Parameter Estimation and Practical Aspects of Modeling Stochastic Volatility

Borus Jungbacker and Siem Jan Koopman

Abstract Estimating parameters in a stochastic volatility (SV) model is a challenging task and therefore much research is devoted in this area of estimation. This chapter presents an overview and a practical guide of the quasi-likelihood and the Monte Carlo likelihood methods of estimation. The concepts of the methods are straightforward and the implementation is based on Kalman filter, smoothing, simulation smoothing, mode calculation and Monte Carlo simulation. These methods are general, transparent and computationally fast; therefore, they provide a feasible way for the estimation of parameters in SV models. Various extensions of the SV model are considered and some details are provided for the effective implementation of the Monte Carlo methods. Some empirical illustrations are given to show that the methods can be successful in measuring the unobserved volatility in financial time series.

1 Introduction

Volatility models are concerned with the analysis of time-varying characteristics of the variance in financial return series. The daily closure log prices of an asset or an index evolve usually as a random walk or a process close to it; therefore, the relative price changes often behave as a white noise series. This empirical finding is consistent with economic theory. When markets operate efficiently, all current information is consolidated in the price. The current

Borus Jungbacker
Department of Econometrics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, e-mail: bjungbacker@feweb.vu.nl

Siem Jan Koopman
Department of Econometrics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, e-mail: s.j.koopman@feweb.vu.nl

T.G. Anderson et al., *Handbook of Financial Time Series*,
DOI: 10.1007/978-3-540-71297-8_13,

asset price is then also the best forecast for the future asset price. The dynamics in the mean of asset prices can therefore be treated as not existing. On the other hand, it is well established that the unobserved volatility in asset returns is subject to a dynamic process. For example, a well-known empirical feature in finance and economics is the temporary clustering of large shocks in a series of daily returns. The clustering of such shocks implies serial correlation in the squared returns.

Although various ad hoc treatments are available for analyzing time-varying features in volatility, a model-based approach has become the industry standard. The observation-driven class of models for volatility dynamics is the well-known generalized autoregressive conditional heteroscedasticity (GARCH) model as developed and popularized by Engle (1982) and Bollerslev (1986). Their applications in empirical finance literature are exhaustive and their extensions in the econometrics and statistics literature are widespread; see Zivot (2008). The parameter-driven counterpart of GARCH is the class of stochastic volatility (SV) models as formulated by Taylor (1986) and Harvey et al. (1994). In various empirical studies it has been shown that the SV model provides a basis for more accurate forecasts of volatility than those provided by GARCH models; see Koopman et al. (2005). Furthermore, SV models have a closer connection with financial economics theory. For example, in the option pricing literature, the asset price is usually modeled by a stochastic differential equation (SDE) such as

$$\begin{aligned} \mathrm{d}\log P(t) &= \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}B_1(t), \\ \mathrm{d}\log \sigma^2(t) &= \left\{\gamma + (\phi - 1)\log \sigma^2(t)\right\}\,\mathrm{d}t + \sigma_\eta\,\mathrm{d}B_2(t), \end{aligned} \tag{1}$$

where $P(t)$ is the asset price at time t, $\mu(t)$ is a drift term and $\sigma(t)$ is the volatility of the asset price at time t. The drift term $\mu(t)$ is likely to be small and is in practice often set to zero. The stochastic properties of the mean and variance of $P(t)$ are determined by the independent Brownian motions $B_i(t)$, for $i = 1, 2$, and the unknown fixed coefficients γ, ϕ and σ_η. The basic SV model is a discrete-time version of the SDE (1).

The estimation of parameters in discretized SV models is not standard since a closed-form expression for the likelihood function does not exist; therefore, different approaches have been considered for inference in SV models. Estimation can be based on approximations (quasi-maximum likelihood), numerical methods for evaluating the likelihood (numerical integration) or simulation methods. Much focus in the econometrics and statistics literature is on the use of Bayesian Markov chain Monte Carlo (MCMC) methods for inference, see, for example, Jacquier et al. (1994) and Kim et al. (1998). In this chapter we focus on the Monte Carlo method of evaluating the likelihood function of the SV model. In particular, we adopt the methods of Shephard and Pitt (1997) and Durbin and Koopman (1997) based on importance sampling. Although these methods are not applicable in all situations, they are computationally fast (in comparison with most other simulation methods)

and relatively easy to implement. So when importance sampling methods can be implemented successfully, they can be regarded as an effective estimation methodology for SV models. Other estimation methods that have generated interest are the method of moments developed by Andersen and Sorensen (1996), frequency-domain estimation considered by Breidt et al. (1998) and the likelihood approaches explored by Fridman and Harris (1998) and Brockwell (2007). A general overview of the SV literature is given by the collection of articles in the book of Shepard (2005).

Consider a time series of asset log-returns y_t that is assumed to have constant mean μ and a time-variance variance $\exp h_t$. The observations are typically sampled at daily intervals. The basic version of the discretized SV model for y_t is given by

$$\begin{aligned} y_t &= \mu + \exp(\tfrac{1}{2}h_t)\varepsilon_t, \;\; \varepsilon_t \sim \mathrm{NID}(0,1), \\ h_{t+1} &= \gamma + \phi h_t + \eta_t, \qquad \eta_t \sim \mathrm{NID}(0,\sigma_\eta^2), \end{aligned} \tag{2}$$

for $t = 1, \ldots, n$ and where ε_t and η_s are independent of each other at all time points $t, s = 1, \ldots, n$. This basic SV model is a nonlinear time series model since both h_t and ε_t in the multiplication $\exp(\frac{1}{2}h_t)\varepsilon_t$ are stochastic. It is usually assumed that the log-volatility process is stationary but persistent, that is, $0 < \phi < 1$ is typically larger than 0.8. The unconditional mean of the log-volatility process is $(1-\phi)^{-1}\gamma$ and can be interpreted as the long-term log variance of the asset return series y_t. The unconditional variance of the log-volatility process is $(1-\phi^2)^{-1}\sigma_\eta^2$ and is sometimes referred to as the "volatility of volatility." Furthermore, the stochastic time-varying variance of the log returns y_t conditional on h_t is given by $\sigma_t^2 = \mathrm{E}(y_t - \mu)^2 = \exp h_t$, for $t = 1, \ldots, n$, where h_t can be any stationary autoregressive process. The conditional log density $p(y_t|h_t)$ is then given by

$$\log p(y_t|h_t) = -\frac{1}{2}\log 2\pi - \frac{1}{2}h_t - \frac{1}{2}\exp(-h_t)(y_t-\mu)^2, \qquad t = 1, \ldots, n. \tag{3}$$

The unknown coefficients that need to be estimated are γ, ϕ and σ_η and are collected in the parameter vector ψ.

The SV model is a nonlinear and non-Gaussian time series model with an observation equation and a state equation. The observation equation describes the relationship between the observations and the latent factors, while the state equation provides a model description for the dynamic properties of the latent factors. Denote $y = (y_1, \ldots, y_n)'$, where y_t is the scalar observation at time t for $t = 1, \ldots, n$, and denote $\theta = (\theta_1, \ldots, \theta_n)'$, where θ_t is the so-called signal at time t that is only determined by the latent factors. For the SV model (2), we have simply $\theta_t = h_t$. Furthermore, the observation density is given by $p(y|\theta)$ and the signal density is given by $p(\theta)$. For the classes of SV models in this chapter, the conditional independence assumption applies to the density $p(y|\alpha)$, that is,

$$p(y|\theta) = \prod_{t=1}^{n} p(y_t|\theta_t).$$

The model densities $p(y|\theta)$ and $p(\theta)$ depend on a set of unknown coefficients. In the case of (2), these are γ, ϕ and σ_η. The estimation of these coefficients will be based on maximum likelihood. The likelihood function can be expressed as

$$p(y) = \int p(\theta, y)\,\mathrm{d}\theta = \int p(y|\theta)p(\theta)\,\mathrm{d}\theta. \tag{4}$$

which is an n-fold integral and is typically not tractable except in the most trivial cases. An analytical expression for $p(y)$ is therefore not available for the SV class of models. We need to rely on numerical techniques for the evaluation of $p(y)$. In this Chapter we consider the method of Monte Carlo integration based on importance sampling. A straightforward Monte Carlo estimator of $p(y)$ in (4) is $\widetilde{p}(y) = M^{-1}\sum_{m=1}^{M} p(y|\theta^m)$ where θ^m is a draw from $p(\theta)$ and with $\widetilde{p}(y) \to p(y)$ as $M \to \infty$. However, this Monte Carlo estimator is not efficient since many draws from $p(\theta)$ will make no contribution to $p(y|\theta)$ and hence the estimate will be poor even for extremely high values of M.

In this article we opt for two approaches that overcome the problems as described above. First we consider a linearization of the nonlinear observation equation. This will lead to a transformation of y_t that has a linear observation equation. Obviously, this approach is an approximation to the SV model. Second we consider the Monte Carlo evaluation of the likelihood function. The method of importance sampling is considered for the evaluation of (4). In both approaches, state-space methods for the linear Gaussian state-space model play a prominent role. We therefore consider a general state-space representation of the SV model although the implications of the results presented for the SV model will be given explicitly. The state-space algorithms are instrumental but do not need to be discussed in much detail since we only need to apply them. As a service to the reader, the algorithms are briefly discussed in the Appendix. Since the SV model is discussed within a more general setting, various extensions of the SV model can be considered as special cases too. The general method applies to each of them but there are some differences in detail which will be reported. To illustrate the effectiveness of the methods, some empirical illustrations are given for stock index return series and for exchange rate series.

2 A Quasi-Likelihood Analysis Based on Kalman Filter Methods

The basic SV model (2) is intrinsically a nonlinear model owing to the multiplication of two stochastic variables in the observation equation, that is,

$y_t - \mu = \exp(\frac{1}{2}h_t)\varepsilon_t$. Since the sample mean of the log returns y_t is a consistent estimator of μ, we can replace μ by the sample mean. Harvey et al. (1994) have pointed out that the basic SV model (2) can be analyzed on the basis of a linearized version of the model. For this purpose, we consider scalar x_t as the transformation of y_t and given by

$$x_t = \log(y_t - \bar{y})^2, \qquad \bar{y} = n^{-1}\sum_{t=1}^{n} y_t, \tag{5}$$

for $t = 1, \ldots, n$. Given the basic SV model for y_t, a reasonable suggestion of a model for x_t is given by

$$x_t = \kappa_1 + h_t + u_t, \qquad h_{t+1} = \gamma + \phi h_t + \eta_t, \tag{6}$$

where $u_t = \log \varepsilon_t^2 - \kappa_1$ is distributed by the centered $\log \chi^2$ density with one degree of freedom. The mean and variance of $\log \varepsilon_t^2$ are given by κ_1 and κ_2, respectively. In this case, $\kappa_1 \approx -1.27$ and $\kappa_2 = \pi^2 \,/\, 2$. Model (6) is linear and the observation disturbance has a non-Gaussian density.

