CHAPTER ON

BAYESIAN INFERENCE FOR STOCHASTIC VOLATILITY MODELING

Hedibert F. Lopes and **Nicholas G. Polson**
The University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, Illinois, 60637.
[hlopes,ngp]@ChicagoBooth.edu

Abstract

This chapter reviews the major contributions over the last two decades to the literature on the Bayesian analysis of stochastic volatility (SV) models (univariate and multivariate). Bayesian inference is performed by tailoring Markov chain Monte Carlo (MCMC) or sequential Monte Carlo (SMC) schemes that take into account the specific modeling characteristics.

The popular univariate stochastic volatility model with first order autoregressive dynamics (SV) is introduced in Section 1, which provides a detailed explanation of efficient MCMC and SMC algorithms. We briefly describe several extensions to the basic SV model that allows for fat-tailed, skewed, correlated errors as well as jumps (Markovian or not, smooth or not) in both observation and volatility equations, and the leverage effect via correlated errors.

Multivariate SV models are presented in Section 2 with particular emphasis on Wishart random processes, cholesky stochastic volatility models and factor stochastic volatility models. Section 3 contains several illustrations of both univariate and multivariate SV models based on both MCMC and SMC algorithms. Section 4 concludes the chapter.

1 Univariate SV models

Univariate stochastic volatility (SV) asset price dynamics results in the movements of an equity index S_t and its stochastic volatility v_t via a continuous time diffusion by a Brownian motion (Rosenberg, 1972, Taylor, 1986, Hull and White, 1987, Ghysels, Harvey and Renault, 1996, Johannes and Polson, 2010):

$$d\log S_t = \mu dt + \sqrt{v_t} dB_t^P \quad (1)$$
$$d\log v_t = \kappa(\gamma - \log v_t)dt + \tau dB_t^V \quad (2)$$

where the parameters governing the volatility evolution are $(\mu, \kappa, \gamma, \tau)$ and Brownian motions (B_t^P, B_t^V) possibly correlated. One extension of the above model is the stochastic volatility jump (SVJ) model that includes the possibility of jumps to asset prices. Here the equity index

S_t and its stochastic variance v_t replaces equation (1) by

$$d\log S_t = \mu dt + \sqrt{v_t} dB_t^P + d\left(\sum_{j=n_{t-1}}^{n_t} Z_j\right) \quad (3)$$

where the additional term in the above equity price evolution describes the jump process with jump sizes Z_j (Eraker, Johannes and Polson, 2003, and Johannes and Polson, 2010).We now show how to perform Bayesian inference for a wide class of models.

1.1 The SV model

Data arises in discrete time so it is natural to take an Euler discretization of equations (1) and (2). This is then commonly referred to as the *stochastic volatility autoregressive* (SV) model and is described by the following non-linear dynamic model (West and Harrison, 1997):

$$y_t = \exp\{x_t/2\}\varepsilon_t \quad (4)$$
$$x_t = \beta_0 + \beta_1 x_{t-1} + \tau\eta_t \quad (5)$$

where y_t are log-returns and log-variances $x_t = \log v_t$, ε_t and η_t iid standard normal errors. We take $\mu = 0$ for simplicity, $\beta_0 = \kappa\gamma$, $\beta_1 = 1 - \kappa$. The initial log-volatility state $x_0 \sim N(m_0, C_0)$, for known prior moments m_0 and C_0. An alternative specification assumes that $(x_0|\beta_0, \beta_1, \tau^2) \sim N(\beta_0/(1-\beta_1), \tau^2/(1-\beta_1^2))$ with $|\beta_1| < 1$; see Kalayloglu and Ghosh (2009) for Bayesian unit root tests regarding β. The centering parameterization moves β_0 to the observation equation and centers log-variances. This parameterization only marginally affects posterior inference in most cases while creating unnecessary computational burden. We will then keep the simpler, less restrictive, more general specification with m_0 and C_0.

The SV model is completed with a conjugate prior distribution for $\theta = (\beta, \tau^2)$, i.e. $p(\theta) = p(\beta|\tau^2)p(\tau^2)$, where $(\beta|\tau^2) \sim N(b_0, \tau^2 B_0)$ and $\tau^2 \sim IG(c_0, d_0)$, for known hyperparameters b_0, B_0, c_0 and d_0. An alternative specification where β and τ^2 are independent *a priori* can be easily implemented with negligible additional computational cost.

Given a set of observed asset returns $y^n = (y_1, \ldots, y_n)$ and equations (4) and (5), the posterior distribution of the hidden volatility states and parameters (x^n, θ) is given by Bayes rule

$$p(x^n, \theta|y^n) \propto p(\theta)\prod_{t=1}^{n} p(y_t|x_t, \theta)p(x_t|x_{t-1}, \theta), \quad (6)$$

which is analytically intractable because of the nonlinearity of equation (4). Approximate posterior inference for the SV model based on a Markov chain Monte Carlo algorithm and a sequential Monte Carlo (SMC) algorithm are discussed in the next two sections. We also provide several references on MCMC and SMC methods.

1.2 Posterior inference via Markov Chain Monte Carlo

Following the seminal paper of Jacquier, Polson and Rossi (1994), an abundance of Markov chain Monte Carlo (MCMC) algorithms have been proposed for the SV model as well as sev-

eral of its univariate and multivariate extensions. In this section we present one of these algorithms and argue that the derivations of majority of the existing alternatives/extensions follow roughly the same route. For further details see section 1.4 (univariate SV) and section 2 (multivariate SV). The MCMC algorithm cycles through the main two full conditional distributions, $p(\theta|y^n, x^n)$ and $p(x^n|y^n, \theta)$, in order to produce draws from $p(x^n, \theta|y^n)$ (Gamerman and Lopes, 2006, Migon, Gamerman, Lopes and Ferreira, 2005).

Sampling parameters. Sampling θ from its full conditional is rather standard since it is based on the Bayesian analysis of the normal linear regression (Gamerman and Lopes, Chapter 2). Given $y^t = (y_1, \ldots, y_t)$ and $x^t = (x_1, \ldots, x_t)$, for $t = 1, \ldots, n$, it is straightforward to show that the full conditional distribution of θ is given by

$$p(\theta|y^t, x^t) = p(\theta|s_t) = f_N(\beta; b_t, \tau^2 B_t) f_{IG}(\tau^2, c_t, d_t) \tag{7}$$

where $f_N(x; \mu, \sigma^2)$ is the density function of a normal distribution with mean μ and variance σ^2 and evaluate at point x. The sufficient statistics $s_t = (b_t, B_t, c_t, d_t)$ can be determined recursively as

$$b_t = B_t^{-1}(B_{t-1}^{-1} b_{t-1} + x_t z_t) \quad \text{and} \quad B_t^{-1} = B_{t-1}^{-1} + z_t z_t' \tag{8}$$

$$c_t = c_{t-1} + 1/2 \quad \text{and} \quad d_t = d_{t-1} + (x_t - b_t' z_t) x_t/2 + (b_{t-1} - b_t)' B_{t-1}^{-1} b_{t-1}/2, \tag{9}$$

for $z_t' = (1, x_{t-1})$. It is worth mention that we keep the recursive nature of these moments since it will be useful when deriving a SMC, or *particle filter*, in the next section.

Sampling states one at a time. Sampling x^n from its full conditional is a bit more complicated because of the nonlinearity in the observation equation (4). Jacquier, Polson and Rossi (1994) introduced the general MCMC algorithm to SV models that sample x_t one at a time, and conditional on $x_{-t} = (x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_n)$, from

$$\begin{aligned} p(x_t|x_{-t}, \theta, y^n) &\propto p_N(y_t; 0, e^{x_t}) p_N(x_t; \beta_0 + \beta_1 x_{t-1}, \tau^2) p_N(x_{t+1}; \beta_0 + \beta_1 x_t, \tau^2) \\ &\propto p_N(y_t; 0, e^{x_t}) f_N(x_t; \mu_t, \omega^2) \end{aligned}$$

as the conditional only depends on x_{t-1} and x_{t+1} and we can combine the state evolution densities. Here, $\mu_t = (\beta_0(1-\beta_1) + \beta_1(x_{t+1} + x_{t-1}))/(1+\beta_1^2)$, for $t = 1, \ldots, n-1$ $\mu_n = \beta_0 + \beta_1 x_{n-1}$ and $\omega^2 = \tau^2/(1+\beta_1^2)$. A simple random walk Metropolis algorithm with tuning variance v_x^2 and current state $x_t^{(j)}$ would work as follows. For $t = 1, \ldots, n$, sample x_t^* from $N(x_t^{(j)}, v_x^2)$ and accept the draw with probability

$$\alpha = \min \left\{ 1, \frac{f_N(x_t^*; \mu_t, \nu_t^2) f_N(y_t; 0, e^{x_t^*})}{f_N(x_t^{(j)}; \mu_t, \nu_t^2) f_N(y_t; 0, e^{x_t^{(j)}})} \right\}.$$

Alternatively, x_t could be sampled via a independent Metropolis-Hastings with a normal proposal density

$$q(x_t|x_{-t}, \theta, y^n) = f_N(x_t; \tilde{\mu}_t, \nu_t^2)$$

where $\tilde{\mu}_t = \mu_t + 0.5\nu_t^2(y_t^2 e^{-\mu_t} - 1)$ and $\nu_t^2 = \nu^2$ for $t = 1, \ldots, n-1$ and $\nu_n^2 = \tau^2$. The independent M-H algorithm would work as follows. For $t = 1, \ldots, n$ and current state $x_t^{(j)}$, sample x_t^* from $N(\tilde{\mu}_t, \nu_t^2)$ and accept the draw with probability

$$\alpha = \min\left\{1, \frac{f_N(x_t^*; \mu_t, \nu_t^2) f_N(y_t; 0, e^{x_t^*})}{f_N(x_t^{(j)}; \mu_t, \nu_t^2) f_N(y_t; 0, e^{x_t^{(j)}})} \times \frac{f_N(x_t^{(j)}; \tilde{\mu}_t, \nu_t^2)}{f_N(x_t^*; \tilde{\mu}_t, \nu_t^2)}\right\}.$$

It has been extensively argued that this is a rather inefficient route bounded to produce highly correlated chains and, consequently, failing to traverse the whole parameter space. The example of Section 3.1 illustrates the performance of both random walk M-H and independent M-H algorithms.

Sampling states jointly. When the model belongs to (or can be well approximated by) the class of conditionally normal dynamic linear models, then it is feasible to jointly sample from x^n from $p(x^n|y^n, \theta)$ recursively sampling x_n, then x_{n-1}, and so on:

$$p(x^n|y^n, \theta) = p(x_n|y^n, \theta) \prod_{t=1}^{n-1} p(x_t|x_{t+1}, \theta, y_t). \tag{10}$$

In the well-known class of normal dynamic linear model (NDLM), where $y_t|x_t \sim N(F_t' x_t, \sigma_t^2)$ and $x_t|x_{t-1} \sim N(G_t x_{t-1}, \tau_t^2)$, where the quadruple $\{F_t, G_t, \sigma_t^2, \tau_t^2\}$, for $t = 1, \ldots, n$ is known, F_t a vector of regressors, G_t driving the dynamic of x_t, and the initial distribution $(x_0|y^0) \sim N(m_0, C_0)$, it is straightforward to show that $x_t|y^{t-1} \sim N(a_t, R_t)$, $y_t|y^{t-1} \sim N(f_t, Q_t)$ and $x_t|y^t \sim N(m_t, C_t)$, for $t = 1, \ldots, n$. The means and variances of the three densities are provided by the Kalman recursions: $a_t = G_t m_{t-1}$, $R_t = G_t C_{t-1} G_t' + \tau_t^2$, $f_t = F_t' a_t$, $Q_t = F_t' R_t F_t + \sigma_t^2$, $m_t = a_t + A_t e_t$ and $C_t = R_t - A_t Q_t A_t'$, where $e_t = y_t - f_t$ is the prediction error and $A_t = R_t F_t Q_t^{-1}$ is the Kalman gain. Two other useful densities are conditional and marginal smoothed densities, i.e.

$$\begin{aligned} x_t|x_{t+1}, y^t &\sim N(h_t, H_t) \\ x_t|y^T &\sim N(m_t^T, C_t^T) \end{aligned}$$

where

$$\begin{aligned} h_t &= m_t + B_t(x_{t+1} - a_{t+1}) \\ H_t &= C_t - B_t R_{t+1} B_t' \\ m_t^T &= m_t + B_t(m_{t+1}^T - a_{t+1}) \\ C_t^T &= C_t - B_t^2(R_{t+1} - C_{t+1}^T) \end{aligned}$$

for $B_t = C_t G_{t+1}' R_{t+1}^{-1}$ $C_T^T = C_t$ and $m_T^T = m_T$ (West and Harrison, 1997, Chapter 4).

