01/16/2001. Figure 2.16 shows the log returns of the index from 01/02/1997 to 01/16/2001 (1,000 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.

Starting with priors similar to the ones used in Carvalho & Lopes (2007) we estimate the parameters both in the log-linear stochastic volatility and in the two-state MSSV. Model selection was done using Bayes factors that revealed strong

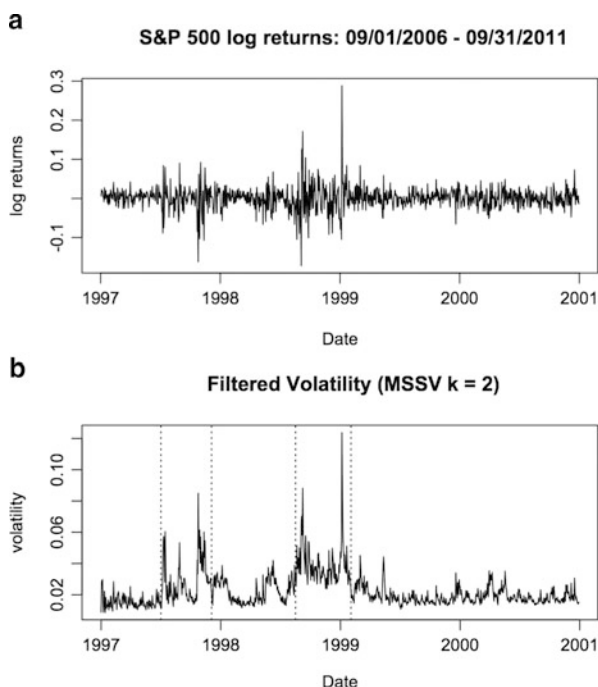


Fig. 2.16 Panel A: log returns for the IBOVESPA stock index between 01/02/1997 and 01/16/2001 for a total of 1,000 observations. Panel B: volatility estimates for the IBOVESPA fitting an MSSV $k = 2$ model. Dotted lines highlight the following dates: (1) 07/02/1997: Thailand devalues the Baht by as much as 20 %; (2) 12/02/1997: IMF and South Korea set a bailout agreement; (3) 08/19/1998: Russia officially falls into default; 02/02/1999: Arminio Fraga is named President of Brazil's Central Bank.

evidence in favor of the two-state MSSV (log Bayes factor of 10.579). The estimated volatility times series is presented in panel B of Fig. 2.16 where one appreciates, like in [Carvalho and Lopes \(2007\)](#), the structural changes that result in periods of higher volatility. The dotted lines in plot highlight four key dates¹² mentioned in [Carvalho and Lopes \(2007\)](#), which appear to coincide with the regime switches in the model and that agents perceive as moments that started or ended a crisis. This

¹² The following key dates are highlighted in Fig. 2.16 panel B that are presented in [Carvalho and Lopes \(2007\)](#)

- (a) 07/02/1997: Thailand devalues the Baht by as much as 20 %
- (b) 12/02/1997: IMF and South Korea set a bailout agreement
- (c) 08/19/1998: Russia officially falls into default
- (d) 02/02/1999: Arminio Fraga is named President of Brazil's Central Bank

behavior matches the one found in [Carvalho and Lopes \(2007\)](#) and corroborates the fact that the sequential estimation is able to identify the structural changes through the discrete state prediction.

Our fixed parameter estimates present behavior similar to the findings by [Carvalho and Lopes \(2007\)](#). Again, the persistence parameter η is not overestimated. The discrete shifts in volatility level the posterior mean for η is no longer close to once. Likewise, the diagonal elements of the transitions probability matrix for the discrete states are high. In our estimation we obtained¹³ $E(p_{11}|D_t) = 0.994$ and $E(p_{22}|D_t) = 0.974$, matching the observation of [Carvalho and Lopes \(2007\)](#) that the duration in each regime is quite long with a predominance of the low volatility regime.

2.4.2 S&P 500

As mentioned before, for the S&P 500 exploration we look at two series: (1) a short, 5-year series between 9/1/2006 and 8/31/2011 and (2) a long 40-year series between 9/1/1971 and 8/31/2011. The shorter series is used as second example of the applicability of the filters and for this one we do a backtest to highlight the accuracy of the estimates in a real data setting. The longer series is used to highlight the benefits of using the APF + SS in more extensive data sets.

Both the log linear stochastic volatility model and the two-state MSSV model are implemented for the two time series. The filters are run using uninformative priors with the same hyperparameter values as the ones implemented in the simulation study. Again a Bayes factor will be used to determine which model better fits the data.