However, we may consider u_t also as a sequence of independent noise terms with mean zero and variance κ_2 to avoid the need to treat the density function of u_t explicitly. Linear methods can then be applied to obtain estimators of h_t that belong to the class of minimum mean squares linear estimators. The metric for estimation can nevertheless be chosen as the Gaussian likelihood function. Such an approach will be referred to as a quasi-maximum likelihood analysis. It effectively considers model (6) with Gaussian disturbance terms for the transformed series x_t, that is,

$$\begin{aligned} x_t &= \kappa_1 + h_t + u_t, \qquad h_{t+1} \;=\; \gamma + \phi h_t + \eta_t, \\ u_t &\sim \mathrm{NID}(0, \kappa_2), \qquad \eta_t \;\sim\; \mathrm{NID}(0, \sigma_\eta^2), \end{aligned} \tag{7}$$

for $t = 1, \ldots, n$. This linearized SV model is an example of the linear Gaussian state space model with the state equation for h_t and the observation equation for x_t. Although the SV model usually has the log volatility h_t modeled as an autoregressive (AR) process (mean-reverting), the state-space framework allows it to be modeled by many other linear Gaussian time series processes.

In the state-space formulation, we define θ_t as the signal and α_t as the state vector. A general observation equation for x_t can be given by

$$x_t = \theta_t + u_t, \qquad u_t \sim \mathrm{NID}(0, H_t), \tag{8}$$

for $t = 1, \ldots, n$. The dynamic properties of the signal are modeled by

$$\theta_t = c_t + Z_t\alpha_t, \qquad \alpha_{t+1} = d_t + T_t\alpha_t + \eta_t, \qquad \eta_t \sim \mathrm{NID}(0, Q_t), \tag{9}$$

for $t = 1, \ldots, n$, where system vectors c_t and d_t and system matrices Z_t, H_t, T_t and Q_t are fixed and known functions of parameter vector ψ. The

observation x_t and signal θ_t are scalar variables, while the state vector α_t together with the disturbance vector η_t have dimensions $q \times 1$. It follows that model (7) is represented by $x_t = \theta_t + u_t$, where $\theta_t = h_t + \kappa_1$ is modeled by (9) with $c_t = \kappa_1$, $Z_t = 1$, $H_t = \kappa_2$, $\alpha_t = h_t$, $d_t = \gamma$, $T_t = \phi$ and $Q_t = \sigma_\eta^2$. The linearized SV model is therefore a time-invariant case of the general linear Gaussian state space model (8) and (9). The Kalman filter and related methods can be applied to this state-space model; see the Appendix.

Stationary autoregressive moving averages processes as well as nonstationary linear processes for the signal θ_t can be formulated as (9). The system vectors and matrices in (9) have appropriate dimensions and the variance matrix Q_t is positive-semidefinite. The initial state vector is normally distributed with mean a and variance matrix P. The disturbances η_t are serially independent and are independent of the initial state vector for $t = 1, \ldots, n$. The joint property of a sequence of n state vectors can be expressed by the multivariate normal density

$$\alpha \sim \mathrm{N}(d, \Omega), \tag{10}$$

where

$$\alpha = (\alpha_1', \ldots, \alpha_n')', d = T\left(a', d_1', \ldots, d_{n-1}'\right)', \Omega = T\mathrm{diag}(P_1, Q_1, \ldots, Q_{n-1})T',$$

with

$$T = \begin{bmatrix} I & 0 & 0 & \cdots & 0 & 0 \\ T_1 & I & 0 & \ldots & 0 & 0 \\ T_2T_1 & T_2 & I & & 0 & 0 \\ & & & \ddots & & \vdots \\ T_{n-2}\ldots T_1 & T_{n-2}\ldots T_2 & T_{n-2}\ldots T_3 & & I & 0 \\ T_{n-1}\ldots T_1 & T_{n-1}\ldots T_2 & T_{n-1}\ldots T_3 & \cdots & T_{n-1} & I \end{bmatrix}, \tag{11}$$

for $t = 1, \ldots, n$.

It further follows that $\theta = (\theta_1, \ldots, \theta_n)'$ has a multivariate normal distribution given by

$$\theta \sim \mathrm{N}(\mu, \Psi), \qquad \mu = c + Zd, \qquad \Psi = Z\Omega Z', \tag{12}$$

where

$$\theta = c + Z\alpha, \qquad c = (c_1', \ldots, c_n')', \qquad Z = \mathrm{diag}(Z_1, \ldots, Z_n).$$

The log density of the signal is given by

$$\log p(\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Psi| - \frac{1}{2}(\theta - \mu)'\Psi^{-1}(\theta - \mu). \tag{13}$$

The prediction error decomposition can be used to evaluate (13). Most of the commonly used linear Gaussian time series models, including autoregres-

sive moving average models, can be represented in state-space form with a positive-definite variance matrix Ψ.

Define $n \times 1$ vector $x = (x_1, \ldots, x_n)'$, then

$$x = \theta + u, \qquad u \sim \mathrm{N}(0, H), \tag{14}$$

with $u = (u_1, \ldots, u_n)'$ and $H = \mathrm{diag}(H_1, \ldots, H_n)$. In case of (7), we have $H_t = \kappa_2$ for $t = 1, \ldots, n$ and $H = \kappa_2 I_n$. The linear Gaussian observation density is given by

$$p(x|\theta) = \mathrm{N}(\theta, H) = \prod_{t=1}^{n} p(x_t|\theta_t). \tag{15}$$

We have shown that the linearized SV model can be represented by the general linear Gaussian state space model with signal density $p(\theta)$ and observation density $p(x|\theta)$.

2.1 Kalman filter for prediction and likelihood evaluation

The linear Gaussian state-space model as formulated in (8) and (9) can be analyzed using computationally efficient and fast recursive algorithms. These methods can be applied to any linear time series model in which the dynamics can be represented in the Markovian form. The computations are carried out by so-called order-n operations, where n is the number of observations. The main attractions of the state-space approach are its generality, its effective treatment of missing observations, its handling of other messy features in time series analysis and its natural way of carrying out one-step-ahead prediction and long-term forecasting.

The Kalman filter is given in the Appendix. It evaluates the estimator of the state vector α_t conditional on the past observations $X_{t-1} = \{x_1, \ldots, x_{t-1}\}$ together with its conditional variance as given by

$$a_t = E(\alpha_t|X_{t-1}), \qquad P_t = \mathrm{Var}(\alpha_t|X_{t-1}) = E[(\alpha_t - a_t|X_{t-1})(\alpha_t - a_t|X_{t-1})'],$$

respectively, for $t = 1, \ldots, n$. It follows that the prediction of the signal is given by $c_t + Z_t a_t$ with mean square error $Z_t P_t Z_t'$. Further, the observation prediction error and its conditional variance are given by

$$v_t = x_t - E(x_t|X_{t-1}) = x_t - c_t - Z_t a_t, \qquad F_t = E(v_t v_t'|X_{t-1}) = Z_t P_t Z_t' + H_t,$$

respectively, for $t = 1, \ldots, n$. The prediction error decomposition allows the log-likelihood function $\ell(\psi)$ to be evaluated analytically as

$$\ell(\psi) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{n}\log F_t - \frac{1}{2}\sum_{t=1}^{n} v_t' F_t^{-1} v_t,$$

where ψ is the parameter vector containing the coefficients γ, ϕ and σ_η. By maximizing the log likelihood $\ell(\psi)$ with respect to ψ, we obtain estimates for the unknown coefficients. Analytical expressions for the estimates at the maximum of $\ell(\psi)$ are not available and therefore we rely on numerical optimization methods such as quasi-Newton methods; see Nocedal and Wright (1999). In the case where x is modeled by the Gaussian density $p(x|\theta)$ as in (14), the estimate of ψ is referred to as the maximum likelihood estimate. However, in the case of the linearized SV model (7) for x_t, the estimate of ψ, obtained by maximizing the Gaussian likelihood function, is referred to as the quasi-maximum likelihood estimate since the assumption of a $\log \chi^2$ density for u_t is replaced by the assumption of a Gaussian density.

Interest in the analysis of log returns often focuses on the forecasting of volatility for the purpose of, for example, determining option prices in the short-term future. Volatility forecasting is based on the prediction of the future signal conditional on the observed time series $X_n = \{x_1, \ldots, x_n\}$, together with its conditional variance, as defined by

$$\widehat{\theta}_{n+j|n} = E(\theta_{n+j}|X_n), \qquad V_{n+j|n} = Var(\theta_{n+j}|X_n), \qquad j = 1, 2, 3 \ldots.$$

The forecasts of the signal from a general state-space model can also be obtained by the Kalman filter. This is due to the ability of the Kalman filter to deal with missing observations whether they occur within sample or out of sample; see the Appendix.

2.2 Smoothing methods for the conditional mean, variance and mode

The actual measurement of the unobserved volatility given all observed log returns is referred to as smoothing or signal extraction in the state-space terminology. More specifically, we aim to evaluate the estimated log volatility as defined by the conditional mean of θ conditional on the observed log returns x (or X_n). These smoothed estimators can be obtained within the general framework of the state-space model as follows. Given parameter estimates of ψ, the unconditional mean, variance and covariance of observation x and signal θ are given by

$$E(x) = \mu, \qquad \mathrm{Var}(x) = \Sigma = \Psi + H, \qquad \mathrm{Cov}(\theta, x) = \Psi. \tag{16}$$

It follows from the standard lemma of the multivariate normal density that the conditional means and variances are given by

$$E(\theta|x) = \widehat{\theta} = \mu + \Psi\Sigma^{-1}(x-\mu), \qquad \mathrm{Var}(\theta|x) = V = \Psi - \Psi\Sigma^{-1}\Psi. \quad (17)$$

The Kalman filter and smoother evaluate the conditional mean $E(\theta_t|x)$ and variance $\mathrm{Var}(\theta_t|x)$ in a recursive and computationally efficient way for a linear Gaussian state-space model, see the Appendix. More specifically, the smoothing algorithm is a backwards-operating recursion that accommodates output of the Kalman filter and it evaluates the quantities

$$\widehat{\theta}_t = E(\theta_t|x), \qquad V_t = \mathrm{Var}(\theta_t|x), \qquad t = 1, \ldots, n.$$

The evaluation of V_t enables the construction of confidence intervals for $\widehat{\theta}_t$ for $t = 1, \ldots, n$.