Kim, Shephard and Chib (1998) introduces an MCMC scheme that approximates the distribution of $\log y_t^2$ by a carefully tuned mixture of normals with seven components. More precisely, the observation equation (4) can be rewritten by

$$\log y_t^2 = x_t + \epsilon_t \tag{11}$$

where $\epsilon_t = \log \varepsilon_t^2$ follows a $\log \chi_1^2$ distribution, a parameter-free left skewed distribution with mean -1.27 and variance 4.94. They argue that $\epsilon = \log \chi_1^2$ can be well approximated by

$$p(\epsilon_t) = \sum_{i=1}^{7} \pi_i p_N(\epsilon_t; \mu_i, v_i^2) \tag{12}$$

where $\pi = (0.0073, 0.1056, 0.00002, 0.044, 0.34, 0.2457, 0.2575)$, $\mu = (-11.4, -5.24, -9.84, 1.51, -0.65, 0.53, -2.36)$ and $v^2 = (5.8, 2.61, 5.18, 0.17, 0.64, 0.34, 1.26)$. Therefore, a standard data augmentation argument allows the mixture of normals to be transformed in individual normals, i.e. $(\epsilon_t|k_t) \sim N(\mu_{k_t}, v_{k_t}^2)$ and $k_t \sim Multinomial(q)$. Conditionally on k^t, the SV model for $z_t = \log y_t^2 - \mu_{k_t}$ can be rewritten as a standard first order dynamic linear model, i.e.

$$\begin{aligned} (z_t|x_t, k_t, \theta) &\sim N(x_t, v_{k_t}^2) & (13)\\ (x_t|x_{t-1}, \theta) &\sim N(\beta_0 + \beta_1 x_{t-1}, \tau^2). & (14) \end{aligned}$$

Then, the standard *forward filtering, backward sampling* (FFBS) scheme of Carter and Kohn (1994) and Frühwirth-Schnatter (1994) can be implemented in order to jointly sample the vector of states x^n conditional on (y^n, k^n, θ). Finally, conditionally on x^n, the indicators k_t are sampled straightforwardly from $\{1, \ldots, 7\}$ with probability $Pr(k_t = j) \propto \pi_j p_N(z_t; x_t, v_j^2)$, for $t = 1, \ldots, n$. The example of Section 3.1 illustrates the performance of this algorithm.

1.3 Posterior inference via sequential Monte Carlo

Let us start by assuming that the vector of static parameter of the SV model, i.e. $\theta = (\beta_0, \beta_1, \tau^2)$ is known. Then, *particles filters* (PF) use Monte Carlo methods, mainly the sampling importance resampling (SIR), to sequentially reweigh and resample draws form the propagation density. The nonlinear Kalman filter is summarized by the predictive and smoothing steps:

$$\begin{aligned} p\left(x_t|y^{t-1}\right) &= \int f_N(x_t; \beta_0 + \beta_1 x_{t-1}, \tau^2) p\left(x_{t-1}|y^{t-1}\right) dx_{t-1} & (15)\\ p\left(x_t|y^t\right) &\propto p_N(y_t; 0, e^{x_t}) p(x_t|y^{t-1}). & (16) \end{aligned}$$

Particle filters, loosely speaking, combine the sequential estimation nature of Kalman-like filters with the flexibility for modeling of MCMC samplers, while avoiding some of the their shortcomings. On the one hand, like MCMC samplers and unlike Kalman-like filters, particle filters are designed to allow for more flexible observational and evolutional dynamics and distributions. On the other hand, like Kalman-like filters and unlike MCMC samplers, particle filters provide online filtering and smoothing distributions of states and parameters. The goal of most particle filters is to draw a set of i.i.d. particles $\{x_t^{(i)}\}_{i=1}^N$ that approximates $p(x_t|y^t)$ by starting with a set of i.i.d. particles $\{x_{t-1}^{(i)}\}_{i=1}^N$ that approximates $p(x_{t-1}|y^{t-1})$. The most popular filters are the bootstrap filter (BF), also known as sequential importance sampling with resampling (SISR) filter, proposed by Gordon, Salmond and Smith (1993), and the auxiliary particle filter (APF), also known as the auxiliary SIR (ASIR) filter, proposed by Pitt and Shephard (1999b).

The BF of Gordon *et al.* (1993) is based on sequential SIR steps over time (Smith and Gelfand, 1992). The Kalman recursions from (15) and (16) are combined in

$$p(x_t, x_{t-1}|y_t, y^{t-1}) \propto \underbrace{p_N(y_t; 0, e^{x_t})}_{2.Resample} \underbrace{p_N(x_t|\beta_0 + \beta_1 x_{t-1}, \tau^2) p(x_{t-1}|y^{t-1})}_{1.Propagate} \tag{17}$$

In words, the BF first propagates particles from the posterior at time $t-1$ in order to generate particles from the prior at time t. Then it resamples the propagated particles with weights proportional to their likelihoods. Similarly, the APF first resamples particles from the posterior at time $t-1$ with weights taking into account the next observed data point, y_t. Then, it propagates the resampled particles. They rewrite the identity from equation (17) as

$$p(x_t, x_{t-1}|y_t, y^{t-1}) \propto \underbrace{p(x_t|x_{t-1}, y^t)}_{2.Propagate} \underbrace{p(y_t|x_{t-1}) p_N(x_t|\beta_0 + \beta_1 x_{t-1}, \tau^2)}_{1.Resample}. \tag{18}$$

The main difficulty in implementing the APF in the SV case is that neither $p(y_t|x_{t-1})$ is available for pointwise evaluation nor $p(x_t|x_{t-1}, y^t)$ is available for sampling. Pitt and Shephard (1999b) suggests approximating $p(y_t|x_{t-1})$ and $p(x_t|x_{t-1}, y^t)$ by $p(y_t|g(x_{t-1}))$ and $p(x_t|x_{t-1})$, respectively, where $g(.)$ is usually the expected value, median or mode of $p(x_t|x_{t-1})$. In this case, the weights of the propagated particles are proportional to

$$w_t \propto \frac{p(y_t|x_t)}{p(y_t|g(x_{t-1}))}. \tag{19}$$

Bootstrap filter for the SV model

1. Propagate $\{x_{t-1}^{(i)}\}_{i=1}^N$ to $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ via $p_N(x_t|\beta_0 + \beta_1 x_{t-1}, \tau^2)$;
2. Resample $\{x_t^{(i)}\}_{i=1}^N$ from $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ with weights $w_t^{(i)} \propto p_N(y_t; 0, e^{\tilde{x}_t^{(i)}})$.

Auxiliary particle filter for the SV model

1. Resample $\{\tilde{x}_{t-1}^{(i)}\}_{i=1}^N$ from $\{x_{t-1}^{(i)}\}_{i=1}^N$ with weights $w_t^{(i)} \propto p_N(y_t; 0, e^{\beta_0 + \beta_1 x_{t-1}^{(i)}})$.
2. Propagate $\{\tilde{x}_{t-1}^{(i)}\}_{i=1}^N$ to $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ via $p_N(x_t; \beta_0 + \beta_1 \tilde{x}_{t-1}, \tau^2)$;
3. Resample $\{x_t^{(i)}\}_{i=1}^N$ from $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ with weights

$$w_t^{(i)} \propto \frac{p_N(y_t; 0, e^{\tilde{x}_t^{(i)}})}{p_N(y_t; 0, e^{\beta_0 + \beta_1 \tilde{x}_{t-1}^{(i)}})}.$$

Pitt and Shephard (1999a) suggest local linearization of the observation equation via an extended Kalman filter-type approximation in order to better approximate $p(x_t|x_{t-1}, y_t)$. See Doucet, Godsill and Andrieu (2000) and Guo, Wang and Chen (2005), amongst others, for

additional discussion on approximations based on local linearization. A more efficient approximation is based on the *mixture Kalman filters* of Chen and Liu (2000), when analytical integration of some components of the state vector is possible by conditioning on some other components. Such filters are commonly refereed to as *Rao-Blackwellized particle filter*. This is also acknowledged in Pitt and Shephard (1999b) and many other references.

Parameter learning involves the sequential and joint learning of x_t and θ. The immediate idea of simply resampling θ over time is bounded to fail since, in general, after a few time steps the particle set will contain only one particle. Gordon, Salmond and Smith (1993), in their seminal paper, suggest incorporating artificial evolution noise for θ when tackling the problem of sequentially learning the static parameters of a state space model. Here, for the sake of brevity, we derive only two well established filters for sequentially learning both x_t and θ in the SV context: the *Liu-West* filter of Liu and West (2001) and the *Particle Learning* of Carvalho, Johannes, Lopes and Polson (2010) and Lopes, Carvalho, Johannes and Polson (2010).

Liu and West filter. Liu and West (2001) combine (i) the APF of Pitt and Shephard (1999b), (ii) a kernel smoothing approximation to $p(\theta|y^{t-1})$ via a mixture of multivariate normals, and (iii) a neat shrinkage idea to incorporate artificial evolution for θ without the associated loss of information; see West (1993a,b). More specifically, let the set of i.i.d. particles $\{x_{t-1}^{(i)}, \theta_{t-1}^{(i)}\}_{i=1}^{N}$ approximates $p(x_{t-1}, \theta|y^{t-1})$ such that

$$p^N(\theta|y^{t-1}) \approx \frac{1}{N}\sum_{j=1}^{N} p_N(\theta; m^{(j)}, V) \tag{20}$$

where $m^{(j)} = a\theta_{t-1}^{(j)} + (1-a)\bar{\theta}$, $\bar{\theta} = \sum_{j=1}^{N}\theta_{t-1}^{(j)}/N$, $V = h^2\sum_{j=1}^{N}(\theta_{t-1}^{(j)} - \bar{\theta})(\theta_{t-1}^{(j)} - \bar{\theta})'/N$ and $h^2 = 1 - a^2$. The subscript t on θ_t is used only to indicate that samples are from $p(\theta|y^t)$. The APF of Pitt and Shephard (1999a) of equation (18) can now be written for the states vector (x_t, θ_t) as $p(x_t, x_{t-1}, \theta_t, \theta_{t-1}|y_t, y^{t-1})$ is decomposed into

$$\begin{aligned} \text{Resampling step} &: p(y_t|x_{t-1}, \theta_{t-1})p(x_{t-1}|\theta_{t-1}, y^{t-1})p(\theta_{t-1}|y^{t-1}) \\ \text{Propagation step} &: p(x_t|x_{t-1}, \theta_t, y^t)p(\theta_t|\theta_{t-1}, y^t). \end{aligned}$$

Again, $p(y_t|x_{t-1}, \theta)$ is not available for point-wise evaluation and/or $p(x_t|x_{t-1}, \theta_t, y^t)$ is not easy to sample from in the SV case. Liu and West (2001) follow Pitt and Shephard's (1999b) steps and resample from the proposal $p(y_t|g(x_{t-1}), m(\theta_{t-1}))$, where $g(\cdot)$ and $m(\cdot)$ are described above. Then, θ_t is sampled from the artificial transition $p(\theta_t|\theta_{t-1})$ and x_t sampled from the evolution density $p(x_t|x_{t-1}, \theta_t)$. The propagated particles (x_t, θ_t) have associated weights $\tilde{\omega}_t$ proportional to $p(y_t|x_t, \theta_t)/p(y_t|g(x_{t-1}), m(\theta_{t-1}))$.

The performance of the LW filter 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 set a around 0.98 or higher. The components of θ can be either transformed in order to accommodate the approximate local normality or the multivariate normal approximation could be replaced by a composition of, say, conditionally normal densities for location parameters and inverse-gamma densities for scale/variance parameters. See, for example, Petris *et al.* (2009,

pp. 222 - 228) for an example based on the local level model and Carvalho and Lopes (2007) for an application on Markov switching stochastic volatility models.

Liu and West filter for the SV model

1. Resample $\{(\tilde{x}_{t-1}, \tilde{\theta}_{t-1})^{(i)}\}_{i=1}^N$ from $\{(x_{t-1}, \theta_{t-1})^{(i)}\}_{i=1}^N$ with weights $w_t^{(i)} \propto p_N(y_t; 0, e^{m_0^{(i)} + m_1^{(i)} x_{t-1}^{(i)}})$ and $m_0^{(i)}$ and $m_1^{(i)}$ defined in Eq. (20).

2. Propagate

   (a) $\{\tilde{\theta}_{t-1}^{(i)}\}_{i=1}^N$ to $\{\hat{\theta}_t^{(i)}\}_{i=1}^N$ via $N(\tilde{m}^{(i)}, V)$, then

   (b) $\{\tilde{x}_{t-1}^{(i)}\}_{i=1}^N$ to $\{\hat{x}_t^{(i)}\}_{i=1}^N$ via $p_N(x_t; \hat{\beta}_0^{(i)} + \hat{\beta}_1^{(i)} \tilde{x}_{t-1}^{(i)}, \hat{\tau}^{2(i)})$.

3. Resample $\{(x_t, \theta_t)^{(i)}\}_{i=1}^N$ from $\{(\hat{x}_t, \hat{\theta})^{(i)}\}_{i=1}^N$ with weights

$$w_t^{(i)} \propto \frac{p_N(y_t; 0, e^{\hat{x}_t^{(i)}})}{p_N(y_t; 0, e^{\tilde{m}_0^{(i)} + \tilde{m}_1^{(i)} \tilde{x}_{t-1}^{(i)}})}.$$

Particle Learning. Carvalho, Johannes, Lopes and Polson (2010) and Lopes, Carvalho, Johannes and Polson (2010) introduce *Particle Learning* (PL) for particle filtering and parameter learning in a rather general state space models. They extend Chen and Liu's (2000) mixture Kalman filter (MKF) methods by allowing parameter learning and utilize the resample-propagate algorithm introduced by Pitt and Shephard (1999b), also in the pure filter context, together with a particle set that includes state sufficient statistics (Storvik, 2002, Fearnhead, 2002). Carvalho *et al.* (2010) and Lopes *et al.* (2010) empirically show that resample-propagate filters tend to outperform propagate-resample ones. They also show via several simulation studies that PL outperforms the LW filter and is comparable to MCMC samplers, even when fully adaptation is considered. The advantage is even more pronounced for large values of n.