2.4.2.1 Short Time Series

The shorter S&P 500 series that we explore in this chapter is summarized graphically in [Fig. 2.17](#) top panel. Here, the reader can appreciate that at the end of 2008 and in the middle of 2011 there are periods with particularly high variability in the returns. This is consistent with the economic climate, as they coincide with the timing of the Lehman Brothers bankruptcy and the European credit crisis, respectively; 5,000 particle APF + SS filters are run for the two stochastic volatility models. Latent state and parameter estimates are obtained fitting $k = 1, 2$ MSSV models. Evaluating the results, we obtained a *log Bayes* factor of 15.52 in favor of the two-state MSSV model, implying that there are periods of time where the substantial and

¹³ $E(.|D_t)$ is the conditional expected value given the data up to time t .

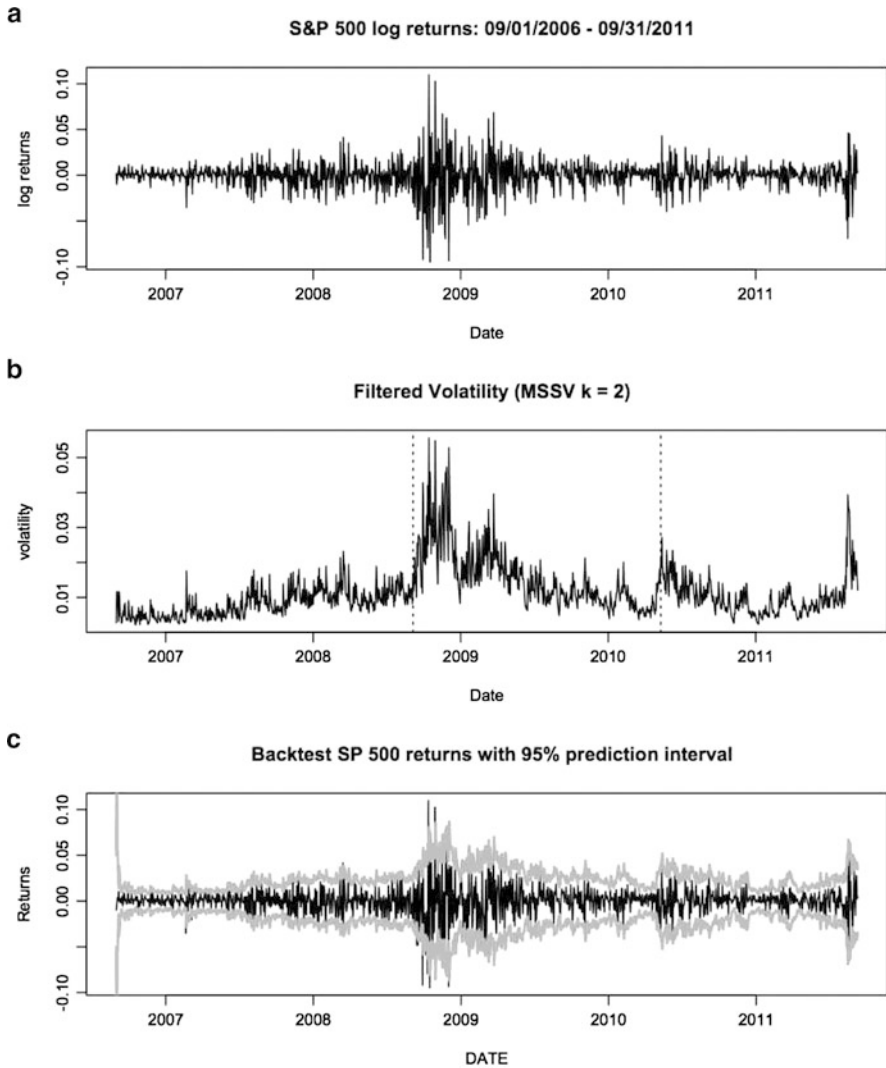


Fig. 2.17 Panel A shows the log returns for the S&P 500 stock index between 09/01/2006 and 08/31/2011. Panel B shows the volatility estimate obtained fitting an MSSV with $k = 2$. Panel C presents a graphical summary of the MSSV $k = 2$ estimates backtest. The *black line* is the true return process and the *dark gray lines* are the 95 % predictive intervals.

sustained volatility increases. Panel B of Fig. 2.17 shows the $k = 2$ MSSV volatility estimates. From the volatility estimates, it is somewhat clear when regime-shifting takes place, which in turn match periods of increased volatility. Dotted lines in the volatility estimate plots highlight an important date around the Lehman Brothers

Bankruptcy filing (September 7, 2008¹⁴) and a key date on the European credit crises (May 10, 2010¹⁵).