Since all densities are Gaussian, the conditional or posterior mode of $p(\theta|x)$, denoted by $\widetilde{\theta}$, is equivalent to the conditional mean of $p(\theta|x)$, that is, $\widetilde{\theta} = E(\theta|x)$. After some minor manipulations, it follows from the first equation in (17) that

$$\widetilde{\theta} = \left(\Psi^{-1} + H^{-1}\right)^{-1}\left(H^{-1}x + \Psi^{-1}\mu\right). \quad (18)$$

It should be emphasized that the Kalman filter and smoother effectively computes $\widetilde{\theta}$ for the linear Gaussian state-space model.

2.3 Practical considerations for analyzing the linearized SV model

The linearized SV model (7) provides a full statistical analysis of log returns by using standard state-space methods. The embedding of the model within the general state-space framework also allows the treatment of messy features in the time series such as missing values, irregularly spaced data, outliers and breaks; see Harvey et al. (1998). Furthermore, the methods are computationally efficient and fast. The necessary computations can be carried out by software packages such as the user-friendly STAMP program of Koopman et al. (2007) and the `SsfPack` library for `Ox` and `S-PLUS`; see Koopman et al. (1999) and Zivot et al. (2004).

However, there is the practical inconvenience that the log returns are transformed by taking logs of the squares (mean-adjusted) as in (5). In the case where the return is very small or even zero (whether or not it is mean-adjusted), this transformation can clearly lead to numerical problems. The problem is often referred to as the inlier problem and has been investigated by Breidt and Carriquiry (1996). They found that a transformation (based on a Taylor series) suggested by Fuller (1996) has proved to be more stable. This transformation is given by

$$x_t = \log(y_t^2 + s) - \frac{s}{y_t^2 + s}, \qquad t = 1, \ldots, n,$$

where s is the sample variance of y_t scaled by a small multiple, say, 0.02. In practical work, this transformation of x_t is preferred over (5). In the approaches described below, we do not require a transformation since the methods work with the untransformed process; see also the discussion in Davis and Rodriguez-Yam (2005).

The mean κ_1 and variance κ_2 of $\log \varepsilon_t^2$ are considered as given in the quasi-maximum likelihood procedure. These moments have known values and do not need to be estimated. However, from Monte Carlo simulation studies it has emerged that the small-sample properties for the estimates of ψ improve when κ_1 and κ_2 are assumed unknown and are estimated together with the coefficients in ψ.

Despite the effectiveness of the linearized SV model, it is unsatisfactory in that the underlying assumptions of the SV model are not properly treated by the approximation. Therefore, other methods have been developed to analyze the SV model without linearization. The remaining part of this contribution presents and discusses methods for the treatment of the SV model based on the nonlinear formulation (2).

3 A Monte Carlo Likelihood Analysis

Given the nonlinear property of the SV model (2), the likelihood function of the SV model does not have a convenient and analytical expression in terms of parameter vector ψ. The likelihood function can be expressed generally by

$$\ell(\psi) = p(y; \psi) = \int_{-\infty}^{\infty} p(y, \theta; \psi)\, \mathrm{d}\theta, \tag{19}$$

where y and θ are vectors formed by stacking the observations and states y_t and θ_t, respectively, as defined earlier. Given the potentially high dimensional vectors y and θ, direct numerical evaluation of the integral is not feasible. We therefore consider Monte Carlo methods for the evaluation of likelihood function (19). In particular, the method of importance sampling is explored in detail. For this purpose we adopt the trivial identity

$$p(y; \psi) = \int p(y, \theta; \psi)\, \mathrm{d}\theta = \int \frac{p(y, \theta; \psi)}{f(\theta; y, \psi)} f(\theta; y, \psi)\, \mathrm{d}\theta = \mathrm{E}_f \left(\frac{p(y, \theta; \psi)}{f(\theta; y, \psi)} \right), \tag{20}$$

where $\mathrm{E}_f(\cdot)$ denotes expectation with respect to some proposal density $f(\theta; y, \psi)$ that is chosen to be as close as possible to $p(\theta|y; \psi)$ but has convenient properties. The expectation $\mathrm{E}_f(\cdot)$ can be evaluated by sampling from $f(\theta; y, \psi)$ and averaging the importance weights $p(y, \theta; \psi)$ / $f(\theta; y, \psi)$.

In the case where we choose $f(\theta; y, \psi) = p(\theta|y; \psi)$, expression (20) reduces obviously to $p(y; \psi)$, which cannot be evaluated analytically. Furthermore, exact sampling from $p(\theta|y; \psi)$ is usually not feasible in the context of non-Gaussian and nonlinear state-space models such as the SV model (2). For the importance sampling evaluation of (20), we need to find a proposal density $f(\theta; y, \psi)$ that is sufficiently close to $p(\theta|y; \psi)$ but from which it is relatively easy to simulate. In most practical applications of importance sampling, the proposal density is found within the class of linear Gaussian densities. In the context of state-space models, the proposal $f(\theta; y, \psi)$ is usually set equal to the multivariate normal $\mathrm{N}(\widetilde{\theta}, V)$, where $\widetilde{\theta}$ is some location measure of θ and V is an appropriate variance matrix. The notation deliberately suggests that location and scale are related to (18) and the second equation of (17), respectively, as will become apparent later. Sampling from this multivariate normal density can be carried out by the simulation smoother; see the Appendix. This method is fast and straightforward and it provides the means for the evaluation of the likelihood (19) by importance sampling.

The remainder of this section discusses the construction of an effective proposal density, sampling from this density and the practical details of evaluating the likelihood function (20) by importance sampling. Other aspects of analyzing an SV model are also discussed, including the measurement and forecasting of volatility.

3.1 Construction of a proposal density

We adopt a multivariate normal density as the proposal density with its mean equal to the mode of the smoothing density $p(\theta|y)$ and its curvature equal to that of $p(\theta|y)$ at the mode. This choice of the proposal density $f(\theta; y)$ is made since it may be sufficiently close to the smoothing density $p(\theta|y)$. It requires us to find the mode $\widehat{\theta}$, by maximizing the smoothing density $p(\theta|y)$ with respect to θ, and its Hessian matrix G, that is $f(\theta; y) = \mathrm{N}(\widehat{\theta}, V)$, where $V = -G^{-1}$. For the basic SV model (2) we consider the general nonlinear non-Gaussian observation model $p(y|\theta)$ and the linear Gaussian signal vector $p(\theta)$. For this class of models, an analytical expression for the posterior mode $\widehat{\theta}$ of $p(\theta|y)$ is not available. We therefore obtain the mode by maximizing $p(\theta|y)$ with respect to θ using the Newton–Raphson method of optimization; see Nocedal and Wright (1999) for a treatment of numerical optimization methods. The dimension of θ is typically $n \times 1$, so matrix dimensions can be high and straightforward matrix manipulations become infeasible; therefore efficient algorithms need to be considered.

For a given guess g of the solution for $\widehat{\theta}$, the Newton–Raphson method proposes the new guess as

$$g^+ = g - \left[\ddot{p}(\theta|y)|_{\theta=g}\right]^{-1} \dot{p}(\theta|y)|_{\theta=g}\,, \tag{21}$$

where we define

$$\dot{p}(\cdot|\cdot) = \frac{\partial \log p(\cdot|\cdot)}{\partial \theta}, \qquad \ddot{p}(\cdot|\cdot) = \frac{\partial^2 \log p(\cdot|\cdot)}{\partial \theta \partial \theta'}. \tag{22}$$

Since $\log p(\theta|y) = \log p(y|\theta) + \log p(\theta) - \log p(y)$, we have

$$\dot{p}(\theta|y) = \dot{p}(y|\theta) - \Psi^{-1}(\theta - \mu), \qquad \ddot{p}(\theta|y) = \ddot{p}(y|\theta) - \Psi^{-1}. \tag{23}$$

The conditional independence assumption of the observation model implies that $\ddot{p}(y|\theta)$ is a block diagonal matrix.

The Newton–Raphson updating step reduces to

$$\begin{aligned} g^+ &= g - \left[\ddot{p}(y|\theta)|_{\theta=g} - \Psi^{-1}\right]^{-1} \left(\dot{p}(y|\theta)|_{\theta=g} - \Psi^{-1}\{g - \mu\}\right) \\ &= \left[\Psi^{-1} - \ddot{p}(y|\theta)|_{\theta=g}\right]^{-1} \left(\dot{p}(y|\theta)|_{\theta=g} - \ddot{p}(y|\theta)|_{\theta=g}\, g + \Psi^{-1}\mu\right) \\ &= \left(\Psi^{-1} + A^{-1}\right)^{-1} \left(A^{-1}x + \Psi^{-1}\mu\right), \end{aligned} \tag{24}$$

where

$$A = -\left[\ddot{p}(y|\theta)|_{\theta=g}\right]^{-1}, \qquad x = g + A\,\dot{p}(y|\theta)|_{\theta=g}\,. \tag{25}$$

We note the similarity of (24) and (18). In the case where $\ddot{p}(y|\theta)$ is negative-semidefinite for all θ, it follows that the Kalman filter and smoother can be used to compute g^+ by applying it to a state-space model with observation equation (14) for x_t as in (25) and $H = A$. The computation of $E(\theta|x)$ for this model returns g^+ as a result. This approach was taken by Shephard and Pitt (1997), Durbin and Koopman (1997) and So (2003). The mode $\widehat{\theta}$ for a non-Gaussian nonlinear observation model is obtained by the Newton–Raphson method where for each step the Kalman filter and smoother computes the new guess g^+. The Hessian matrix of the mode estimator $\widehat{\theta}$ is given by $G = \ddot{p}(\theta|x) = -\Psi^{-1} - A^{-1}$.