For the basic SV model, PL takes advantage of the Kalman recursions produced by equations (11) to (14) and the recursive sufficient statistics (equations 7 to 9) of the conditionally dynamic linear model Recall that $s_t = (b_t, B_t, c_t, d_t)$ are the parameter sufficient statistics from equations (8) and (9) and let $s_t^x = (m_t, C_t)$ for m_t and C_t derived in the paragraph between equations (10) and (11). Both s_t and s_t^x satisfy deterministic updating rules, i.e. $s_t = \mathcal{S}(s_{t-1}, x_t, y_t)$, as in the Storvik filter from the previous subsection, and $s_t^x = \mathcal{K}(s_{t-1}^x, \theta, y_t)$, for $\mathcal{K}(\cdot)$ mimicking the Kalman filter recursions. The example of Section 3.1 illustrates the performance of particle filters introduced here. See Lopes and Tsay (2010) for a thorough review of particle filters via examples (and `R` code) for Bayesian inference in financial econometrics.

Particle Learning for the SV model

1. Resample $(s_{t-1}, s^x_{t-1}, \theta)$ with weights proportional to

$$p(\log y_t^2 | s^x_{t-1}, \theta) = \sum_{i=1}^{7} \pi_i p_N(\log y_t^2; \mu_i + \beta_0 + \beta_1 m_{t-1}, \beta_1^2 C_{t-1} + v_i^2 + \tau^2)$$

2. Sample (x_{t-1}, x_t) from $p(x_{t-1}, x_t | s^x_{t-1}, \theta, y^t)$:
   - Sample x_{t-1} from $p(x_{t-1} | s^x_{t-1}, \theta, y^t)$, and
   - Sample x_t from $p(x_t | x_{t-1}, \theta, y^t)$.
3. Update parameter sufficient statistics: $s_t = \mathcal{S}(\tilde{s}_{t-1}, x_t, y_t)$.
4. Sample θ from $p(\theta | s_t)$.
5. Update state sufficient statistics: $s^x_t = \mathcal{K}(\tilde{s}^x_{t-1}, \theta, y_t)$.

The distributions from step 2 are

$$\begin{aligned} p(x_{t-1} | s^x_{t-1}, \theta, y^t) &= \sum_{i=1}^{7} f_N(x_{t-1}, \hat{x}_{t-1,i}, V_{t-1,i}) \\ p(x_t | x_{t-1}, \theta, y^t) &= \sum_{i=1}^{7} f_N(x_t, \tilde{x}_{ti}, W_{ti}) \end{aligned}$$

where, from equation (12), $V_{t-1,i} = 1/(1/C_{t-1} + \beta_1^2/(v_i^2 + \tau^2))$, $\hat{x}_{t-1,i} = V^{-1}_{t-1,i}(m_{t-1}/C_{t-1} + (\log y_t^2 - \mu_i - \beta_0)\beta_1/(v_i^2 + \tau^2))$, $W_{ti} = 1/(1/v_i^2 + 1/\tau^2)$ and $\tilde{x}_{ti} = W_{ti}^{-1}((\log y_t^2 - \mu_i)/v_i^2 + (\beta_0 + \beta_1 x_{t-1})/\tau^2)$.

1.4 Other univariate SV models

Correlated errors. Jacquier, Polson and Rossi (2004) provide an MCMC algorithm for the leverage stochastic volatility (SVL) model. This extends the basic SV model to accommodate nonzero correlation ρ between ε_t and η_t from equations (4) and (5). Now the specification becomes

$$y_t = \exp\{x_{t-1}/2\} u_t \quad (21)$$

$$x_t = \beta_0 + \beta_1 x_{t-1} + \phi u_t + \omega v_t \quad (22)$$

$\phi = \tau\rho$ and $\omega^2 = \tau^2(1 - \rho^2)$ and u_t and v_t are iid standard normal errors. When $\rho < 0$ characterizes a *leverage effect*, so a negative shock in the observation y_t is associated to higher x_{t+h} for $h \geq 0$ and a positive shock in y_t is associated to lower x_t. They study weekly data on the equal and value weighted CRSP indices and daily data on the S&P500 and Deutsch Mark and Canadian dollar exchange rates relative to the US dollar. In their study, the posterior means of ρ range roughly between -0.48 and -0.2 for the daily data and between -0.47 and -0.41 for the weekly data. See also Omori and Watanabe (2008).

Fat-tailed, skewed and scale mixture of normals. Fat-tailed distribution for ε_t of equation (4) can be obtained by a continuous scale mixture of normals (Carlin and Polson, 1991, Geweke, 1993, Jacquier, Polson and Rossi, 2004).

$$\begin{aligned} y_t &= \exp\{x_t/2\}\varepsilon_t && (23)\\ x_t &= \beta_0 + \beta_1 x_{t-1} + \tau\eta_t && (24)\\ \varepsilon_t &= \sqrt{\lambda_t} z_t && (25)\\ \lambda_t &\sim IG(\nu/2, \nu/2), && (26) \end{aligned}$$

so that $\varepsilon_t \sim t_\nu(0,1)$, a standard Student's t distribution with ν degrees of freedom. The SV model with fat-tailed error can accommodate a wide range of kurtosis and is particularly important when dealing with extreme observations or outliers The example in Section 3.3 compares the SV and SV with t-errors models for monthly log returns of GE stock from January 1926 to December 1999 for 888 observations. Additional contributions to the theme are Steel (1998), Omori, Chib, Shephard and Nakajima (2007), Asai (2009), Nakajima and Omori (2009) and Abanto-Valle, Bandyopadhyay, Lachos and Enriquez (2009). Lopes and Polson (2010) provide a sequential analysis of this model.

Dirichlet process mixture. Jensen and Maheu (2008) use Dirichlet process mixture (DPM) prior to semi-parametrically model the observational error in the SV model ε_t in the SV model (equation 4):

$$\begin{aligned} \varepsilon_t &\sim N(0, \lambda_t^2) && (27)\\ \lambda_t^2|G &\sim G && (28)\\ G|G_0, \alpha &\sim DP(G_0, \alpha) && (29)\\ G_0(\lambda_t^2) &\equiv IG(\nu_0/2, \nu_0 s_0^2/2). && (30) \end{aligned}$$

where G_0 is the base distribution G0 and $\alpha > 0$ is the scalar precision parameter. They name this class of models the SV-DPM models and show that the above representation can be rewritten as

$$y_t|x_t \sim \sum_{j=1}^{\infty} \pi_j p_N(y_t; 0, \lambda_j^2 \exp\{x_t\}) \quad (31)$$

so revealing the nonparametric nature of the DPM prior with weights π_j derived by the *stick-break* recursion where $\pi_1 = \omega_1$ and $\pi_j = \omega_j \prod_{s=1}^{j-1}(1-\omega_s)$ where $\omega_j \sim Beta(1, \alpha)$. For more details on Bayesian nonparametric and semiparametric models see, amongst others, Dey, Müller and Sinha (1998), Ghosh and Ramamoorthi (2003), Hjort, Holmes, Müller and Walker (2010) and Carvalho, Lopes, Polson and Taddy (2009).

Long memory SV models. So (2002) and Jensen (2004) propose parametric and semiparametric Bayesian inference for long-memory SV models where the log-volatilities exhibit long-memory properties (LMSV):

$$\begin{aligned} y_t &= \exp\{x_t/2\}\varepsilon_t && (32)\\ (1-L)^d x_t &= \tau\eta_t && (33) \end{aligned}$$

where the fractional differencing operator is $(1-L)^d$, where L is the lag operator and $x_{t-s} = L^s x_t$, is defined by its binomial expansion. The MCMC/FFBS algorithm presented in Section 1.2 is not available when x_t follows a long-memory process. Jensen (2004) argues that the simulation smoother for a LMSV model is computationally very expensive and memory intensive. He goes on and propose a sampling scheme that takes advantage of the properties of the long-memory process orthonormal wavelet coefficients.

SV with jumps. Similar to the previous SV model, the Euler discretization of continuous time jump (SVJ) process leads to a specification of the form

$$\begin{aligned} y_t &= \exp\{x_t/2\}\varepsilon_t + J_t z_t && (34)\\ x_t &= \beta_0 + \beta_1 x_{t-1} + \tau\eta_t && (35)\\ J_t &\sim Ber(\lambda) && (36)\\ z_t &\sim N(\mu_z, \sigma_z^2), && (37) \end{aligned}$$

where J_t is the indicator of jump and Z_t the jump size. For the jump specification, one can use the conditionally conjugate prior structure for parameters $(\lambda, \mu_z, \sigma_z^2)$, where $\lambda \sim Beta(a,b)$, $\mu_z \sim N(c,d)$ and $\sigma_z^2 \sim IG(\nu/2, \nu\bar{\sigma}_z^2/2)$, respectively. For instance, when $c=-3$ and $d=0.01$ and $a=2$ and $b=100$ the prior mean and standard deviation of λ are around 0.02 and 0.014. The parameters ν and $\bar{\sigma}_z^2$ can be set, for instance, at 20 and 0.05, respectively, such that the prior mean and standard deviation of σ_z^2 are roughly 0.05 and 0.02. These prior specifications predict around five large negative jumps per year (roughly 250 business days) whose magnitude are around an additional three percent. This structure naturally leads to conditional posterior distributions that can be easily simulated to form a Gibbs sampler (Eraker, Johannes and Polson, 2003). The example of Section 3.4 estimate volatility with jumps for the S&P500 index and the Nasdaq NDX100 index to study the early part of the 2007-2008 credit crisis. In this case, jump probabilities are about 0.04 or 10 jumps per year, with the largest jump sizes around -2.14% for the S&P500 and -1.98% for the NDX100.

Additional Bayesian literature on SV jump models, continuous-time jump diffusion models and related models are Polson and Stroud (2003), Stroud, Müller and Polson (2003), Raggi (2005), Li, Wells and Yu (2006), Polson, Stroud and Müller (2008), Johannes, Polson and Stroud (2009), Li (2009) and Szerszen (2009).

Markov switching stochastic volatility. So, Lam and Li (1998) and Carvalho and Lopes (2007) propose MCMC and SMC algorithms, respectively, to estimate the Markov switching stochastic volatility model, which is an extension of the basic SV model to allow time-varying parameters in the dynamic of the log-volatilities, so equation (5) is replaced by (38) and the

model becomes:

$$
\begin{aligned}
y_t &= \exp\{x_t/2\}\varepsilon_t && (38)\\
x_t &= \beta_{0s_t} + \beta_1 x_{t-1} + \tau\eta_t && (39)\\
p_{ij} &= Pr(s_t = j | s_{t-1} = i) \qquad \text{for } i,j = 1,\ldots,k && (40)\\
\beta_{0s_t} &= \gamma_1 + \sum_{j=1}^{k} \gamma_j I_{jt} && (41)
\end{aligned}
$$

and regime variables s_t following a k-state first order Markov process, $I_{jt} = 1$ if $s_t \geq j$ and zero otherwise, γ_1 real and $\gamma_i > 0$ for $i > 1$. Carvalho and Lopes (2007) analyze the Brazilian Ibovespa stock index, from the São Paulo Stock Exchange, for daily data between 1997 and 2001. They are able to identify the major currency crises of the period, such as the Asian crisis in 1997, the Russian crisis in 1998 and the Brazilian crisis in 1999 all of which directly affected Brazil and other emerging economies.

Smooth transition SV models. Lopes and Salazar (2006a) extends the basic SV model by allowing smoothing transition in the the log-volatility dynamics (5). The first order logistic smooth transition autoregressive stochastic volatility (LSTAR-SV) model is

$$
\begin{aligned}
x_t &= \beta_{01} + \beta_{11} x_{t-1} + \pi(\gamma, c, s_t)(\beta_{02} + \beta_{12} x_{t-1}) + \tau\eta_t && (42)\\
\pi(\gamma, c, x_{t-d}) &= \frac{1}{1 + \exp\{\gamma(x_{t-d} - c)\}} && (43)
\end{aligned}
$$

The parameter $\gamma > 0$ is responsible for the smoothness of π, while c is a location or *threshold* parameter and d is the delay parameter. When $\gamma \to \infty$, the LSTAR model reduces to the well known self-exciting TAR (SETAR) model (Tong, 1990) and when $\gamma = 0$ the standard AR(k) model arises. Finally, s_t is called the transition variable, with $s_t = y_{t-d}$ commonly used (Teräsvirta, 1994, Lopes and Salazar, 2006b). Lopes and Salazar (2006a) compare several LSTAR-SV configurations when modeling the log-returns on the S&P500 index for roughly 3000 daily observed data between January 1986 and December 1997. See section 3.2 for more details.