The sequential strategy used here is able to capture the structural changes in the volatility, by accurately identifying the moments of higher volatility through the discrete state prediction. The diagonal elements of the transition probability matrix for the discrete states are estimated to be high, with $E(p_{1,1}|D_t) = 0.981$ and $E(p_{2,2}|D_t) = 0.994$. This implies that the duration in each regime is quite long which is a fact also encountered by So et al. (1998) when analyzing the US S&P 500 series.

Further assessment of the parameters reveals, as expected, that there is no evidence of parameter degeneracy. Point estimates and 95 % credible intervals for the components of the Θ_{MSSV} are presented in Table 2.7. To test the accuracy of the parameter estimates we run a backtest of the data. Panel C of Fig. 2.17 presents the 95 % predictive intervals compared to the real returns. To test the accuracy of the predictions, we calculate the percentage of observations that lie outside the predictive interval. In this case we have 9.308 % of the observations outside the interval. This means that the estimates are underestimating the volatility. This is consistent with the observations of the simulation study that revealed that all filtering strategies had limitations with correctly estimating the latent process when there were sudden large increases.

Table 2.7 APF + SS fixed parameter estimates and 95 % credible intervals for the $k = 2$ MSSV models fitted to the short S&P 500 data set. $E(\cdot|D_t)$ is the conditional expected value given the data up to time t .

Parameter	95 % credible interval	$E(\cdot D_t)$
α_1	(-2.733, -1.892)	-2.313
α_2	(-2.289, -1.609)	-1.948
η	(0.734, 0.814)	0.775
τ^2	(0.493, 0.656)	0.592
p_{11}	(0.961, 0.994)	0.982
p_{22}	(0.982, 0.998)	0.994

¹⁴ The Federal Housing Finance Agency (FHFA) places Fannie Mae and Freddie Mac in government conservatorship. The U.S. Treasury Department announces three additional measures to complement the FHFAs decision: (1) preferred stock purchase agreements between the Treasury/FHFA and Fannie Mae and Freddie Mac to ensure the GSEs positive net worth; (2) a new secured lending facility which will be available to Fannie Mae, Freddie Mac, and the Federal Home Loan Banks; and (3) a temporary program to purchase GSE MBS.

¹⁵ The Federal Housing Finance Agency (FHFA) places Fannie Mae and Freddie Mac in government conservatorship. The U.S. Treasury Department announces three additional measures to complement the FHFAs decision: (1) Preferred stock purchase agreements between the Treasury/FHFA and Fannie Mae and Freddie Mac to ensure the GSEs positive net worth; (2) a new secured lending facility which will be available to Fannie Mae, Freddie Mac, and the Federal Home Loan Banks; and (3) a temporary program to purchase GSE MBS.

2.4.2.2 Long Time Series

To further assess the applicability of the APF + SS filter we implemented the filter on a significantly longer data set. We use a 40-year or 10,094 observation data set to explore the performance of the filter of interest on longer time series, as the likelihood of parameter collapses increases with the length of the data. We implemented 5,000 and 10,000 particle APF + SS filters and evaluated the results. In both cases we observed that the filter performed well with no collapses or parameter degeneracy present in either setting which corroborates the wide applicability scope of the APF + SS filter.