This approach of finding the mode $\widehat{\theta}$ is clearly not valid when $p(x|\theta)$ is not log-concave, so $\ddot{p}(x|\theta)$ is not negative-definite. This implies that the variance matrix H for the observation model (14) is not positive-definite. However, it was argued by Jungbacker and Koopman (2007) that in cases where $\ddot{p}(y|\theta)$ is not negative-definite, the Kalman filter and smoothing recursions can still be used for the computation of (24). The special structure of variance matrix $\Psi = Z\Omega Z'$ of the signal θ nevertheless allows the use of decompositions based on triangular matrices such as T in (11). Hence, it can be shown that the Kalman and smoothing recursions can also be used for the general model.

Although matrix $\Psi + A$ can be indefinite, the Hessian matrix $-\Psi^{-1} - A^{-1}$ should always be seminegative-definite for θ at or in the close neighborhood of $\widehat{\theta}$ by construction. In cases where the Hessian matrix is not negative-definite,

the Newton–Raphson step does not progresses to the maximum of $p(\theta|y)$ with respect to θ. To enforce global convergence, the algorithm can be modified by line-search and other numerical methods; see Nocedal and Wright (1999). In general, line-search strategies often speed up the maximization and stabilize the algorithm. A line search can be implemented by introducing the scalar $0 < \lambda \leq 1$ in (21) and defining

$$g_\lambda^+ = g - \lambda \left[\ddot{p}(\theta|y)|_{\theta=g} \right]^{-1} \dot{p}(\theta|y)|_{\theta=g} \,. \tag{26}$$

The line search consists of finding a value for λ so that

$$p(\theta|y)|_{\theta=g_\lambda^+} > p(\theta|y)|_{\theta=g} \,.$$

By combining (26) and (21), the line search computations are straightforward and are given by

$$g_\lambda^+ = g + \lambda(g^+ - g),$$

where $g^+ = g_\lambda^+\big|_{\lambda=1}$ is computed by (24) only once for different values of $0 < \lambda \leq 1$. Global convergence is ensured when an appropriate set of regularity conditions for the line search is fulfilled; see Nocedal and Wright (1999) for a detailed discussion. To check these conditions, it is usually necessary to evaluate the score function.

The score vector of $p(\theta|y)$ is defined in (22) and is given by (23), that is,

$$\frac{\partial \log p(\theta|y)}{\partial \theta} = \dot{p}(\theta|y) = \dot{p}(y|\theta) - \Psi^{-1}(\theta - \mu).$$

The score can also be evaluated by the Kalman filter and smoother algorithms; see Jungbacker and Koopman (2007). Given the computational device for evaluating the score, other maximization methods may also be considered to obtain the mode of $p(\theta|y)$. It is noted that different numerical problems can occur during the maximization of $p(\theta|y)$ with respect to the high-dimensional vector θ. Although line-search methods can stabilize the Newton–Raphson method, it may be necessary to switch to other score-based or quasi-Newton optimization methods. Therefore, this computationally efficient method of computing the score is important in practical work.

3.2 Sampling from the importance density and Monte Carlo likelihood

The likelihood function $\ell(\psi) = p(y) = \int p(\theta, y)\, \mathrm{d}\theta$ is estimated via importance sampling using the expression

$$\widehat{\ell}(\psi) = M^{-1} \sum_{i=1}^{M} \frac{p(\theta^i, y)}{f(\theta^i; y)}, \qquad \theta^i \sim f(\theta^i; y). \tag{27}$$

The computation of (27) requires algorithms to sample from the importance density $f(\theta^i; y)$ and to evaluate for $i = 1, \ldots, M$ the so-called importance weight $p(\theta^i, y) \,/\, f(\theta^i; y)$. In cases where the observation density $p(y|\theta)$ is log-concave, the importance density can be represented as a linear Gaussian state space model and the simulation smoothers of de Jong and Shephard (1995) and Durbin and Koopman (2002) can be used to simulate from $f(\theta; y)$ in a computationally efficient way. The derivations of these methods rely on a properly defined linear Gaussian observation equation with positive definite matrices for Σ and $H = A$. In the general case, in particular when $p(y|\theta)$ is not log-concave, simulations can be carried out by the modified simulation smoother proposed by Jungbacker and Koopman (2007), see Appendix. We note that the Hessian of $p(\theta|y)$ is evaluated at $\theta = \widehat{\theta}$ and therefore matrix G is guaranteed to be negative definite and V is positive definite as a result.

The computation of (27) further requires evaluating the importance weight $p(\theta^i, y) \,/\, f(\theta^i; y)$ for $i = 1, \ldots, M$. Given the linear Gaussian signal vector θ, the evaluation of the nominator is based on the identity $p(\theta, y) = p(y|\theta)p(\theta)$, where $p(y|\theta)$ is defined by the model and is usually straightforward to compute. The density of the signal $p(\theta)$ for $\theta = \theta^i$ is evaluated by the Kalman filter since $\theta = c + Z\alpha$ has the Markov property and the prediction error decomposition can be applied to $p(\theta)$. Given the draw $\theta^i \sim f(\theta; y)$ with $f(\theta; y) = \mathrm{N}(\widehat{\theta}, V)$ obtained from the simulation smoother, the denominator $f(\theta^i; y)$ in (27) can be evaluated using the output of the simulation smoother; see the Appendix.

The estimator (27) is subject to Monte Carlo error. A strong law of large numbers insists that $\widehat{\ell}(\psi) \to \ell(\psi)$ as $M \to \infty$ with a rate of convergence depending on the precision of the proposal density; see Geweke (1989). The choice of M can be relatively small when an accurate proposal density is chosen. To identify appropriate proposal densities we may rely on a set of tests and diagnostics; see Koopman et al. (2007). In practical work, it is often sufficient to take M equal to 100 or 500.

For the purpose of parameter estimation, we maximize the Monte Carlo estimator of the likelihood function $\widehat{\ell}(\psi)$ in (27) with respect to the unknown vector ψ. The Newton–Raphson method can be used to maximize the likelihood function directly; see also the discussion in Section 2. During the numerical search, the same random seed is used each time the Monte Carlo likelihood function (27) is computed for a different ψ. This ensures that $\widehat{\ell}(\psi)$ is continuous in ψ.

4 Some Generalizations of SV Models

The basic SV model (2) was the motivating example for the discussion of estimation methods in the previous section. However, these methods can also be adopted for a wider set of SV models. Some generalizations of the SV model are presented in this section together with the details of implementing the estimation methods.

We consider the observed time series of asset log returns y_t which are typically sampled at daily intervals but not necessarily. A general SV model is given by

$$y_t = \mu_t + \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim IID(0,1), \qquad t = 1, \ldots, n, \tag{28}$$

where σ_t is the unobserved volatility process and μ_t is the possibly time varying mean of log returns. In this chapter we assume

$$\sigma_t = \sigma \exp\left(\frac{1}{2}\theta_t\right),$$

where θ_t is an unobserved linear Gaussian process and σ^2 is a fixed unknown parameter.

4.1 Basic SV model

In the basic SV model (2) the log-volatility process θ_t is assumed to be the Gaussian AR process as given by $\theta_t = h_t$, where

$$h_{t+1} = \phi_1 h_t + \ldots + \phi_p h_{t+1-p} + \xi_t, \qquad \xi_t \sim \mathrm{NID}(0, \sigma_\xi^2), \qquad t = 1, \ldots, n, \tag{29}$$

and with $p = 1$. The AR(1) process for h_t is given by (29) with $p = 1$ and is assumed to be independent of the Gaussian innovation sequence for ε_t. The mean is fixed, that is, $\mu_t = \mu$. Alternative specifications for μ_t and θ_t are discussed below.

The quasi-maximum likelihood analysis of the basic SV model is discussed in Section 2. It is made clear that such an analysis can be fully based on standard state-space methods applied to a linear Gaussian model. The transformation $y_t^* = \log(y_t - \bar{y})^2$ as in (5), for $t = 1, \ldots, n$, leads to the linear model (6) for $x_t = y_t^*$, where the disturbance $u_t + \kappa_1 = \log \varepsilon_t^2$ is $\log \chi^2$-distributed. The quasi-likelihood approach then replaces the $\log \chi^2$ by a Gaussian density for u_t with mean and variance equal to those of the $\log \chi^2$ density. However, Sandmann and Koopman (1998) considered model (6) as a linear model with the non-Gaussian $\log \chi^2$ density given by

$$\log p(y_t^*|\theta_t) = -\frac{1}{2}\log 2\pi + \frac{1}{2}\left(z_t - \exp z_t\right), \qquad z_t = y_t^* - \theta_t, \qquad t = 1, \ldots, n,$$

where $\theta_t = h_t$. We can adopt the Monte Carlo likelihood analysis of Section 2. The methods presented are fairly easy to implement in this case. The posterior mode calculations require expressions for

$$\dot{p}(y_t^*|\theta_t) = \frac{\partial}{\partial \theta_t}\log p(y_t^*|\theta_t), \quad \ddot{p}(y_t^*|\theta_t) = \frac{\partial^2}{\partial \theta_t \partial \theta_t'}\log p(y_t^*|\theta_t), \quad t = 1, \ldots, n.$$

In the case of the $\log \chi^2$ density $p(y_t^*|\theta_t)$, we have

$$\dot{p}(y_t^*|\theta_t) = \frac{1}{2}\left(\exp z_t - 1\right), \qquad \ddot{p}(y_t^*|\theta_t) = -\frac{1}{2}\exp z_t, \qquad t = 1, \ldots, n.$$

It follows that we can obtain the posterior mode via the Newton–Raphson updating steps (24) with the tth element of x given by $x_t = g_t + 1 - \exp(-z_t)$ and the tth diagonal element of A given by $A_t = 2\exp(-z_t)$. Once the posterior mode $\widetilde{\theta}$ has been obtained, simulation from the multivariate normal density $\mathrm{N}(\widetilde{\theta}, V)$ can take place via the simulation smoother and the Monte Carlo likelihood estimator (27) can be computed as a result. Further, it can be numerically maximized with respect to parameter vector ψ.

However, it is preferred to apply the methods of Section 3 directly on the SV model without transforming the log returns; see Shephard and Pitt (1997). In this case we need to treat the nonlinear observation model with a Gaussian density for ε_t. The conditional density $p(y_t|\theta_t)$ of the basic SV model is given by (3) with $h_t = \theta_t$. The posterior mode can be obtained as described in the previous paragraph but the derivatives of the model density are different and are given by

$$\dot{p}(y_t|\theta_t) = \frac{1}{2}\left\{\exp(-\theta_t)(y_t - \mu)^2 - 1\right\}, \qquad \ddot{p}(y_t|\theta_t) = -\frac{1}{2}\exp(-\theta_t)(y_t - \mu)^2,$$

for $t = 1, \ldots, n$. The methods for a Monte Carlo likelihood analysis can be implemented in a way similar to what has just been described.