Volatility-volume models. Mahieu and Bauer (1998) are among the first to perform Bayesian inference in the modified mixture model (MMM) of Andersen (1996) that model the volatilities based on a bivariate Gaussian-Poisson system both log-returns, y_t, and trading volume, v_t,

$$
\begin{aligned}
y_t | x_t, \theta &\sim N(0, \exp\{x_t\}) && (44)\\
v_t | x_t, \theta &\sim Poi(m_0 + m_1 \exp\{x_t\}) && (45)\\
x_t | x_{t-1}, \theta &\sim N(\beta_0 + \beta_1 x_{t-1}, \tau^2) && (46)
\end{aligned}
$$

where the parameter m_0 reflects the uninformed component of trading volume and is related to liquidity traders. The remaining part of trading volume that is induced by new information is represented by $m_1 \exp\{x_t\}$. Abanto-Valle, Migon and Lopes (2009) extend the model to allow

for Student's t errors (equations 25 and 26) and/or Markov switching dynamics (equations 39 to 41). They analyze daily closing prices and trading volume corrected by dividends and stock splits for the BP Company stock series listed on the London Stock Exchange (LSE), from 1999 to 2008 (around 2400 observations).

2 Multivariate SV models

Let y_t be a p-dimensional vector of (financial) time series. The majority of the existing multivariate stochastic volatility models assume that

$$y_t \sim N(0, \Sigma_t) \tag{47}$$

and focus on modeling the dynamic behavior of the covariance matrix Σ_t. Two challenges arise in the multivariate context. Firstly, the number of distinct elements of Σ_t equals $p(p+1)/2$. This quadratic growth has made the modeling Σ_t computationally very expensive and, consequently, has created up to a few years ago a practical upper bound for p. The vast majority of the papers we cite illustrate their methods and models with $p < 100$. Secondly, the distinct elements of Σ_t can not be modeled independently since positive definiteness has to be satisfied.

There are at least three ways to decompose the covariance matrix Σ_t. In the first case,

$$\Sigma_t = D_t R_t D_t \tag{48}$$

where D_t is a diagonal matrix with the standard deviations, $D_t = \text{diag}(\sigma_{1t}, \ldots, \sigma_{pt})$, and R_t is the correlation matrix. The above two challenges remain in this parametrization, i.e. the number of parameters increases with p^2 and R_t has to be positive definite. In the second case,

$$\Sigma_t = A_t H_t A_t' \tag{49}$$

where $A_t H_t^{1/2}$ is the lower triangular Cholesky decomposition of Σ_t. H_t is a diagonal matrix, the diagonal elements of A_t are all equal to one and, more importantly, its lower diagonal elements are unrestricted since positive definiteness is guaranteed by (49). Finally, in the third case (also the most popular) a standard factor analysis structure is used:

$$\Sigma_t = \beta_t H_t \beta_t' + \Psi_t \tag{50}$$

where β_t is the $p \times k$ matrix of factor loadings and, similar to A_t, is lower block triangular with diagonal elements equal to one. Ψ_t and H_t are the diagonal covariance matrices of the specific factors and common factors, respectively. One of the main reasons for the popularity of this decomposition, which became to be known as *factor stochastic volatility*, is that usually k is much smaller than p leading to a drastic reduction in the number of free parameters necessary to estimate Σ_t, i.e. $(k+1)p$. In two fairly realistic situations $(p,k) = (10,3)$ and $(p,k) = (100,10)$. In the first case, $p(p+1)/2 = 45$ and $p(k+1) = 40$, so the difference in number of parameters is not very significant. In the second example though, $p(p+1)/2 = 4950$ and $p(k+1) = 1100$, which translates to roughly 80% less parameters whose dynamics need to be estimated. Still under the first two decompositions, $p = 1000$ and $p = 5000$ generate of 0.5 and 13 million parameters, respectively, against 10% under the factor decomposition. A thorough review of the multivariate stochastic volatility literature up to a few years is provided in Asai, McAleer and Yu (2006).

2.1 Wishart random processes

Uhlig (1997) and Philipov and Glickman (2006a) proposed models for the covariance matrix based on the temporal update of the parameters of a Wishart distribution. See also Asai and McAleer (2009). Uhlig (1997) proposed the following recursion for the Cholesky decomposition of the precision matrix in structural vector autoregressions:

$$y_t = \sum_{i=1}^{q} B_i y_{t-i} + B_t \epsilon_t \tag{51}$$

where $B_t = A_t H_t^{1/2}$ (see the equation (49)), $\epsilon_t \sim N(0, I_p)$ and

$$\begin{aligned} \Sigma_t^{-1} &= \frac{\nu}{\nu+1} B_{t-1}^{-1} \Theta_{t-1} (B_{t-1}^{-1})' & (52) \\ \Theta_{t-1} &\sim Beta\left(\frac{\nu+pq}{2}, \frac{1}{2}\right), & (53) \end{aligned}$$

with $Beta$ denoting the the multivariate Beta distribution (Uhlig, 1994). See also Triantafyllopoulos (2008) for a similar derivation in the context of multivariate dynamic linear models. They model daily/current prices per tonne of aluminium, copper, lead and zinc exchanged in the London Metal Exchange from 4 January 2005 to 28 April 2006, or 334 trading days.

Philipov and Glickman's (2006a) *Wishart random process* is given by the observational equation (47) combined with equation (54) below:

$$\begin{aligned} (\Sigma_t^{-1} | \Sigma_{t-1}^{-1}, \theta) &\sim W(\nu, S_{t-1}^{-1}) & (54) \\ S_{t-1}^{-1} &= \frac{1}{\nu} (A^{1/2}) (\Sigma_{t-1}^{-1})^d (A^{1/2})' & (55) \end{aligned}$$

where $\theta = (\nu, A)$ and

$$\begin{aligned} E(\Sigma_t^{-1} | \Sigma_{t-1}^{-1}, \theta) &= (A^{1/2}) (\Sigma_t^{-1})^d (A^{1/2})' & (56) \\ E(\Sigma_t | \Sigma_{t-1}, \theta) &= \frac{\nu}{\nu-p-1} (A^{-1/2}) (\Sigma_{t-1})^d (A^{-1/2})'. & (57) \end{aligned}$$

A constant covariance model arises when $d = 0$, so $E(\Sigma_t) = \nu A^{-1}/(\nu - p - 1)$. Then, A plays the role of a precision matrix. When $d = 1$ and $A = I_p$, it follows that $E(\Sigma_t^{-1}) = \Sigma_{t-1}^{-1}$ so generating random walk evolution for the covariance. They fit their model to 240 monthly return data for $p = 5$ industry portfolios, so a relatively small dimensional problem.

2.2 Cholesky Stochastic Volatility (CSV)

Lopes, McCulloch and Tsay (2008) introduced the class of *Cholesky stochastic volatility* (CSV) models by exploring a triangular and recursive representation of the multivariate model in equation (47). More precisely, they use the decomposition (49) where

$$A_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1t} & a_{p2t} & \cdots & 1 \end{pmatrix} \qquad \Phi_t = A_t^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\phi_{21t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\phi_{p1t} & \phi_{p2t} & \cdots & 1 \end{pmatrix},$$

and $H_t = \text{diag}(h_{1t}, \ldots, h_{pt})$. The system $\Phi_t y_t \sim N(0, H_t)$ generates the following conditionally independent CSV recursions (for $t = 1, \ldots, n$):

$$y_{1t} = \exp\{h_{1t}/2\}\varepsilon_{1t} \quad (58)$$

$$y_{it} = \phi_{i1t} y_{1t} + \cdots + \phi_{i,i-1,t} y_{i-1,t} + \exp\{h_{it}\}\varepsilon_{it} \quad (i = 2, \ldots, p) \quad (59)$$

where $\varepsilon_t \sim N(0, I_p)$. The model is completed with standard SV structures for h_{it}, i.e. $\log h_{it} \sim N(\beta_{0i} + \beta_{1i} \log h_{i,t-1}, \tau_i^2)$, $i = 1, \ldots, p$, and first order autoregressive structures for $\phi_{ijt} \sim N(\beta_{0ij} + \beta_{1ij}\phi_{ij,t-1}, \tau_{ij}^2)$ for $i = 2, \ldots, p$ and $j = 1, \ldots, i-1$. They show that the prior on the parameters driving the dynamics of the hs and ϕs play an important role in producing more parsimonious models, particularly important when p is moderately large, say $p = 100$. In fact, there are $p(p+1)/2$ dynamic linear models to be estimated and, therefore, $3p(p+1)/2$ static parameters. When $p = 30$ and $p = 100$, for example, there are 465 and 5050 latent states, respectively, and 1395 and 15150 static parameters. Lopes, McCulloch and Tsay (2008) implement their model to the 100 components of the S&P100 index and 30 components of the Dow Jones Industrial Average index. See Dellaportas and Pourahmadi (2004) for a similar model for GARCH-type dynamics and $A_t = A$ for all t.

2.3 Factor stochastic volatility (FSV)

The literature on factor-based multivariate stochastic volatility models is now abundant, with Harvey, Ruiz and Shephard (1994), Pitt and Shephard (1999b), Aguilar and West (2000), Lopes and Migon (2002), Chib, Nardari and Shephard (2006) and Lopes and Carvalho (2007) just a few references. Loosely speaking, they model the levels (or first differences) of a set of (financial) time-series by a standard normal factor model (Lopes and West, 2004) in which both the common factor variances and the specific (or idiosyncratic) time-series variances are modeled as univariate or multivariate (of low dimension) SV processes. The main practical and computational advantage of the factor stochastic volatility (FSV) model is its parsimony, where all the variances and covariances of a vector of time-series are modeled by a low dimensional stochastic volatility structure dictated by common factors. It is fairly common to find that, for large vectors of time series, the number of common factors is usually one or two order of magnitude smaller, which speeds up computation and estimation considerably.

The large class of factor stochastic volatility (FSV) models, reviewed here and based on the decomposition of equation (50), is written as

$$(y_t | f_t, \beta_t, \Sigma_t) \sim N(\beta_t f_t; \Phi_t) \quad (60)$$

$$(f_t | H_t) \sim N(0; H_t) \quad (61)$$

where, as before, H_t is diagonal contains the variances of the common factors and Ψ_t is diagonal and contains the variances of the specific or idiosyncratic factors. The elements Ψ_t are modeled by conditionally independent univariate SV structures, while $\log h_t = (\log h_{1t}, \ldots, \log h_{kt})'$ follows a first-order vector autoregression:

$$(\log h_t | h_{t-1}, \beta_0, \beta_1, U) \sim N(\beta_0 + \beta_1 \log h_{t-1}, U) \quad (62)$$

with correlated innovations characterized by the non-diagonal matrix U (Aguilar and West, 2000). When U is a diagonal matrix, the above multivariate model is reduced to p univariate conditionally independent autoregressive models (Pitt and Shephard, 1999a). Both Pitt and Shephard (1999a) and Aguilar and West (2000) consider $\beta_t = \beta$ for all t time periods. Lopes (2000) and Lopes, Aguilar and West (2002) extend the previous works by modeling the evolution of the unconstrained loadings by univariate first order autoregressions. See Section 3.5 for a brief review of their exchange rate example.

Philipov and Glickman (2006b) extends the above FSV model (with $\Sigma_t = \Sigma$) and model H_t as a full covariance matrix via their Wishart random process (see equations 54 and 55). They implement their model on return series 324 monthly observations of 88 individual companies from the S&P500 and used $k = 2$ common factors. Han (2006) implements a similar FSV model to form portfolio based on 36 stocks, 1200 observations collected from the Center for Research in Security Prices (CRSP). Chib, Nardari and Shephard (2006) introduce fat-tailed errors and jumps in the FSV model as well as efficient and fast MCMC algorithm. They implement their extension to simulated data ($p = 50$) and and real data on international weekly stock index returns where $p = 10$ (see also Nardari and Scruggs, 2007). Finally, Lopes and Carvalho (2007) extend the FSV model to allow for Markovian regime shifts in the dynamic of the variance of the common factors and apply their model to study Latin America's main markets ($p = 5$).

2.4 Additional MSV references

Yu and Meyer (2006) compare several bivariate SV models, i.e. $p = 2$, when studying weekly data on the Australian dollar and the New Zealand dollar both against the US dollar for the period of January 1994 to December 2003. They use the deviance information criterion (DIC) of Spiegelhalter *et al.* (2002) and comparisons are, consequently, made via the Bayesian software `WinBUGS`[1]. On a related paper, Meyer and Yu (2000) used `BUGS`, which is an older version of `WinBUGS`, when comparing univariate SV model.

Asai, McAleer and Yu (2006) reviews the literature on specification, estimation, and evaluation of MSV models and divide the models according to various categories: (i) asymmetric models, (ii) factor models, (iii) time-varying correlation models, and (iv) alternative MSV specifications. Liesenfeld and Richard (2006) use efficient Importance Sampling (EIS) to perform Bayesian analysis of relatively low dimensional ($p = 4$) multivariate SV models.

3 Applications

In this section we illustrate the use of SV models in a series of different contexts. The first example compares random walk Metropolis-Hastings, independent Metropolis-Hastings, naïve

[1] `WinBUGS` is a Bayesian software whose development started two decades ago as part of the "Bayesian inference Using Gibbs Sampling (BUGS) project"' in the MRC Biostatistics Unit. `WinBUGS` can be freely downloaded from `http://www.mrc-bsu.cam.ac.uk/bugs`. See Spiegelhalter *et al.* (2003) and the previous webpage for more details.

normal approximation and mixture of seven normals approximation MCMC/FFBS via mixture of seven normals for the SV model for simulated data. A few variants of the SV model are applied in examples 3.2, 3.3 and 3.4. Example 3.2 deals with SV models with smooth transition between competing regimes, while example 3.3 models GE stock returns with normal and Student's t errors and compute sequential Bayes factors. In Example 3.4, the credit crisis of 2007-2008 is analyzed and monitored by particle filters. Finally, the popular and parsimonious class of factor stochastic volatility is used in example 3.5 to model multivariate exchange rate data.