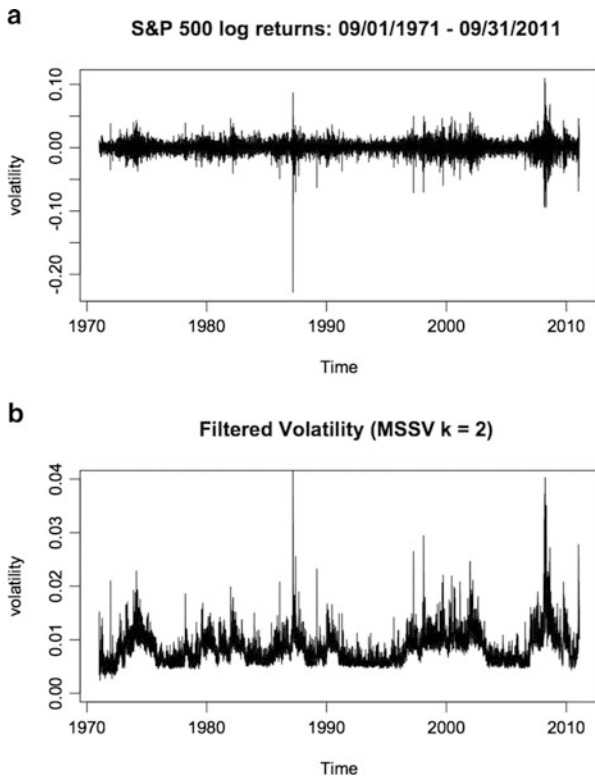


Fig. 2.18 Panel A shows the log returns for the S&P 500 stock index between 09/01/1971 and 08/31/2011. Panel B shows the volatility estimate obtained fitting an MSSV with $k = 2$.

Here, we present a summary of the 10,000 particle estimation, given that results are very similar regardless of the amount of particles used. Figure 2.18 panel A presents the log returns of the S&P 500 in the analyzed period. After fitting the $k = 1, 2$ MSSV models, a Bayes factor analysis revealed strong evidence supporting the two-state MSSV. Panel B of Fig. 2.18 shows the volatility estimates for this model. Again, it is interesting to observe how in the return and volatility plots the

periods of high variability match. Once more, this proves that the sequential learning characteristic of the filter allows to detect the structural changes in the data through the discrete state prediction.

A summary of the parameter estimates and 95 % credible intervals is presented in Table 2.8. Like in the short series, the diagonal elements in the transition probabilities are quite high, again being consistent with the [So *et al.* \(1998\)](#) findings. Furthermore, by allowing discrete shifts in the volatility level the posterior mean for η is no longer close to one which means that the persistence parameter is not being overestimated as discussed in [Carvalho and Lopes \(2007\)](#).

Table 2.8 APF + SS fixed parameter estimates and 95 % credible intervals for the $k = 2$ MSSV models fitted to the long S&P 500 data set. $E(.|D_t)$ is the conditional expected value given the data up to time t .

Parameter	95 % credible interval	$E(. D_t)$
α_1	(-4.488, -4.151)	-4.321
α_2	(-4.013, -3.710)	-3.863
η	(0.559, 0.592)	0.576
τ^2	(0.410, 0.435)	0.422
p_{11}	(0.997, 0.999)	0.998
p_{22}	(0.997, 0.999)	0.998

2.5 Conclusions

Our main contribution is to extend the APF and BF filters to accommodate sequential parameter learning via conditional sufficient statistics. We showed that, among a group of four filters, the APF + SS filter outperforms while the LW filter underperforms for the standard MSSV models. Our APF + SS filter also avoids or at least reduces the dependence of the PF on the somewhat arbitrariness of selecting a shrinkage/smoothness parameter. This is particularly important when dealing with variance parameters, such as the volatility of the volatility parameter in the SV and MSSV models.

The simulation study highlighted some of the important shortcomings in the LW filter. As expected, there is strong evidence of collapses in the parameter estimates and high Monte Carlo error. The only dimension in which this filter appeared to be efficient was in computational time. On the other hand, the APF + SS filter produces accurate parameter estimates that have the lowest Monte Carlo variability and show no evidence of particle degeneracy. Furthermore, for the two-state MSSV this filter is able to correctly track the regime changes. Nonetheless, given the complexity of the calculation of the conditional sufficient statistics, the APF + SS filter is not efficient in terms of computational time. As such we believe that compared to the other filters considered for analysis here the APF + SS the arguably the best.

Using real data examples we confirmed the applicability of the Bayesian filtering strategies presented here. We were able to fit and estimate the MSSV models with the APF + SS filter for both long and short return series. Results strongly suggested fitting two-state MSSV model in the three analyzed cases. For the IBOVESPA and S&P 500 analyses regime shifts matched key economic dates linked both to rises and decreases in market volatility.

Despite our concentration on stochastic volatility models, the APF + SS filter can be useful for many other statistical problems, such as (dynamic) factor models, space–time models, hierarchical models and several classes of time-series models, to name but a few. We believe that the flexibility of the APF + SS filter should be combined with careful (and potentially optimal) choice of resample-propagate proposal distributions (see [Carvalho *et al.*, 2010](#)) for efficient sequential learning in large and/or more complex dynamic systems.