4.2 Multiple volatility factors

In empirical work it has been observed that volatility often exhibits long-range dependence; see Andersen et al. (2003). Ideally, log volatility θ_t is modeled by a fractionally integrated process; see Granger and Joyeau (1980). Inference for the SV model (28) with a long-memory process for θ_t is often based on the spectral likelihood function; see Breidt et al. (1998) and Ray and Tsay (2000). Exact maximum likelihood methods were recently considered by Brockwell (2007). In our framework, we can approximate the long-range

dependence in the log volatility θ_t by considering it as a sum of independent autoregressive factors, that is,

$$\theta_t = \sum_{i=1}^{q} h_{it},$$

where each h_{it} represents the autoregressive process (29). The most commonly used specification is the two-factor model ($q = 2$), where one factor models the long-run dependence and the other the short-run dependence. The details of a Monte Carlo likelihood analysis are not different from those for a basic SV model since $p(y_t|\theta_t)$ remain unaltered.

4.3 Regression and fixed effects

It is often desirable to include regression effects in the specification of the volatility. Tsiakas (2006) introduced dummy effects to account for a seasonal pattern in the volatility of his periodic SV model. Koopman et al. (2005) considered a regression variable that contains information on the unobserved log-volatility process. Such regression effects can be incorporated into the SV model by extending the log-volatility signal specification by

$$\theta_t = W_t^{\theta}\beta + h_t,$$

where h_t is the autoregressive process (29), W_t^{θ} is a $1 \times k^{\theta}$ vector of covariates and β is a $k^{\theta} \times 1$ vector of regression coefficients.

The estimation of regression effects in the volatility process can be carried out in two ways. First, the coefficients in β can be treated as unknown parameters and incorporated in the parameter vector ψ. The Monte Carlo likelihood function (27) is maximized with respect to ψ that includes β. The methods of Section 3 can be applied straightforwardly. Second, the state vector α_t in the signal model (9) can be augmented so that it includes the coefficients in β. In this case (9) becomes

$$\alpha_t = \begin{pmatrix} h_t \\ \beta \end{pmatrix}, \quad \theta_t = (1, W_t^{\theta})\alpha_t, \quad \alpha_{t+1} = \begin{bmatrix} \phi & 0 \\ 0 & I \end{bmatrix} \alpha_t + \begin{pmatrix} \xi_t \\ 0 \end{pmatrix}, \quad Q_t = \begin{bmatrix} \sigma_{\xi}^2 & 0 \\ 0 & 0 \end{bmatrix},$$

for $t = 1, \ldots, n$, where h_t is taken here as an AR process of order $p = 1$. The initial variance matrix of the state vector P is block-diagonal with the two blocks given by $\sigma_{\xi}^2 \,/\, (1-\phi^2)$ and qI, where $q \to \infty$ and I is the $k^{\theta} \times k^{\theta}$ identity matrix. The diffuse prior for β reflects that β is fixed and unknown. State-space methods can be adjusted to handle diffuse prior conditions exactly; see Chapter 5 in Durbin and Koopman (2001). The Monte Carlo likelihood analysis of Section 3 also applies in this case. Depending on which way β is

estimated, the estimates will not be the same since the conditional likelihood approach produces a Monte Carlo maximum likelihood estimate, while the marginal likelihood approach produces a Monte Carlo conditional mean of β given the observations y; however, the differences are likely to be small. A similar discussion applies to linear Gaussian state-space models; see de Jong (1991).

Finally, regressors for the expected return can be included by the specification

$$\mu_t = \mu + W_t^{\mu}\delta,$$

where μ is an unknown fixed constant, W_t^{μ} is a $1 \times k^{\mu}$ vector of covariates and δ is a $k^{\mu} \times 1$ vector of regression coefficients. In this case we can regard model (28) as a regression model with SV errors. We include μ and δ in ψ and estimate ψ by Monte Carlo maximum likelihood using the methods of Section 3.

4.4 Heavy-tailed innovations

The excess kurtosis found in financial time series is often larger than can be explained by the basic SV model. This is caused by the fact that the excess kurtosis in the SV model is generated solely by randomness in the volatility. The model can be generalized by assuming that the innovations ε_t have a scaled t distribution. In this way, the dynamic properties of log volatility and and the thickness of tails are modeled separately. Examples of this approach can be found in Fridman and Harris (1998), Liesenfeld and Jung (2000) and Lee and Koopman (2004). We consider the SV model (28) with a scaled t distribution for ε_t. Its observation density is given by

$$\log p(y_t|\theta_t) = \log \frac{\Gamma(\frac{\nu}{2} + \frac{1}{2})}{\Gamma(\frac{\nu}{2})} - \frac{1}{2}\left\{\log \sigma^2(\nu-2) + \theta_t + (\nu+2)\log\left(1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right)\right\},$$

for $t = 1, \ldots, n$, while the first and the second derivative with respect to θ_t are given by

$$\dot{p}(y_t|\theta_t) = \frac{1}{2} + \frac{1}{2}\frac{(\nu+1)y_t^2}{(\nu-2)\sigma_t^2 + y_t^2}, \qquad \ddot{p}(y_t|\theta_t) = -\frac{1}{2}\frac{\sigma_t^2(\nu-2)(\nu+1)y_t^2}{\{(\nu-2)\sigma_t^2 + y_t^2\}^2},$$

for $t = 1, \ldots, n$. The further details of a Monte Carlo likelihood analysis for this model are not different from those provided for a basic SV model in Section 3.

4.5 Additive noise

The basic SV model assumes that there is only one source of error. The SV model with additive noise assumes that there is an additional Gaussian noise term with constant variance, more specifically

$$\mu_t = \zeta_t, \qquad \zeta_t \sim \text{NID}(0, \sigma_\zeta^2).$$

This model was used in the context of high-frequency returns in Jungbacker and Koopman (2005). In the special case $\sigma_\zeta^2 = 0$ this model reduces to the basic SV model. The observation density is given by

$$\log p(y_t|\theta_t) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\left(\sigma_\zeta^2 + \sigma_t^2\right) - \frac{1}{2}\frac{y_t^2}{(\sigma_\zeta^2 + \sigma_t^2)},$$

while the first and the second derivative with respect to θ_t are given by

$$\dot{p}(y_t|\theta_t) = \frac{\sigma_t^2}{2(\sigma_t^2 + \sigma_\zeta^2)}\left\{\frac{y_t^2}{(\sigma_t^2 + \sigma_\zeta^2)} - 1\right\}$$

and

$$\ddot{p}(y_t|\theta_t) = \frac{\sigma_t^2}{2(\sigma_t^2 + \sigma_\zeta^2)}\left\{\frac{\sigma_t^2 + y_t^2}{(\sigma_t^2 + \sigma_\zeta^2)} - \frac{\sigma_t^2 y_t^2}{(\sigma_t^2 + \sigma_\zeta^2)^2} - 1\right\},$$

for $t = 1, \ldots, n$, respectively. Since $\ddot{p}(y_t|\theta_t)$ is not necessarily negative for all $t = 1, \ldots, n$, the observation density is not necessarily log-concave. We therefore need to rely on the arguments of Jungbacker and Koopman (2007) to carry out a Monte Carlo likelihood analysis. However, the methods and techniques have not changed intrinsically and the descriptions in Section 3 still apply.

4.6 Leverage effects

The leverage effect occurs if a negative return increases the volatility more than a positive return of the same magnitude decreases it; see the seminal paper of Black (1976) where this phenomenon was described originally. The leverage effect is incorporated in the SV model by allowing correlation between the innovations of the state and the observation equation; see Yu (2005) for a detailed discussion. The SV model with leverage and based on an AR(1) process for log volatility is given by

$$y_t = \sigma \exp(\frac{1}{2}h_t)\varepsilon_t, \qquad h_{t+1} = \phi h_t + \xi_t, \qquad \begin{pmatrix}\varepsilon_t \\ \xi_t\end{pmatrix} \sim \text{NID}\left(0, \begin{bmatrix}1 & \sigma_\xi\rho \\ \sigma_\xi\rho & \sigma_\xi^2\end{bmatrix}\right),$$

for $t = 1, \ldots, n$. The correlation coefficient ρ is typically negative, implying that negative shocks in the return are accompanied by positive shocks in the volatility and vice versa.

The general formulation of the SV model with leverage requires both h_t and ξ_t with $t = 1, \ldots, n$ in θ since h_t appears directly in the observation equation and ξ_t is required to measure the correlation with ε_t. The variance matrix Ψ of the signal θ in (12) is therefore singular and the methods of Section 3 clearly break down. We treat this problem by following Jungbacker and Koopman (2007) and reformulate the model by

$$y_t = \sigma \exp(\frac{1}{2}h_t^*) \left\{\varepsilon_t^* + \text{sign}(\rho)\xi_{2t}\right\}, \qquad \varepsilon_t^* \sim \text{NID}(0, 1 - |\rho|),$$

where

$$h_{t+1}^* = \phi h_t^* + \sigma_\xi \left(\xi_{1,t} + \xi_{2t}\right), \quad \xi_{1t} \sim \text{NID}(0, 1 - |\rho|), \quad \xi_{2t} \sim \text{NID}(0, |\rho|),$$

for $t = 1, \ldots, n$, with $h_1^* \sim N\{0, \sigma_\xi^2(1 - \phi^2)^{-1}\}$. The disturbances ε_t^*, ξ_{1t} and ξ_{2t} are mutually and serially independent disturbances for $t = 1, \ldots, n$. The signal vector for this model formulation contains h_t^* and ξ_{2t} with $t = 1, \ldots, n$ and the corresponding variance matrix Ψ is nonsingular as required. In terms of the general formulation (9), we have $\alpha_t = (h_t^* \ , \ \sigma_\xi \xi_{2,t})'$, $\xi_t = \sigma_\xi(\xi_{1,t} \ , \ \xi_{2,t+1})'$ and

$$\theta_t = \alpha_t, \quad \alpha_{t+1} = \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix} \alpha_t + \xi_t, \quad \begin{aligned} \xi_t &\sim \text{NID}\left\{0, \sigma_\xi^2 \text{diag}(1 - |\rho|, |\rho|)\right\}, \\ \alpha_1 &\sim \text{NID}\left\{0, \sigma_\xi^2 \text{diag}([1 - \phi^2]^{-1}, |\rho|)\right\}, \end{aligned}$$

for $t = 1, \ldots, n$. The observations $y_1, \ldots, y_n$ have the conditional density $\log p(y|\theta) = \sum_{t=1}^n \log p(y_t|\theta_t)$, where

$$\log p(y_t|\theta_t) = c - \frac{1}{2}h_t^* - \frac{1}{2}\sigma^{-2} \exp(-h_t^*)(1-|\rho|)^{-1}\{y_t - \sigma \exp(\frac{1}{2}h_t^*)\text{sign}(\rho)\xi_{2,t}\}^2,$$

for $t = 1, \ldots, n$, where c is some constant. Expressions for the 2×1 vector $\dot{p}(y_t|\theta_t)$ and the 2×2 matrix $\ddot{p}(y_t|\theta_t)$, as defined in (22), can be obtained straightforwardly. It turns out that density $p(y|\theta)$ is not necessarily log-concave. The Monte Carlo estimator (27) of the likelihood function can be evaluated by using the methods of Section 3 and by adopting the arguments of Jungbacker and Koopman (2007), who also presented an empirical illustration.