3.1 Simple SV

This example illustrates the performance of four MCMC algorithms to estimate the parameters $(\beta_0, \beta_1, \tau^2)$ and states $x^n = (x_1, \ldots, x_n)$ given y^n in the SV model (Section 1.1). Namely, the random walk Metropolis-Hastings, independent Metropolis-Hastings, naïve normal approximation to $\log \chi_1^2$ and MCMC/FFBS via mixture of seven normals (Section 1.2) and Liu and West filter and particle learning (Section 1.3). A time series of length $n = 500$ is simulated from $x_0 = 0.0$, $\beta = (-0.00645, 0.99)'$ and $\tau^2 = 0.15^2$. Figure 1 exhibits the simulated time series and volatilities. The prior hyperparameters are $m_0 = -0.8$, $C_0 = 100$, $b_0 = (-0.013, 0.962)'$, $C_0 = 100I_2$, $c_0 = 5$ and $d_0 = 0.1078$.

The MCMC schemes are based on $M = 3000$ draws, after discarding the initial $M_0 = 1000$ draws. Posterior inference based on the four MCMC algorithms is summarized in figure 2. As expected, both random-walk and independent Metropolis-Hastings algorithms behave very similarly. In terms of mixing of the chains, both are outperformed by the FFBS based on the mixture of seven normals, in terms of mixing chains. The FFBS based on the normal approximation produce chains with good mixing properties, but the approximation to the marginal posterior distributions of the volatilities is rather crude.

3.2 SV with smooth transition

Lopes and Salazar (2006b) use LSTAR(k)-stochastic volatility models to analyze log-returns on the S&P500 index for roughly 3000 daily observed data between January 1986 and December 1997. They compared six SV models based on the Akaike's (1974) information criterion (AIC), the Schwarz's (1978) information criterion (BIC) and Spiegelhalter *et al.*'s (2002) deviance information criterion (DIC)[2]. The six models are: $\mathcal{M}_1 : AR(1)$, $\mathcal{M}_2 : AR(2)$, $\mathcal{M}_3 : LSTAR(1)$ with $d = 1$, $\mathcal{M}_4 : LSTAR(1)$ with $d = 2$, $\mathcal{M}_5 : LSTAR(2)$ with $d = 1$, and $\mathcal{M}_6 : LSTAR(2)$ with $d = 2$. The arrive at an $LSTAR(1)$ with $d = 1$ as the best model under three criteria. One can argue that the linear relationship prescribed by an AR(1) structure is insufficient to capture the dynamic behavior of the log-volatilities. The LSTAR structure brings more flexibility to the

[2]For data y and parameter θ, these criteria are defined as follows: $AIC = -2\log(p(y|\hat{\theta})) + 2p$ and $BIC = -2\log(p(y|\hat{\theta})) + p\log n$, p is the dimension of θ, sample size n and maximum likelihood estimator, $\hat{\theta}$. The DIC is defined as $DIC = D(\tilde{\theta}) + 2p_D = \bar{D} + p_D$, where $D(\theta) = -2\log p(y|\theta)$ is the *deviance*, $p_D = \bar{D} - D(\tilde{\theta})$, (measure of model complexity), $\tilde{\theta} = E(\theta|y)$ and $\bar{D} = E(D(\theta)|y)$ (measure of model fit).

modeling. Table 1 present the posterior mean and standard devations of all parameters for each one of the six models listed above.

	Models					
	$AR(1)$	$AR(2)$	$LSTAR(1)$	$LSTAR(1)$	$LSTAR(2)$	$LSTAR(2)$
			$d=1$	$d=2$	$d=1$	$d=2$
Parameter	Posterior mean (standard deviation)					
β_{01}	-0.060	-0.066	0.292	-0.354	-4.842	-6.081
	(0.184)	(0.241)	(0.579)	(0.126)	(0.802)	(1.282)
β_{11}	0.904	0.184	0.306	0.572	-0.713	-0.940
	(0.185)	(0.242)	(0.263)	(0.135)	(0.306)	(0.699)
β_{21}	-	0.715	-	-	-1.018	-1.099
		(0.248)			(0.118)	(0.336)
β_{02}	-	-	-0.685	0.133	4.783	6.036
			(0.593)	(0.092)	(0.801)	(1.283)
β_{12}	-	-	0.794	0.237	0.913	1.091
			(0.257)	(0.086)	(0.314)	(0.706)
β_{22}	-	-	-	-	1.748	1.892
					(0.114)	(0.356)
γ	-	-	118.18	163.54	132.60	189.51
			(16.924)	(23.912)	(10.147)	(0.000)
c	-	-	-1.589	0.022	-2.060	-2.125
			(0.022)	(0.280)	(0.046)	(0.000)
τ^2	0.135	0.234	0.316	0.552	0.214	0.166
	(0.020)	(0.044)	(0.066)	(0.218)	(0.035)	(0.026)
DIC	7223.1	7149.2	7101.1	7150.3	7102.4	7159.4

Table 1: *LSTAR-SV for S&P500:* Posterior means and posterior standard deviations for the parameters from all six entertained models plus deviance information criterion.

3.3 SV with fat-tailed errors

We revisit the simple SV model with normal innovations of Example 2 and compute sequential Bayes factor against the alternative SV model with Student-t innovations (Chib, Nardari and Shephard, 2002, Jacquier, Polson and Rossi, 2004). We assume initially that the number of degrees of freedom is known. We use monthly log returns of GE stock from January 1926 to December 1999 for 888 observations. This series was analyzed in Example 12.6 of Tsay (2005, ch. 12)[3]. The competing models are defined by the number of degrees of freedom:

$$\begin{aligned} \text{Observation equation} &: y_t|(x_t,\theta) \quad \sim t_\eta(0,\exp\{x_t\}), \\ \text{System equation} &: x_t|(x_{t-1},\theta) \quad \sim N(\alpha+\beta x_{t-1},\tau^2), \end{aligned}$$

[3] `http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts2/m-geln.txt`.

where $t_\eta(\mu, \sigma^2)$ denotes the Student-t distribution with η degrees of freedom, location μ and scale σ^2. The number of degrees of freedom η is treated as known. Sequential posterior inference is based on the Liu and West filter with $N = 100000$ particles. The shrinkage constant a is set at $a = 0.95$, whereas prior hyperparameters are $m_0 = 0$, $C_0 = 10$, $\nu_0 = 3$, $\tau_0^2 = 0.01$, $b_0 = (0, 1)'$ and $B_0 = 10I_2$. Particle approximation to the sequential posterior model probabilities, assuming uniform prior for η over models $\{t_\infty, t_2, \ldots, t_{20}\}$, appears in Figure 3, where t_∞ denotes normal distribution. Figure 3(d) shows percentiles of $p(\sigma_t|y^t)$ when integrating out over all competing models in $\{t_\infty, t_2, \ldots, t_{20}\}$. One can argue that the data slowly moves over time from a more t-like, heavy tail model towards a more Gaussian, thin tail model. Figures 4 and 5 present posterior summaries for the volatilities and parameters of a few competing models.

3.4 SV with jumps

The credit crisis of 2007-2008 is analyzed and monitored by particle filters in Lopes and Polson (2010). They sequentially estimate the volatility and examine the volatility dynamics for three major financial time series during the early part of the , namely the Standard and Poors S&P500 index, the Nasdaq NDX100 index and the financial index XLF. Sequential model choice is a natural outcome of our application and they show how the evidence in support of the SV with jumps model accumulates over time as market turbulence increases.

Figure 6 shows that before August 2007 the Bayes factor favors the stochastic volatility jump model. The market volatility risk premia is effectively constant for this model over this data period, except at the very end of the period where the implied option volatility decay quickly and the estimated volatility does not. This is also coincident with the Bayes factor decaying back in favor of the pure SV model for the NDX100 index. By the end of 2007, the odds favor the pure SV model over the SVJ model for the NDX100 index. For the XLF, most of the evidence for jumps is again contained in the February move. The sequential Bayes factor tends to lie in between the strong evidence for the S&P500 and weaker evidence for the NDX100 index. The story for the S&P500 is different. The sequential Bayes factor of figure 6 shows that after the February shock, the SVJ model is preferred to the SV model for the whole period. When comparing with VIX the jump model seems to track the option implied volatility with an appropriate market price of volatility risk.

3.5 FSV with time-varying loadings

Lopes (2000) and Lopes, Aguilar and West (2002) analyzed daily log-returns on weekday closing spot prices for six currencies relative to the US dollar: German Mark, British Pound, Japanese Yen, French Franc, Canadian Dollar and Spanish Peseta. They used the factor stochastic volatility model with time-varying loadings presented in Section 3.5. The data analyzed spans rom January 1st 1992 to October 31st 1995 in order to keep the analysis somewhat comparable Aguilar and West (2000). They consider a $k = 3$ factor model with relatively vague priors for all model parameters and run their MCMC scheme for 35000 iterations and several initial values. All chains converged, in practical terms, after around 20000 iterations.

An interesting observation that highlights the importance of time-varying loadings in the

context of this example is the change in the explanatory power of factor 1 (Figure 7), the "European factor" on the British Pound. The final months of 1992 marks the withdrawal of Great Britain from the European Union exchange-rate agreement (ERM), fact that is captured in our analysis by changes in the British loading in factor 1 and emphasized by the changes in the percentage of variation of the British Pound explained by factors 1 and 2 (Figure 8). If temporal changes on the factor loadings were not allowed, the only way the model could capture this change in Great Britain's monetary policy would be by a "shock" on the idiosyncratic variation of the Pound, reducing, in turn, the predictive ability of the latent factor structure.

4 Final remarks

This chapter reviews the major contributions over the last two decades to the literature on Bayesian analysis of stochastic volatility models, both in the univariate (Jacquier, Polson and Rossi, 1994) and the multivariate context (Shephard, 2005, Shephard and Andersen, 2009). Posterior inference for the majority of the models is performed by tailored MCMC schemes that take into account specific modeling characteristics. Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998) are amongst the most influential contributions when dealing with univariate SV models, with Jacquier, Polson and Rossi (1995), Pitt and Shephard (1999a) and Aguilar and West (2000) playing similar roles in the multivariate case.

These and other early contributions have, over the last decade, played fundamental role in helping shaping the field of financial time series and econometrics. See Johannes and Polson (2010), for instance, for a thorough review of MCMC methods for continuous-time financial econometrics. Factor and Cholesky stochastic volatility models for high dimensional systems have also become fairly popular. See, amongst others, Chib, Nardari and Shephard (2006), Lopes and Carvalho (2007) and Lopes, McCulloch and Tsay (2008).

Markov chain Monte Carlo methods, again over the last decade, have started to share the Bayesian computational stage with efficient sequential Monte Carlo methods, with a detailed illustration in the SV context introduced in Section 1.3 along with additional SMC references. Other successful implementation in the univariate SV literature are Pitt and Shephard (1999b), Stroud, Polson and Müller (2004), Carvalho and Lopes (2007) and Johannes, Polson and Stroud (2009), to name but a few. Berg, Meyer and Yu (2004) and Raggi and Bordignon (2006) compare the performance of several univariate SV models. In the multivariate SV case, Liu and West (2001) and Lopes (2000, Chapter 6) implement particle filter with parameter learning for two variants of the FSV model.

References

Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V. H. and Enriquez (2009). Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics and Data Analysis* (published online on 26 June 2009).

Abanto-Valle, C. A., Migon, H. S. and Lopes, H. F. (2009). Bayesian modeling of financial

returns: A relationship between volatility and trading volume. *Applied Stochastic Modeling in Business and Industry* (published online on 8 Jun 2009).

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics*, **18**, 338-57.

Akaike, H. (1974). New look at the statistical model identification. *IEEE Transactions in Automatic Control*, **19**, 716-23.

Andersen, T. (1996). Return volatility and trading volume: an information flow interpretation of stochastic volatility. *Journal of Finance*, **51**, 169-204.

Asai, M. (2009). Bayesian analysis of stochastic volatility models with mixture-of-normal distributions. *Mathematics and Computers in Simulation*, **79**, 2579-96.

Asai, M. and McAleer, M. (2009). The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics*, **150**, 182-92.

Asai, M., McAleer, M. and Yu, Jun. (2006). Multivariate stochastic volatility: a review. *Econometric Reviews*, **25**, 145-75.

Berg, A., Meyer, R. and Yu, J. (2004). Deviance Information Criterion for Comparing Stochastic Volatility Models. *Journal of Business & Economic Statistics*, **22**, 107-20.

Carlin, B. P. and Polson, N. G. (1991). Inference for non-conjugate Bayesian models using the Gibbs sampler. *Canadian Journal of Statistics*, **19**, 399-405.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, **81**, 541-53.

Carvalho, C., Johannes, M. Lopes, H. F. and Polson, N. G. (2010). Particle learning and smoothing. *Statistical Science* (to appear).

Carvalho, C., Lopes, H. F., Polson, N. G. and Taddy, M. (2009). Particle Learning for General Mixtures. *Working Paper*. University of Chicago Graduate School of Business.