One of the drawbacks of the SS estimation technique is that its application is highly dependent on fixed parameters admitting recursive conditional sufficient statistics. To overcome this a hybrid parameter learning technique can be implemented in situations where the vector of fixed parameters can be divided into two vector components, θ and η , where η , conditional on θ , admits recursive conditional sufficient statistics. More specifically, the prior for (θ, η) is

$$p(\theta, \eta) = p(\theta)p(\eta|x_0, \theta) = p(\theta)p(\eta|s_0, \theta) \quad (2.21)$$

and $s_t = \mathcal{S}(s_{t-1}, x_t, \theta, y_t)$ is a vector of recursive sufficient statistics, for $t = 1, \dots, n$.

References

1. Bodart, V., Kholodilin, K., Shadman-Mehta, F.: Identifying and forecasting the turning points of the Belgian business cycle with regime-switching and logit models. Working Paper, Universite catholique de Louvain, 2005
2. Bruno, G., Otranto, E.: Models to date the business cycle: The Italian case. *Economic Modelling*, **25**, 899–911, 2008
3. Carvalho, C.M., Johannes, M., Lopes, H.F., Polson, N.G.: Particle learning and smoothing. *Statistical Science*, **25**, 88–106, 2010
4. Carvalho, C.M., Lopes, H.F.: Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics & Data Analysis*, **51**, 4526–4542, 2007
5. Diebold, F.X., Rudebusch, G.D.: Scoring the Leading Indicators, *The Journal of Business*, University of Chicago Press, **62**(3), 369–91, 1989
6. Doucet, A., de Freitas, J., and Gordon, N. (eds.): *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001
7. Doucet, A., Johansen, A. M.: A tutorial on particle filtering and smoothing: Fifteen years later. In: Crisan, D., Rozovsky, B. (eds.) *Handbook of Nonlinear Filtering*, Oxford University Press, 2010
8. Eraker, B., Johannes, M., Polson, N.: The impact of jumps in volatility and returns. *Journal of Finance*, **58**, 1269–1300, 2003
9. Fearnhead, P.: Markov chain Monte Carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics*, **11**, 848–862, 2002
10. E. Ghysels, E., Harvey, A. C., Renault, E.: Stochastic volatility. In: Maddala, G.S., Rao, C.R. (eds) *Handbook of Statistics*, pp. 119–191. Amsterdam: North-Holland, 1996

11. Gordon, N., Salmond, D., Smith, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings*, **F-140**, 107–113, 1993
12. Jacquier, E., Polson, N.G., Rossi, P.E.: Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, **12**, 371–389, 1994
13. Kim, S., Shephard, N., Chib, S.: Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review, Econom. Stud.*, **65**, 361–393, 1998
14. Kong, A., Liu, J.S., W. Wong, W.: Sequential imputation and Bayesian missing data problems. *Journal of the American Statistical Association*, **89**, 590–599, 1994
15. Liu, J., West, M.: Combined parameter and state estimation in simulation-based filtering. In: Doucet, A., de Freitas, J., Gordon, N.J. (eds.), *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001
16. Lopes, H.F., Carvalho, C.M.: Online Bayesian learning in dynamic models: An illustrative introduction to particle methods. In: West, M., Damien, P., Dellaportas, P., Polson, N.G., Stephens, D.A. (eds.) *Bayesian Dynamic Modelling, Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian Smith*, Oxford University Press, 2011
17. Lopes H.F., Tsay, R.S.: Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting*, **30**, 168–209, 2011
18. Olsson, J., Cappé, O., Douc, R., Moulines, E.: Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models. *Bernoulli*, **14**, 155–179, 2008
19. Otranto, E.: The stock and Watson model with Markov switching dynamics: an application to the Italian business cycle. *Statistica Applicata*, **13**, 413–429, 2001
20. Pitt, M.K., Shephard, N., Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, **94**, 590–599, 1999
21. Polson, N.G., Stroud, J.R., Muller, P.: Practical filtering with sequential parameter learning. *Journal of the Royal Statistical Society, Series B*, **70**, 413–428, 2008
22. So, M., Lam, K., Li, W.K.: A stochastic volatility model with Markov switching. *Journal of Business & Economic Statistics*, **16**, 244–253, 1998
23. Storvik, G.: Particle filters in state space models with the presence of unknown static parameters. *IEEE: Trans. of Signal Processing*, **50**, 281–289, 2002
24. West, M.: Approximating posterior distributions by mixture. *Journal of the Royal Statistical Society, Series B*, **55**, 409–422, 1993



<http://www.springer.com/978-1-4614-7788-4>

State-Space Models

Applications in Economics and Finance

Zeng, Y.; Wu, S. (Eds.)

2013, XXI, 347 p., Hardcover

ISBN: 978-1-4614-7788-4