4.7 Stochastic volatility in mean

As investors require a larger expected return if the risk is large, it seems reasonable to expect a positive relationship between volatility and returns. Empirical evidence however points to a negative influence of volatility on returns; see French et al. (1987). This effect can be explained by assuming a positive relationship between expected return and *ex ante* volatility. Koopman and Hol-Uspensky (2002) proposed capturing this so-called volatility feedback effect by including volatility as a regression effect in the mean:

$$\mu_t = a + by_{t-1} + d\sigma \exp\left(\frac{1}{2}\theta_t\right),$$

where a, b, d and σ^2 are parameters. We assume that both mean and variance processes are stationary. The volatility feedback effect coefficient d is typically negative, if not zero.

The observation density is given by

$$\log p(y_t|\theta_t) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}h_t - \frac{\left(y_t - a - by_{t-1} - d\sigma_t^2\right)^2}{2\sigma_t^2},$$

while the first and the second derivative with respect to θ_t are given by

$$\dot{p}(y_t|\theta_t) = -\frac{1}{2} + \left(y_t - a - by_{t-1} - d\sigma_t^2\right)d + \frac{\left(y_t - a - by_{t-1} - d\sigma_t^2\right)^2}{2\sigma_t^2}$$

and

$$\ddot{p}(y_t|\theta_t) = -\frac{1}{2} - d^2\sigma_t^2 - \dot{p}(y_t|\theta_t),$$

for $t = 1, \ldots, n$, respectively. It can be shown that density $p(y|\theta)$ is log-concave, so $\ddot{p}(y_t|\theta_t) < 0$ for all $t = 1, \ldots, n$. The further details of a Monte Carlo likelihood analysis for this model are therefore not different from those provided for a basic SV model in Section 3.

5 Empirical Illustrations

Three financial time series are analyzed to illustrate the methods presented in this chapter. The daily returns of the Standard & Poor's 500 (S&P500) stock index (January 3, 1991 to October 20, 2006: 3,985 observations, weekends and holidays excluded), the daily changes in the US dollar–pound sterling exchange rates and the daily changes in the US dollar–Japanese yen exchange rates (both for January 3, 1990 to October 20, 2006: 4,383 observations,

weekends excluded but with missing values for holidays). These time series are obtained from Datastream and are presented graphically in Fig. 1.

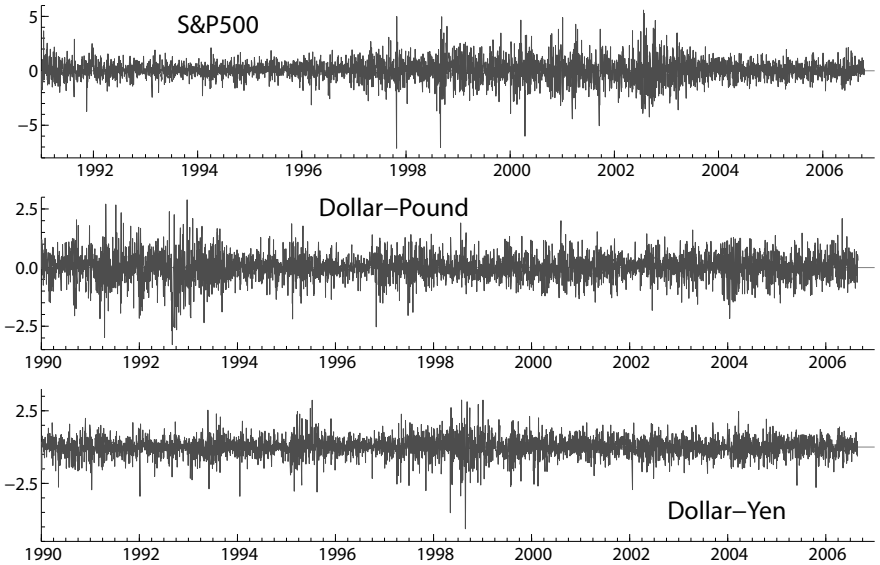

Fig. 1 Daily returns of S&P500 and daily changes of dollar–pound and dollar–yen exchange rate series

5.1 Standard & Poor's 500 stock index: volatility estimation

The volatility in the S&P500 returns is subject to some breaks: low volatility before 1997, high volatility between 1997 and 2004 with peaks in the aftermath of September 11, 2001 and moderate volatility from the end of 2003. Further, the clustering of periods with low and high volatility is clearly visible in the series. An exploratory data analysis is carried out firstly and the STAMP program of Koopman et al. (2007) is used for the analysis, estimation and signal extraction on the basis of the linearized SV model. The estimation results and the estimated volatility are indicative of the salient features in the volatility of S&P500 returns.

The Monte Carlo likelihood methods of Section 3 are used for the estimation of parameters with $M = 100$. Various SV models are considered for this purpose. First, the parameters in the SV models with log volatility modeled by a single AR(1) term (basic SV) and by two AR(1) terms (multiple volatility factors) are estimated by Monte Carlo maximum likelihood and the

results are presented in the second and third columns of Table 1. The standard errors of the estimates are based on the information matrix, which is evaluated numerically. The estimated coefficients are reported together with their standard errors in parentheses and it appears that for both models the estimates are significantly different from zero. However, the estimated autoregressive coefficients are close to unity and it is known that standard errors for parameter estimates close to boundary values are not reliable; see also the extensive literature on unit root testing. However, the likelihood ratio test value for the inclusion of a second AR(1) term in the SV model with multiple volatility factors is 18.8 and has a p value of 0.0001. The estimated volatilities are presented in Fig. 2. The main patterns reflect the salient features of the dynamics in volatility. The volatility increase in 1997, the peak in the early years of the twenty-first century and the slowdown in 2003 can be observed clearly from the estimated volatility patterns. The distinct difference between the estimated volatility patterns of the one- and two-factor volatilities is that the signal is noisier for the multiple-factor SV model; however, the main patterns are similar. Given the likelihood improvement, the noisier volatility estimate for the two-factor model appears to provide a better local estimate of the underlying volatility.

The basic SV model with a t density for ε_t is also considered for the S&P500 series. The parameter estimates can be found in the fourth column of Table 1. It is expected that the large shocks in the return series will be captured more effectively in the tails of the t density and therefore the estimated volatility is more persistent and smoother. The estimation results confirm this. The estimated value for ϕ is 0.995, which is slightly larger than the value of 0.991 for the basic SV model. Furthermore, the estimate of σ_ξ^2 (the volatility of volatility) has a smaller value for the SV model with t density than for the basic SV model and hence the estimated volatility is smoother. This is confirmed by the bottom graph in Fig. 2. The estimated underlying volatility pattern for the S&P500 return series is smoothest for the SV model with t density.

5.2 Standard & Poor's 500 stock index: regression effects

SV models can incorporate regression effects in different ways as shown in Section 4. The SV in the mean model is considered to investigate the feedback between returns and volatility. For this purpose we consider daily excess returns for the S&P500 stock index series where excess return is defined as the return minus the risk-free return. The estimation is carried out using the Monte Carlo likelihood methods for which some details are given in Section 4. The estimation results are presented in the final column of Table 1. Apart from the basic SV parameters, the parameters of interest are a, b and d,

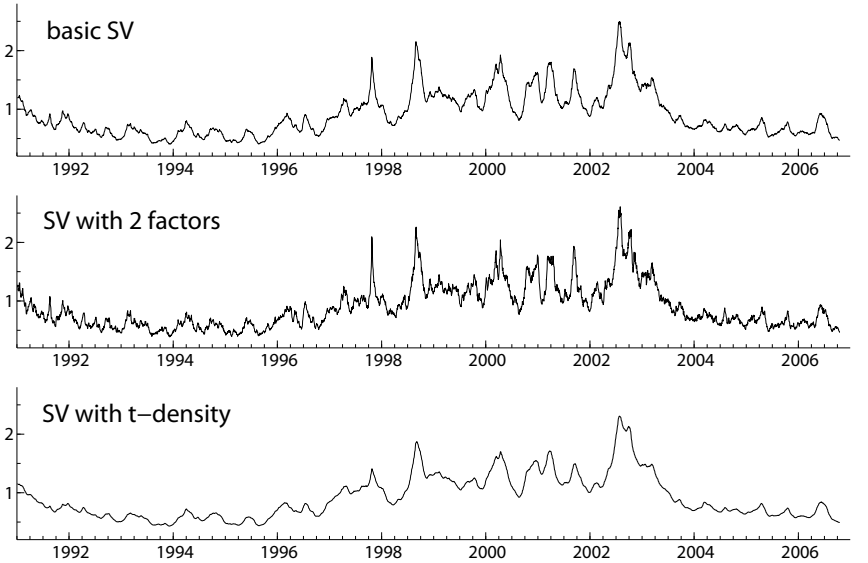

Fig. 2 Estimated volatilities σ_t for S&P500 returns and for three stochastic volatility (SV) models: basic, with two factors and with t density

Table 1 Monte Carlo maximum likelihood estimates of parameters (with standard errors) for stochastic volatility (SV) models applied to S&P 500 daily returns (SV-t: SV model with t density; for the SV in mean model applied to excess returns, the deviations are from risk-free returns)

Coefficients	Basic SV	SV–2 factors	SV-t	SV in mean
ϕ	0.991 (0.003)	0.998 (0.001)	0.995 (0.002)	0.976 (0.006)
σ_ξ^2	0.0127 (0.003)	0.0014 (0.0008)	0.006 (0.00186)	0.028 (0.006)
ϕ	—	0.93 (0.029)	—	—
σ_ξ^2	—	0.028 (0.01)	—	—
σ^2	0.681 (0.129)	0.643 (0.227)	0.695 (0.17401)	0.672 (0.072)
ν	—	—	10.651 (2.1058)	—
a	—	—	—	0.072 (0.018)
b	—	—	—	0.051 (0.016)
d	—	—	—	−0.038 (0.025)

which are estimated significantly except parameter d; however, the estimate of d is not too far away from its significance level and it is a negative value. This confirms the empirical findings of French et al. (1987) and Koopman and Hol-Uspensky (2002), among others.