Carvalho, C. and Lopes, H. F. (2007) Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics & Data Analysis*, **51**, 4526-42.

Chen, R. and Liu, J. S. (2000). Mixture Kalman filter. *Journal of the Royal Statistical Society, Series B*, **62**, 493-508.

Chib, S., Nardari, F. and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, **108**, 281-316.

Chib, S., Nardari, F. and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, **134**, 341-71.

Dellaportas, P. and Pourahmadi, M. (2004). Large Time-Varying Covariance Matrices with Applications to Finance. *Working paper*, Department of Statistics, Athens University of Economics and Business.

Dey, D., Müller, P. and Sinha, D. (1998). *Practical Nonparametric and Semiparametric Bayesian Statistics*. New York: Springer-Verlag.

Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte-Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, **10**, 197-208.

Eraker, B., Johannes, M. and Polson, N.G. (2003). The impact of jumps in volatility and returns. *Journal of Finance*, **59**, 227-60.

Fearnhead, P. (2002). Markov chain Monte Carlo, sufficient statistics and particle filter. *Journal of Computational and Graphical Statistics*, **11**, 848-62.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 183-202.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd edition)*. Chapman & Hall/CRC.

Geweke, J. (1993) Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, **8**, S19S40.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. New York: Springer-Verlag.

Ghysels, E., Harvey, A. C. and Renault, E. (1996). Stochastic Volatility. In Rao, C. R. and Maddala, G. S., editors, *Handbook of Statistics: Statistical Methods in Finance*. Amsterdam: North-Holland, 119-91.

Gordon, N., Salmond, D. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, **140**, 107-13.

Guo, D., Wang, X. and Chen, R. (2005). New sequential Monte Carlo methods for nonlinear dynamic systems. *Statistics and Computing*, **15**, 135-47.

Han, Y., (2006). Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model. *The Review of Financial Studies*, **19**, 237-71.

Harvey, A.C., Ruiz, E., Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247-64.

Hjort, N. L., Holmes, C., Müller, P. and Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge: Cambridge University Press.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, **42**, 281-300.

Jacquier, E., Polson, N. G. and Rossi. P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, **20**, 69-87.

Jacquier, E., Polson, N. G. and Rossi. P. E. (1995). Models and priors for multivariate stochastic volatility models. *Working paper*, The University of Chicago Booth School of Business.

Jacquier, E., Polson, N. G. and Rossi. P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, **122**, 185-212.

Jensen, M. J. (2004). Semiparametric Bayesian Inference of Long-Memory Stochastic Volatility. *Journal of Time Series Analysis*, **25**, 895-922.

Jensen, M. J. and Maheu, J. M. (2008). Bayesian Semiparametric Stochastic Volatility Modeling *Working Paper*, Federal Reserve Bank of Atlanta.

Johannes, M. and Polson, N. (2010). MCMC methods for continous-time financial econometrics. In YAït-Sahalia, Y. and Hansen, L. P., editors, *Handbook of Financial Econometrics, Volume 2*. Princeton: University Press, 1-72.

Johannes, M. S., Polson, N. G. and Stroud, J. R. (2009). Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices. *Review of Financial Studies*, **22**, 2559-99.

Kalayloglu, Z. I. and Ghosh, S. K. (2009). Bayesian unit-root tests for Stochastic Volatility models. *Statistical Methodology*, **6**, 189-201.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic Volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, **65**, 361-93.

Li, H., Wells, M. T. and Yu, C. L. (2006). A Bayesian Analysis of Return Dynamics with Lévy Jumps. *Review of Financial Studies*, **21**, 2345-78.

Li, H. (2009). Sequential Bayesian Analysis of Time-Changed Infinite Activity Derivatives Pricing Models. *Working paper*. ESSEC Business School, Paris-Singapore.

Liesenfeld, R. and Richard, J.-F. (2006). Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Review*, **25**, 335-60.

Liu, J. and West, M. (2001). Combined parameters and state estimation in simulation-based filtering. In Doucet, A., De Freitas, N. and Gordon, N., editors, *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 197-223.

Lopes, H. F. (2000). *Bayesian analysis in latent factor and longitudinal models*. Unpublished Ph.D. Thesis. Institute of Statistics and Decision Sciences, Duke University, USA.

Lopes, H. F., Aguilar, O. and West, M. (2002). Time-varying covariance structures in currency markets. *Working paper*, Department of Statistical Methods, Federal University of Rio de Janeiro.

Lopes, H. F. and Carvalho, C. M. (2007). Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*, **137**, 3082-91.

Lopes, H. F., Carvalho, C. M., Johannes, M. and Polson, N. G. (2010) Particle Learning for sequential Bayesian computation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, editors, *Bayesian Statistics 9*. Oxford: Oxford University Press (to appear).

Lopes, H. F., McCulloch, R. E. and Tsay, R. (2008). Choleski Multivariate Stochastic Volatility. *Working paper*, The University of Chicago Booth School of Business.

Lopes, H. F. and Migon, H. S. (2002). Comovements and contagion in emergent markets: stock indexes volatilities. *Case Studies in Bayesian Statistics*, **6**, 285-300.

Lopes, H. F. and Polson, N. G. (2010). Extracting SP500 and NASDAQ volatility: The credit crisis of 2007-2008. In OHagan, A. and West, M., editors, *The Oxford Handbook of Applied Bayesian Analysis*. Oxford: Oxford University Press, 319-42.

Lopes, H. F. and Salazar, E. (2006a). Time series mean level and stochastic volatility modeling by smooth transition autoregressions: a Bayesian approach. In Fomby, T. B., editor, *Advances in Econometrics: Econometric Analysis of Financial and Economic Time Series, Volume 20, Part B*. Amsterdam: Elsevier, 229-42.

Lopes, H. F. and Salazar, E. (2006b) Bayesian model uncertainty in smooth transition autoregressions. *Journal of Time Series Analysis*, **27**, 99-117.

Lopes, H. F. and Tsay, R. S. (2010). Particle Filters and Bayesian Inference In Financial Econometrics. *Journal of Forecasting* (to appear).

Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, **14**, 41-67.

Mahieu, R. and Bauer, R. (1998). A Bayesian analysis of stock return volatility and trading volume. *Applied Financial Economics*, **8**, 671-87.

Meyer, R. and Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal*, **3**, 198-215.

Migon, H. S., Gamerman, D., Lopes, H. F. and Ferreira, M. A. R. (2005). Dynamic models. In Dey, D. and Rao, C.R., editors, *Handbook of Statistics, Volume 25: Bayesian Thinking, Modeling and Computation*. Amsterdam: Elsevier, 553-88.

Nakajima, J. and Omori, Y. (2009). Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics & Data Analysis*, **53**, 2335-53.

Nardari, F. and Scruggs, J. T. (2007). Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions. *Journal of Financial and Quantitative Analysis*, **42**, 857-92.

Omori, Y. and Watanabe, T. (2008). Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics and Data Analysis*, **52**, 2892-910.

Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, **140**, 425-49.

Petris, G., Petrone, S., Campagnoli, P. (2009). *Dynamic Linear Models with R*, New York: Springer.

Philipov, A. and Glickman, M. E. (2006a). Multivariate stochastic volatility via Wishart processes. *Journal of Business and Economic Statistics*, **24**, 313-28.

Philipov, A. and Glickman, M. E. (2006b). Factor multivariate stochastic volatility via Wishart processes. *Econometric Reviews*, **25**, 311-34.

Pitt, M. and Shephard, N. (1999a). Time varying covariances: a factor stochastic volatility approach (with discussion). In Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., editors, *Bayesian Statistics 6*. Oxford: University Press, 547-70.

Pitt, M. and Shephard, N. (1999b). Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association*, 94, 590-9.

Polson, N. G. and Stroud, J. R. (2003). Bayesian Inference for Derivative Prices. Bernardo, J. M, Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M. and West, M., editors, *Bayesian Statistics 7*. Oxford: University Press, 641-50.

Polson, N. G., Stroud, J. R. and Müller, P. (2008). Practical Filtering with Sequential Parameter Learning. *Journal of the Royal Statistical Society, Series B*, **70**, 413-28.

Raggi, D. (2005). Adaptive MCMC methods for inference on affine stochastic volatility models with jumps. *Econometrics Journal*, **8**, 235-50.

Raggi, D. and Bordignon, S. (2006). Comparing stochastic volatility models through Monte Carlo simulations. *Computational Statistics & Data Analysis*, **50**, 1678-99.

Rosenberg, B. (1972). The behaviour of random variables with nonstationary variance and the distribution of security prices. *Working Paper*.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-4.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: University Press.

Shephard, N. and Andersen, T. A. (2009). Stochastic Volatility: Origins and Overview. In Andersen, T. G., Davis, R. A. and Kreiss, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*. New York: Springer.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling-resampling perspective. *American Statistician*, **46**, 84-8.

So, M. K. P. (2002) Bayesian analysis of long memory stochastic volatility models. *Sankhya: The Indian Journal of Statistics*, **64**, 1-10.

So, M. K. P., Lam, K. and Li, W. K. (1998). A stochastic volatility model with Markov switching. *Journal of Business and Economic Statistics*, **16**, 244-53.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583-639.

Spiegelhalter, D. J., Thomas, A., Best, N. G. and Gilks, W. R. (2003). WinBUGS User Manual (Version 1.4). MRC Biostatistics Unit, Cambridge, UK.

Steel, M. F. J. (1998). Bayesian analysis of stochastic volatility models with flexible tails. *Econometric Reviews*, **17**, 109-43.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, **50**, 281-89.

Stroud, J. R., Müller, P. and Polson, N. G. (2003). Nonlinear State-Space Models with State-Dependent Variances *Journal of the American Statistical Association*, **98**, 377-86.

Stroud, J. R., Polson, N. G. and Müller, P. (2004). Practical Filtering for Stochastic Volatility Models. In Harvey, A., Koopman, S. J. and Shephard, N., editors, *State Space and Unobserved Components Models*, 236-47. Cambridge: University Press.

Szerszen, P. J. (2009). Bayesian Analysis of Stochastic Volatility Models with Lévy Jumps: Application to Risk Analysis. *Working paper*. Divisions of Research & Statistics and Monetary Affairs. Federal Reserve Board, Washington, D.C.

Taylor, S. J. (1986). *Modelling Financial Time Series*. New York: Wiley.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208-18.

Tong, H. (1990). *Non-linear time series: A dynamical systems approach*. Oxford: Oxford University Press.

Triantafyllopoulos, K. (2008). Multivariate stochastic volatility with Bayesian dynamic linear models. *Journal of Statistical Planning and Inference*, **138**, 1021-37.

Uhlig, H. (1994). On Singular Wishart and Singular Multivariate Beta Distributions. *The Annals of Statistics*, **22**, 395-405.

Uhlig, H. (1997). Bayesian Vector Autoregressions with Stochastic Volatility. *Econometrica*, 65, 59-73.

West, W. (1993a). Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society, Series B*, **54**, 553-68.

West, M. (1993b). Mixture models, Monte Carlo, Bayesian updating and dynamic models. *Computing Science and Statistics*, **24**, 325-33.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models (2nd edition)*. New York: Springer.

Yu, J. and Meyer, R. (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, **25**, 361-84.

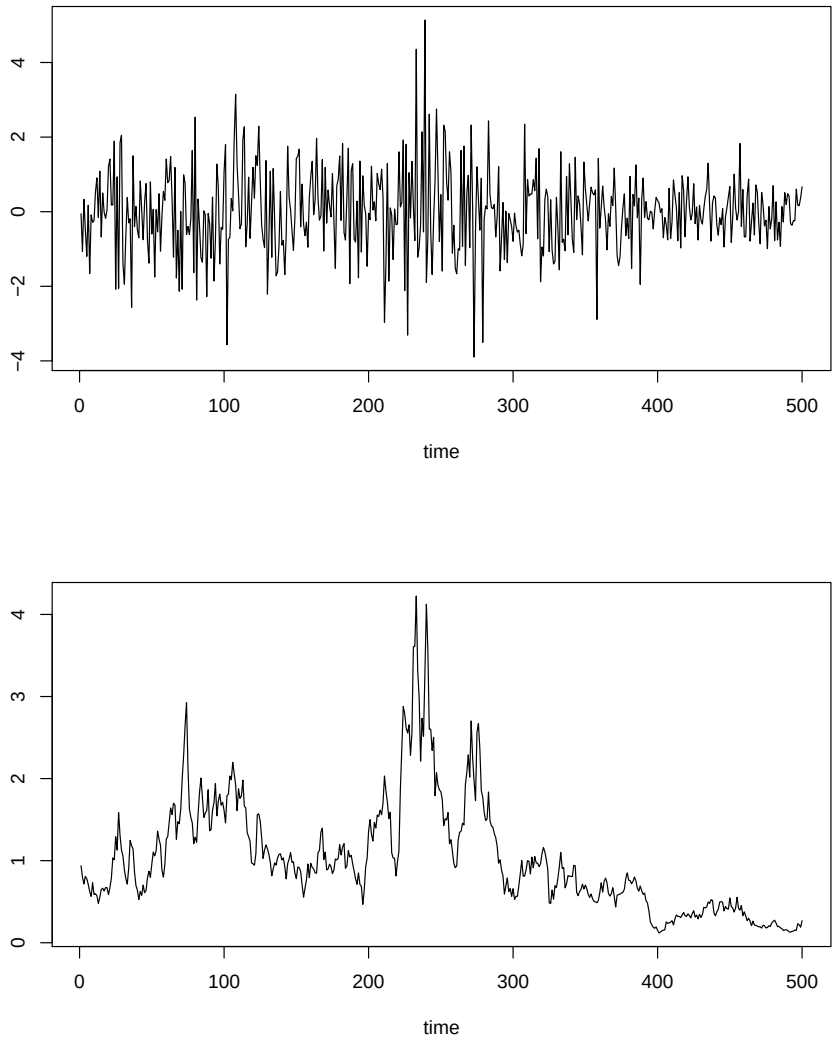

Figure 1: *Simulated SV data.* Time series of length $n = 500$ is based on $x_0 = 0.0$, $\beta = (-0.00645, 0.99)'$ and $\tau^2 = 0.15^2$. *Top:* simulated data. *Bottom:* simulated volatilities.

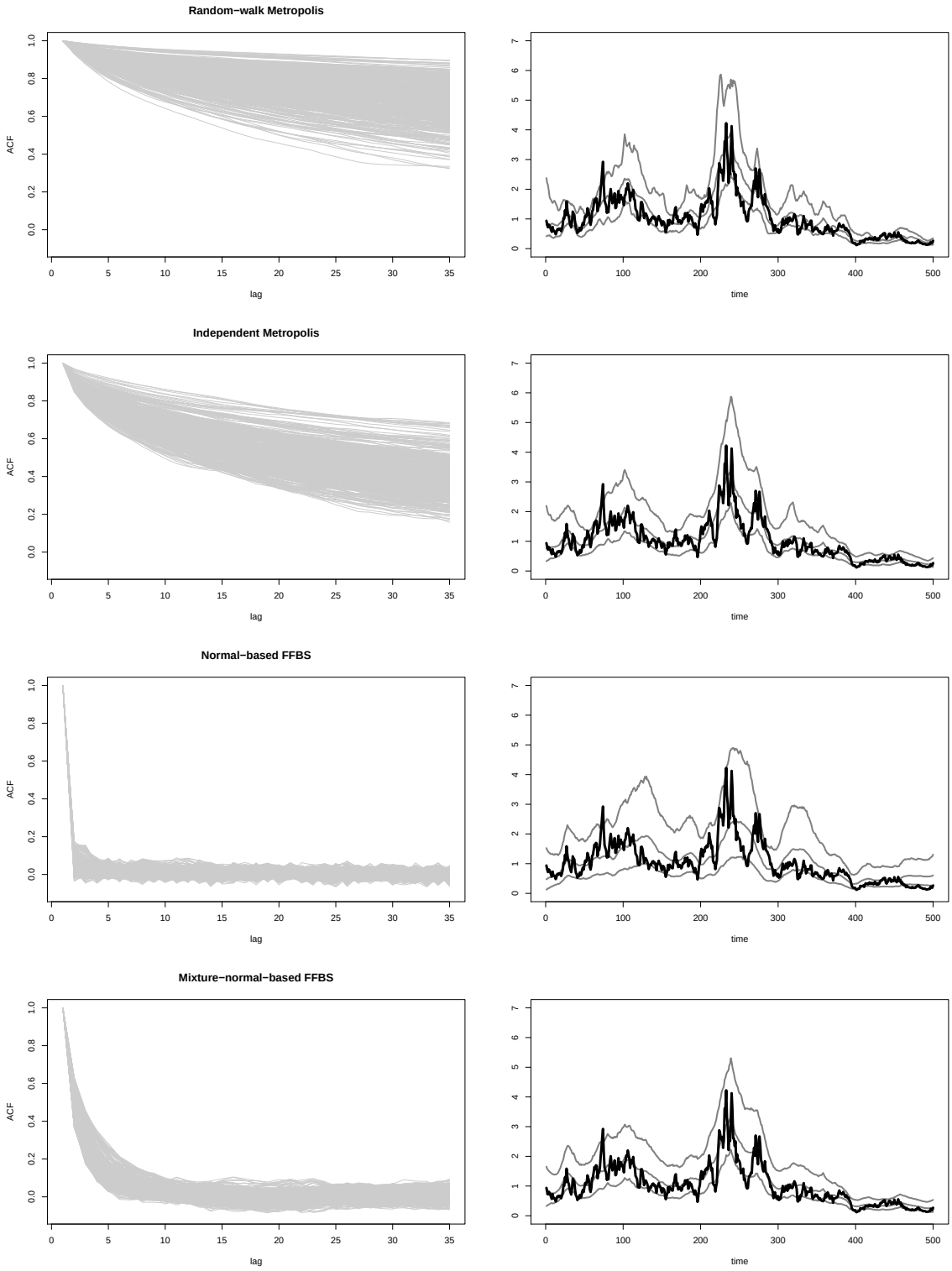

Figure 2: *Simulated SV data - MCMC. Left column:* Chain autocorrelations for each x_t, $t = 1, \ldots, n$. *Right column:* The 2.5th, 50th and 97.5th percentiles of $p(\sigma_t|y^n)$. Rows are the random-walk Metropolis, the independent Metropolis, the normal-based FFBS and the mixture-normal-based FFBS.

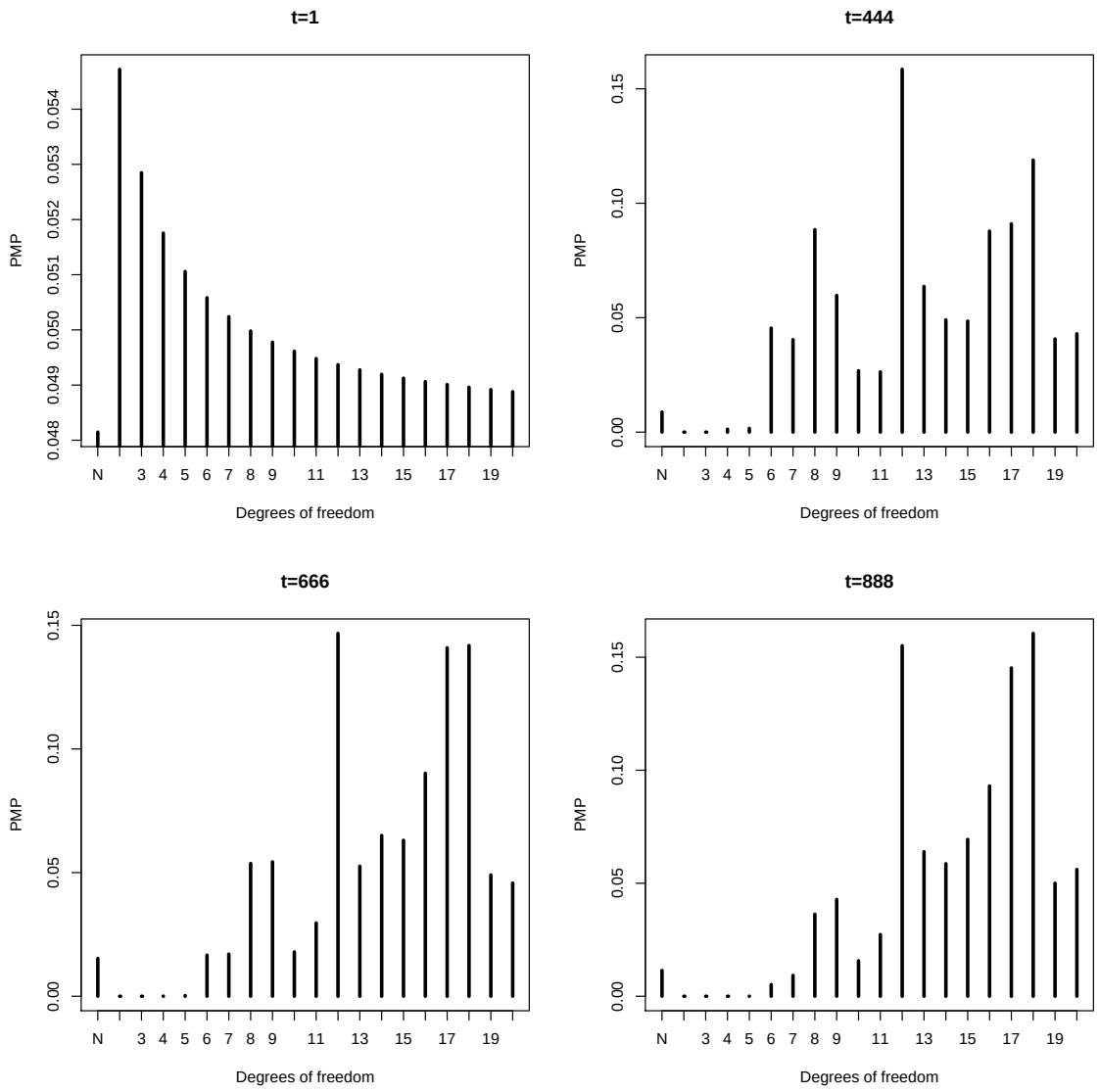

Figure 3: *Stochastic volatility model.* Sequential posterior model probability for the number of degrees of freedom η.

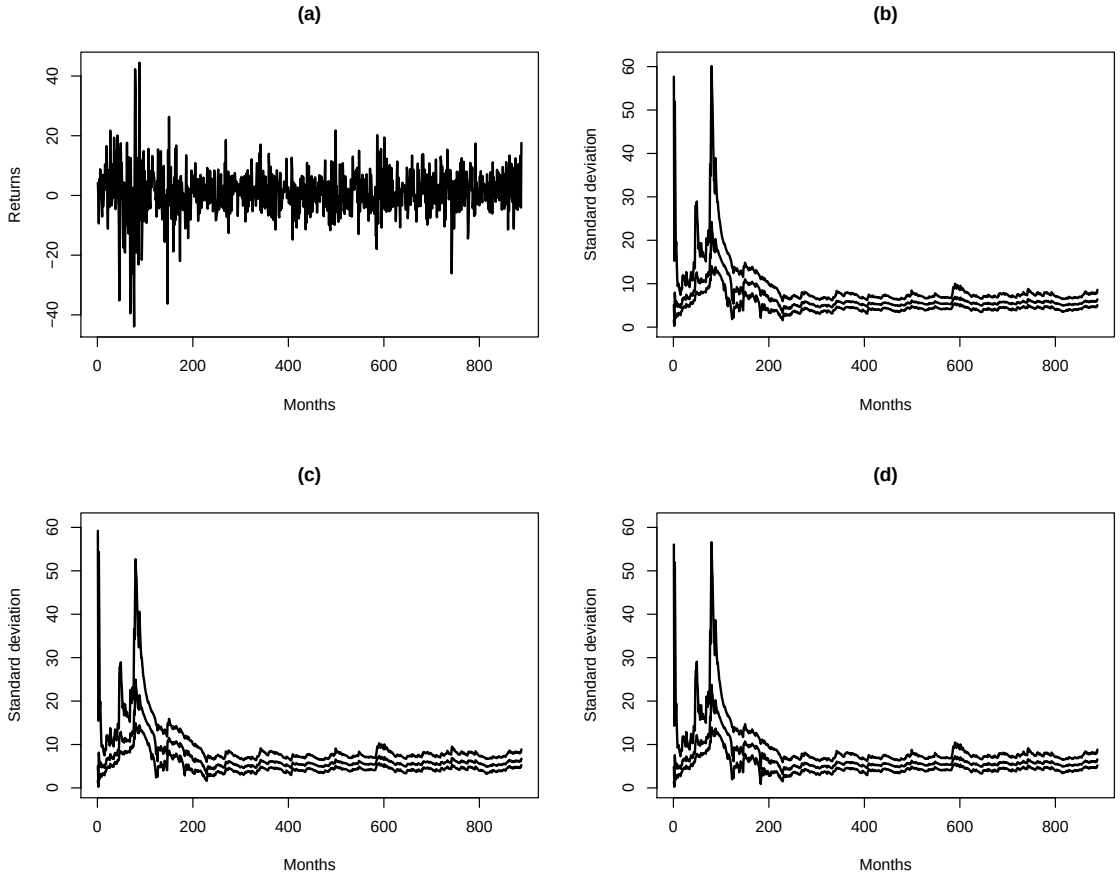

Figure 4: *Stochastic volatility model.* (a) GE returns; (b) an (c) 2.5th, 50th and 97.5th percentiles of $p(\sigma_t|y^t, M)$, where $\sigma_t^2 = \exp\{x_t\}$, for $M = t_{12}$ and $M = t_{18}$, respectively. (d) 2.5th, 50th and 97.5th percentiles of $p(\sigma_t|y^t)$ by integrating out over all competing models in $\{\text{Normal}, t_2, \ldots, t_{20}\}$.