An alternative volatility indicator is the daily price range defined as $Z_t^\beta = \log(\max(\log Pr_t) - \min(\log Pr_t))$ where Pr_t is the vector of realized stock prices within a certain day t. This daily S&P500 stock index price range is also indicative of the amount of volatility and can be used to detect volatility in daily returns. We therefore consider the SV model with $\theta = h_t + Z_t^\beta \beta$, the estimation results for which are presented in Table 2. Although similar results are obtained as for the SV models without the inclusion of the covariate Z_t^β in the signal θ_t, the estimates for β are rather mixed. The SV model with one volatility factor produces a highly significant and positive estimate for β, the model with two volatility factors produces a highly significant but negative estimate for β, while the SV model with a t density produces an insignificant estimate for β. Further research and empirical evidence is required to assess the increased performance of incorporating range-based volatility measures in SV models.

Table 2 Monte Carlo maximum likelihood estimates of parameters (with standard errors) from *SVX* models for S&P 500 returns

Coefficients	Basic SV	SVX	SVX–2 factors	SVX-t
ϕ	0.991 (0.003)	0.991 (0.0029)	0.998 (0.0009)	0.996 (0.0019)
σ_ξ^2	0.0127 (0.003)	0.0124 (0.003)	0.0014 (0.0008)	0.006 (0.0017)
ϕ	—	—	0.929 (0.029)	—
σ_ξ^2	—	—	0.031 (0.011)	—
σ^2	0.681 (0.129)	0.71 (0.234)	0.564 (0.267)	0.911 (0.256)
ν	—	—	—	10.541 (0.351)
β	—	0.0156 (0.003)	−0.033 (0.005)	0.002 (0.002

5.3 Daily changes in exchange rates: dollar–pound and dollar–yen

For the analysis of volatility in exchange rate series, the interest focuses solely on the signal extraction of volatility. For this purpose we have considered the exchange rates for pound sterling and Japanese yen, both against the US dollar. Three different SV model specifications are applied to the two daily change series. The first model is the basic SV model, the second is the SV

model with additive noise and the third is the SV model with a t density for the model equation. In all cases the estimation methods of Section 3 are used. In the case of the SV with noise model, the observation density $p(y|\theta)$ is not log-concave and the recent modifications need to be applied. The implementation in all three cases has been successful and the estimation results are presented in Table 3. The persistency of the volatility clearly increases for an SV model with a t density, while the additive noise does not seem to affect the dynamic properties of volatility. Although the additive noise is highly significant for the dollar–yen series, it is not significant for the dollar–pound series. This empirical finding may be explained by factors such as trading volumes and information flows.

Table 3 Monte Carlo maximum likelihood estimates of parameters (with standard errors) from SV models for daily changes in exchange rates (*SV-t*: SV model with t density; *SVN*: SV model with additive noise

	Dollar–pound			Dollar–yen		
Coefficients	Basic SV	SVN	SV-t	Basic SV	SVN	SV-t
ϕ	0.977 (0.00606)	0.977 (0.00609)	0.986 (0.00435)	0.934 (0.02)	0.933 (0.02)	0.987 (0.0041)
σ_ξ^2	0.0179 (0.00477)	0.0200 (0.0075)	0.00976 (0.00294)	0.0521 (0.018)	0.115 (0.037)	0.0064 (0.002)
σ^2	0.265 (0.0245)	0.248 (0.0487)	0.274 (0.0306)	0.371 (0.0222)	0.224 (0.031)	0.404 (0.041)
σ_ζ^2	—	0.0144 (0.0348)	—	—	0.115 (0.021)	—
ν	—	—	10.459 (1.843)	—	—	6.561 (0.681)

The estimated volatilities obtained from the three models and the two exchange series are presented in Fig. 3 for the dollar–pound series and in Fig. 4 for the dollar–yen series. The salient features of the volatility for both series are clearly captured and it is interesting that the volatility patterns for both series are distinct from each other. In the early years of the 1990s, the dollar–pound series is subject to higher volatility. After 1994, the volatility is moderate for 10 years but increases somewhat from 2004. Throughout the 1990s, the volatility is relatively high for the dollar–yen series and the volatility of volatility is also high. A clearly high period of volatility occurred during the Asian financial crises in the late 1990s. However, in the early years of the new millennium, the volatility stabilizes and seems to behave more in par with the dollar–pound series. This may be indicative of the convergence of international financial markets that is discussed in the economics and finance literature.

The differences in the estimated volatilities for the three models are similar for both exchange series. The volatility patterns for the SV model with noise also turn out to be noisier than those obtained for the basic SV model. The smoothed patterns of volatility are obtained from the SV model with a t density. Given the estimation results presented in Table 3, the estimated

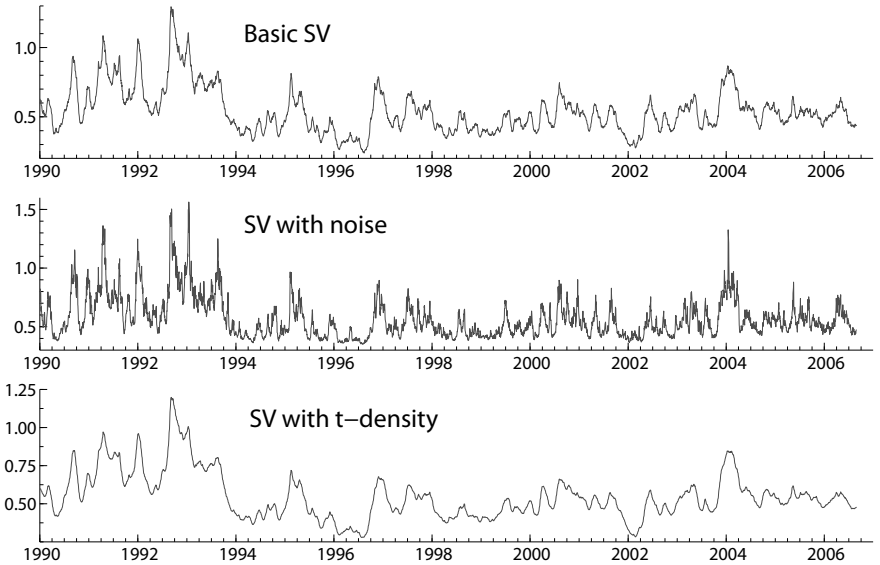

Fig. 3 Volatility estimates for daily changes in dollar–pound exchange rates

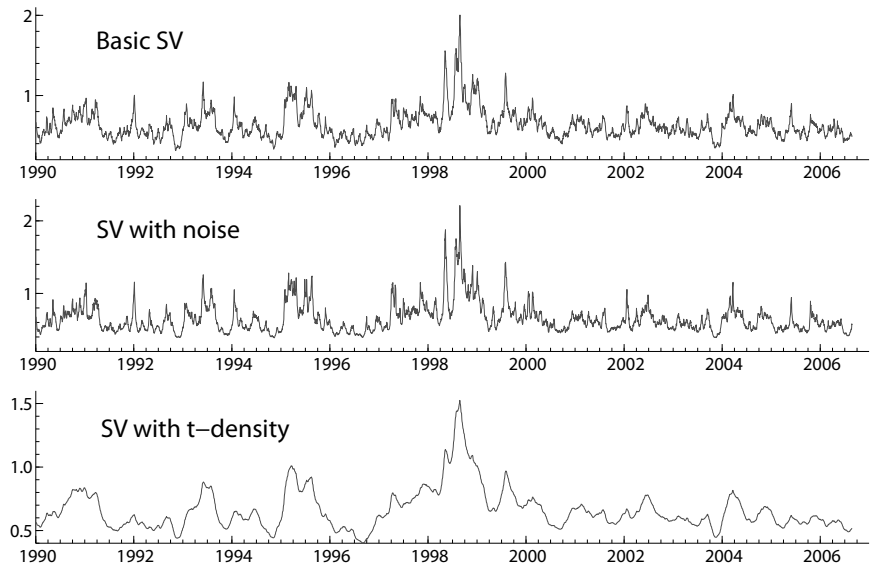

Fig. 4 Volatility estimates for daily changes in dollar–yen exchange rates

volatility patterns confirm that the different SV models capture different features of volatility in the series.

6 Conclusions

This chapter has reviewed parameter estimation methods for the basic SV model. We have restricted ourselves to estimation methods based on the linearization of the SV model and methods based on Monte Carlo simulations. In the former case, fast and linear estimation methods for the standard linear Gaussian state-space model can be adopted, such as the Kalman filter and the associated smoothing algorithm. Classical maximum likelihood methods are used for parameter estimation and for which standard software tools are available. In the latter case, Monte Carlo simulations are used for the evaluation of the likelihood function that is expressed as an integral. A convenient analytical expression is not available and therefore one needs to rely on numerical methods. Importance sampling methods have been suggested as a feasible way to evaluate the likelihood function so that it can be maximized numerically with respect to a set of parameters. The details of this approach were discussed in this chapter and particularly in Section 3. This chapter further reviewed some interesting extensions of the SV model, including models with explanatory variables, with additive noise, with leverage and with a t density for the observation model. It was shown that parameters in these more general SV models can also be estimated by the Monte Carlo methods discussed in this chapter. The empirical results illustrate that the general SV models and the associated methods can successfully capture interesting aspects of volatility.