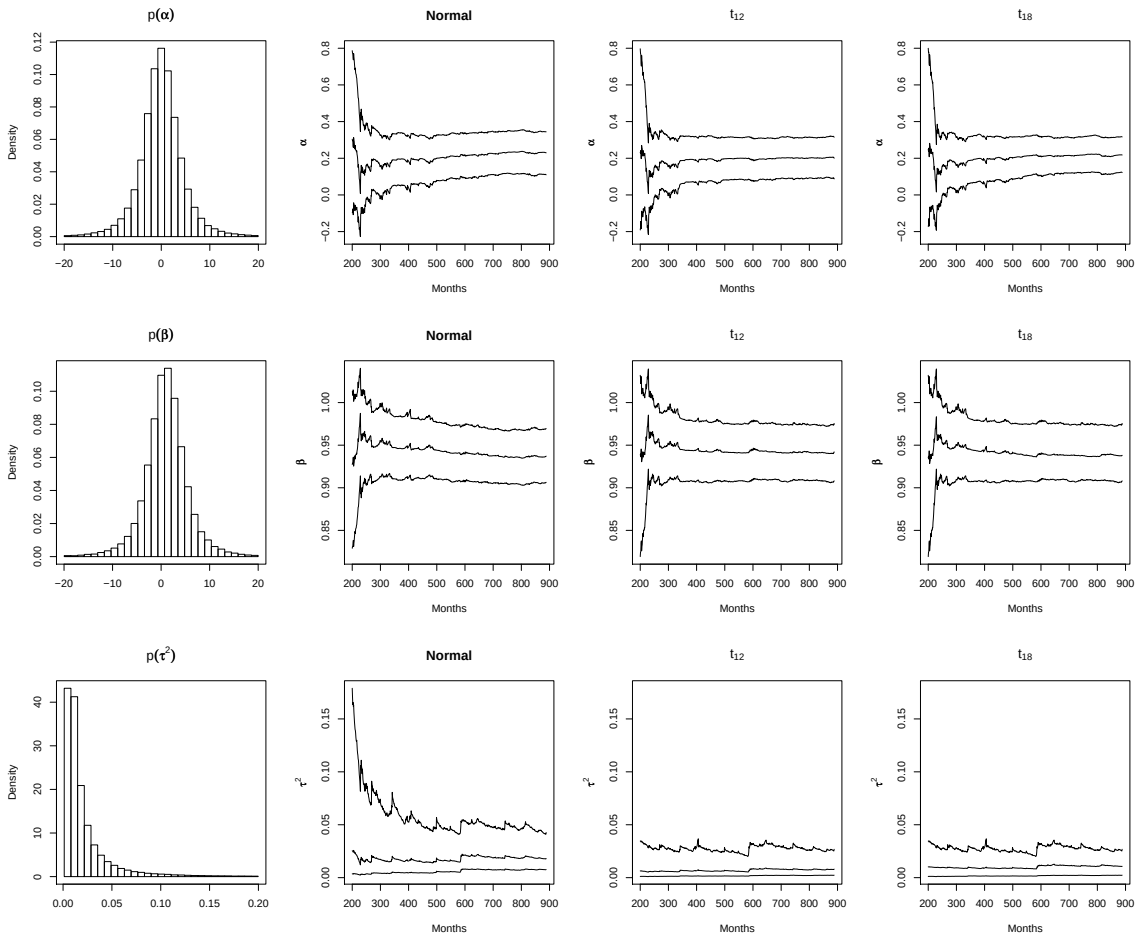

Figure 5: *Stochastic volatility model. 1st column:* Marginal prior distributions for α, β and τ^2. *2nd, 3rd and 4th columns:* Sequential 2.5th, 50th and 97.5th percentiles of $p(\gamma|y^t, M_1)$, for γ in $(\alpha, \beta, \tau^2, M)$ and model $M \in \{\text{Normal}, t_{12}, t_{18}\}$.

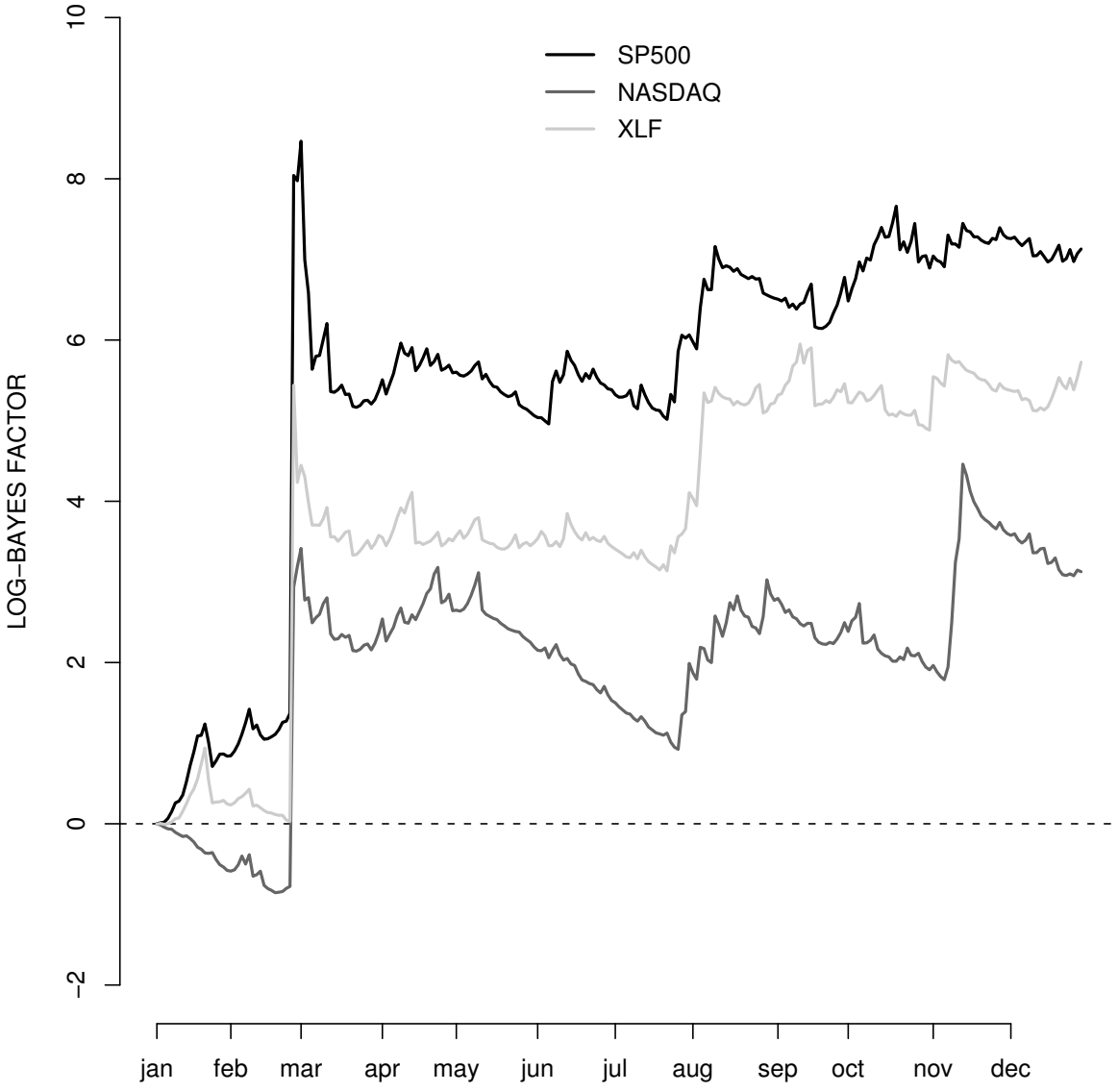

Figure 6: *Stochastic volatility with jumps.* Sequential (log) Bayes factor, $BF(M_1, M_0)$. $M_1 \equiv$SVJ model $M_0 \equiv$SV model. See Lopes and Polson (2010).

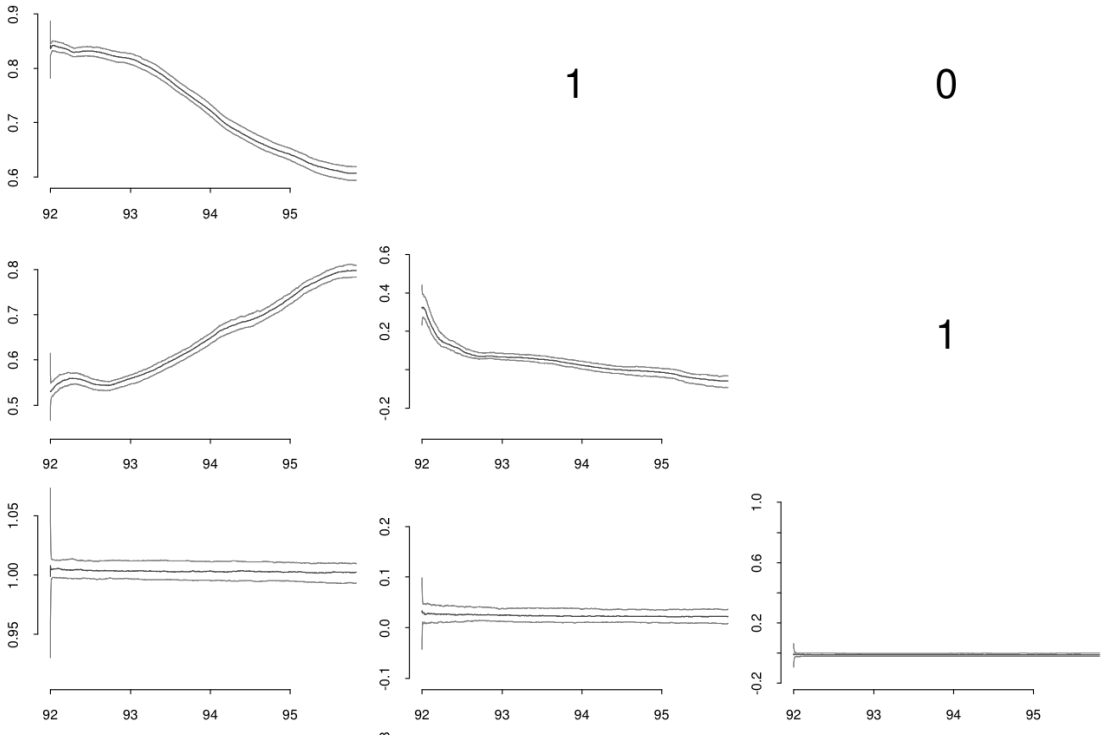

Figure 7: *Factor stochastic volatility with time-varying loadings.* Posterior 2.5th, 50th and 97.5th percentiles of the posterior distribution for the unconstrained elements of β_t, $p(\beta_t|y^n)$, for the first three time series for the period from January 1st 1992 to October 31st 1995. Top row: German Mark (DEM). Middle row: British Pound (GBP). Bottom row: Japanese Yen (JPY). See Lopes (2000) and Lopes, Aguilar and West (2002).

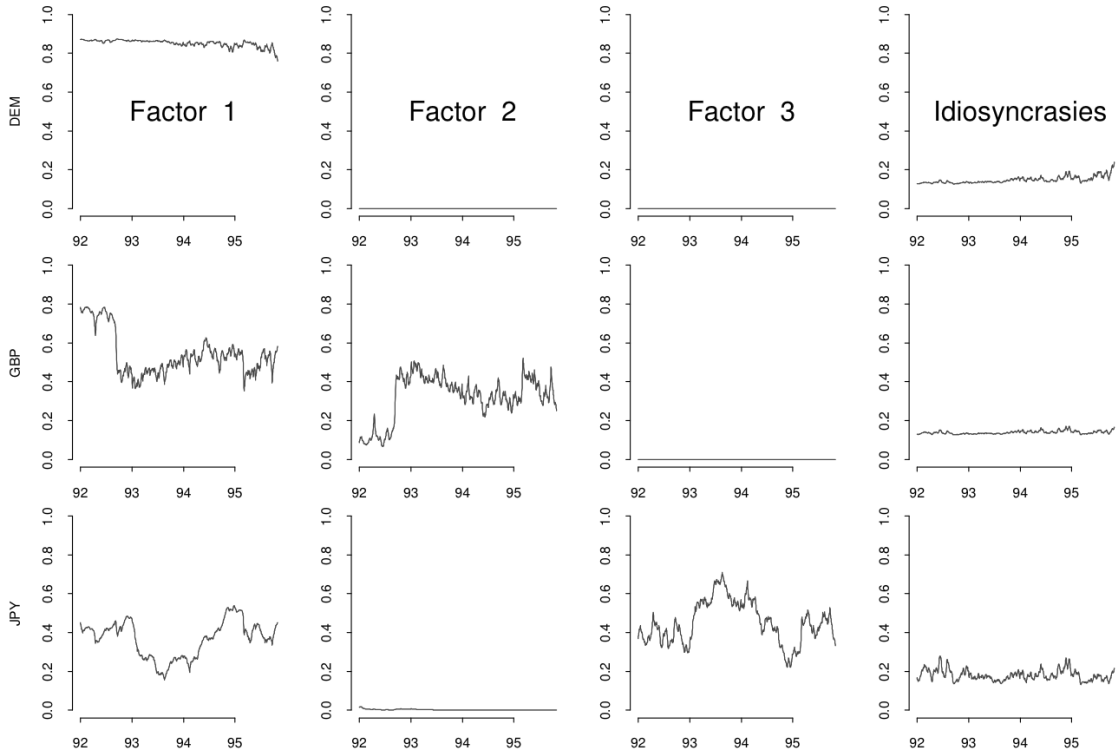

Figure 8: *Factor stochastic volatility with time-varying loadings.* Proportion of the variances of the first three time series explained by the three common factors and the idiosyncratic or specific factor for the period from January 1st 1992 to October 31st 1995. Top row: German Mark (DEM). Middle row: British Pound (GBP). Bottom row: Japanese Yen (JPY). See Lopes (2000) and Lopes, Aguilar and West (2002).