Acknowledgement We gratefully thank the editors for their insightful comments on an earlier draft of this chapter. All remaining errors are ours.

Appendix: State-Space Methods

In this chapter we consider the linear Gaussian state space model with

- State equation:

$$\alpha_{t+1} = d_t + T_t\alpha_t + \eta_t, \qquad \eta_t \sim \mathrm{NID}(0, Q_t).$$

- Initial condition: $\alpha_1 \sim \mathrm{N}(a, P)$.
- Signal equation:

$$\theta_t = c_t + Z_t\alpha_t.$$

- Observation equation:

$$x_t = \theta_t + u_t, \qquad u_t \sim \text{NID}(0, H_t), \qquad t = 1, \ldots, n.$$

The system vectors c_t and d_t and system matrices Z_t, H_t, T_t and Q_t are fixed and known functions of parameter vector ψ. The observation x_t and signal θ_t are assumed to be scalar variables, while the state vector α_t together with the disturbance vector η_t have dimensions $q \times 1$. All state-space quantities have appropriate dimensions.

Kalman filter

The Kalman filter equations are given by

$$\begin{aligned} v_t &= x_t - c_t - Z_t a_t, & F_t &= H_t + Z_t P_t Z_t', \\ & & K_t &= T_t P_t Z_t' F_t^{-1}, & t &= 1, \ldots, n, \\ a_{t+1} &= d_t + T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t T_t' - K_t F_t K_t' + Q_t, \end{aligned} \tag{30}$$

for $t = 1, \ldots, n$. The Kalman filter is initialized by $a_1 = a$ and $P_1 = P$, where a and P are the mean and variance of the initial state vector α_1, respectively. The quantities of the Kalman filter are usually stored so they can be used for other purposes such as smoothing. The Kalman filter is a forwards recursion. The one-step-ahead prediction errors v_t and their variances F_t are used for the evaluation of the Gaussian likelihood function and as residuals for diagnostic checking.

Smoothing algorithm

Once the Kalman filter has been carried out and the quantities K_t, F_t and v_t have been stored for $t = 1, \ldots, n$, the smoothing recursions enable the estimation of the smoothed estimate of θ_t and its variance matrix given by $\widehat{\theta}_t = E(\theta_t | y)$ and $V_t = E(\theta_t - \widehat{\theta}_t)(\theta_t - \widehat{\theta}_t)'$. The recursions operate backwards and are given by

$$\begin{aligned} \widehat{\theta}_t &= x_t - H_t u_t, & V_t &= H_t D_t H_t, \\ u_t &= F_t^{-1} v_t - K_t' r_t & D_t &= F_t^{-1} - K_t' N_t K_t, \\ r_{t-1} &= Z_t' u_t + T_t' r_t, & N_{t-1} &= Z_t' D_t Z_t + T_t' N_t T_t - Z_t' K_t' N_t T_t - T_t' N_t K_t Z_t, \end{aligned} \tag{31}$$

for $t = n,\ n-1, \ldots, 1$ and with the initializations $r_n = 0$ and $N_n = 0$.

Simulation smoothing algorithm

Once the Kalman filter has been carried out and the quantities K_t, F_t and v_t have been stored for $t = 1, \ldots, n$, the simulation smoothing recursions enable the generation of draws from the density $f(\theta; y)$. The recursions operate backwards as in smoothing and are given by

$$\begin{aligned} C_t = B_t B_t' = H_t^{-1} - F_t^{-1} - K_t' N_t K_t, \quad & R_t = C_t^{-1}(H_t^{-1} Z_t - K_t' N_t T_t), \\ o_t \sim \mathrm{N}(0, I), \qquad w_t \;=\; B_t o_t, \qquad & u_t = H_t(w_t + F_t^{-1} v_t - K_t' r_t), \\ r_{t-1} = Z_t' H_t^{-1} u_t - R_t' w_t + T_t' r_t, \qquad & N_{t-1} = R_t' C_t R_t - Z_t' H_t^{-1} Z_t + T_t' N_t T_t, \end{aligned} \tag{32}$$

for $t = n,\ n-1, \ldots, 1$ and with the initializations $r_n = 0$ and $N_n = 0$. The simulation θ^i from $p(\theta|y)$ is computed by $\widehat{\theta} + (u_1', \ldots, u_n')'$, where $\widehat{\theta} = E(\theta|y)$ and is obtained from the smoothing algorithm. Given the simulation θ^i, the importance density function $f(\theta^i; y)$ in (27) can be evaluated by

$$f(\theta^i; y) = \exp\left(-\frac{mn}{2} \log 2\pi - \sum_{t=1}^{n} \log |H_t| - \sum_{t=1}^{n} \log |B_t| - \frac{1}{2} \sum_{t=1}^{n} o_t^i \, {}' o_t^i \right). \tag{33}$$

References

Andersen, T., Bollerslev, T., Diebold, F. and Labys, P. (2003): Modelling and forecasting realized volatility. *Econometrica* **71**, 529–626.

Andersen, T. G. and Sorensen, B. (1996): GMM estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Business and Economic Statistics* **14**, 328–352.

Black, F. (1976): Studies of stock price volatility changes. *Proceedings of the Business and Economic Statistics Section*, 177–181.

Bollerslev, T. (1986): Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics* **51**, 307–327.

Breidt, F. J. and Carriquiry, A. L. (1996): Improved quasi-maximum likelihood estimation for stochastic volatility models. *In: Jack, W. O. J., Lee, C. and Zellner, A. (Eds.): Modelling and Prediction: Honoring Seymour Geisser*. Springer, New York, 228–247.

Breidt, F. J., Crato, N. and de Lima, P. (1998): On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* **83**, 325–348.

Brockwell, A. E. (2007): Likelihood-based analysis of a class of generalized long-memory time series models. *Journal of Time Series Analysis* **28**, 386–407.

Davis, R. A. and Rodriguez-Yam, G. (2005): Estimation for state-space models based on a likelihood approximation. *Statistica Sinica* **15**, 381–406.

de Jong, P. (1991): The diffuse Kalman filter. *Annals of Statistics* **19**, 1073–83.

de Jong, P. and Shephard, N. (1995): The simulation smoother for time series models. *Biometrika* **82**, 339–50.

Durbin, J. and Koopman, S. J. (1997): Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika* **84**, 669–84.

Durbin, J. and Koopman, S. J. (2001): *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.

Durbin, J. and Koopman, S. J. (2002): A simple and efficient simulation smoother for state space time series analysis. *Biometrika* **89**, 603–16.

Engle, R. F. (1982): Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**, 987–1007.

French, K. R., Schwert, G. W. and Stambaugh, R. F. (1987): Expected stock returns and volatility. *J. Financial Economics* **19**, 3–29. Reprinted as pp. 61–86 in Engle, R.F.(1995): *ARCH: Selected Readings*. Oxford University Press, Oxford.

Fridman, M. and Harris, L. (1998): A maximum likelihood approach for non-gaussian stochastic volatility models. *Journal Business and Economic Statistics* **16**, 284–291.

Fuller, W. A. (1996): *Introduction to Time Series* (2nd ed.) Wiley, New York.

Geweke, J. (1989): Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57**, 1317–39.

Granger, C. W. J. and Joyeau, R. (1980): An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**, 15–39.

Harvey, A. C., Koopman, S. J. and Penzer, J. (1998): Messy time series. *In: Fomby, T. B. and Hill, R. C. (Eds.): Advances in Econometrics* **13**, 103–143. JAI Press, New York.

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

Jacquier, E., Polson, N. and Rossi, P. (1994): Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics* **12**, 371–417.

Jungbacker, B. and Koopman, S. J. (2005): Model-based measurement of actual volatility in high-frequency data. *In: Fomby, T. B. and Terrell, D. (Eds.): Advances in Econometrics*. JAI Press, New York.

Jungbacker, B. and Koopman, S. J. (2007): Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika* **94**, 827–39.

Kim, S., Shephard, N. and Chib, S. (1998): Stochastic volatility: likelihood inference and comparison with arch models. *Review of Economic Studies* **65**, 361–393.

Koopman, S. J., Harvey, A. C., Doornik, J. A. and Shephard, N. (2007): *Stamp 8.0: Structural Time Series Analyser, Modeller and Predictor*. Timberlake Consultants, London.

Koopman, S. J. and Hol-Uspensky, E. (2002): The Stochastic Volatility in Mean model: Empirical evidence from international stock markets. *Journal of Applied Econometrics* **17**, 667–89.

Koopman, S. J., Shephard, N. and Doornik, J. A. (1999): Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal* **2**, 113–66. http://www.ssfpack.com/.

Koopman, S. J., Jungbacker, B. and Hol, E. (2005): Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* **12**, 445–75.

Koopman, S. J., Shephard, N. and Creal, D. (2007): Testing the assumptions behind importance sampling. *Discussion paper, VU University, Amsterdam*.

Lee, K. M. and Koopman, S. J. (2004): Estimating stochastic volatility models: a comparison of two importance samplers. *Studies in Nonlinear Dynamics and Econometrics* **8**, Art 5.

Liesenfeld, R. and Jung, R. (2000): Stochastic volatility models: conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* **15**, 137–160.

Nocedal, J. and Wright, S. J. (1999): *Numerical Optimization* Springer, New York.

Ray, B. and Tsay, R. (2000): Long- range dependence in daily stock volatilities. *Journal of Business and Economic Statistics* **18**, 254–62.

Sandmann, G. and Koopman, S. J. (1998): Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics* **87**, 271–301.

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

Shephard, N. and Pitt, M. K. (1997): Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653–67.

So, M. K. P. (2003): Posterior mode estimation for nonlinear and non-Gaussian state space models. *Statistica Sinica* **13**, 255–274.

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

Tsiakas, I. (2006): Periodic stochastic volatility and fat tails. *Journal of Financial Econometrics* **4**, 90–135.

Yu, J. (2005): On leverage in a stochastic volatility model. *Journal of Econometrics* **127**, 165–78.

Zivot, E. (2008): Practical aspects of GARCH-Modelling. *In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.): Handbook of Financial Time Series*, 112–155. Springer, New York.

Zivot, E., Wang, J. and Koopman, S. J. (2004): State space modeling in macroeconomics and finance using SsfPack in S+FinMetrics. *In: Harvey, A. C., Koopman, S. J. and Shephard, N. (Eds.): State space and unobserved components models*. Cambridge University Press, Cambridge.