12 Stochastic Volatility Estimation Using Markov Chain Simulation

Nikolaus Hautsch and Yangguoyi Ou

Stochastic volatility (SV) models are workhorses for the modelling and prediction of time-varying volatility on financial markets and are essential tools in risk management, asset pricing and asset allocation. In financial mathematics and financial economics, stochastic volatility is typically modeled in a continuous-time setting which is advantageous for derivative pricing and portfolio optimization. Nevertheless, since data is typically only observable at discrete points in time, in empirical applications, discrete-time formulations of SV models are equally important.

SV models can be economically motivated by the mixture-of-distribution hypothesis (MDH) postulated by Clark (1973), whereby asset returns follow a mixture of normal distributions with a mixing process depending on the (unobservable) information arrival process. If the mixing process is positively autocorrelated, the resulting return process reveals volatility clustering which is a well-known and typical feature of financial return series. The MDH gives rise to the idea that asset return volatility follows its own stochastic process which is updated by unobservable innovations. This is in contrast to an autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982), where the conditional variance given the available information set is a function of past observations. Denote h_t as the time-t conditional variance of asset return y_t with conditional mean μ_t and $y_t - \mu_t = h_t^{1/2} z_t$, $z_t \sim IID(0,1)$, and let $\mathcal{F}_t$ denote the time-t information set. Then, ARCH processes imply $\mathsf{Var}[h_t|\mathcal{F}_{t-1}] = 0$, i.e., the variance is conditionally deterministic given the (observable) history of the process. Conversely, SV models can be characterized by the property $\mathsf{Var}[h_t|\mathcal{F}_{t-1}] \neq 0$, i.e., there is an unpredictable component in h_t.

A main difficulty of the SV framework compared to the widely used (Generalized) ARCH model is that the likelihood of SV models is not directly available. This requires the use of simulation techniques, like simulated maximum likelihood, method of simulated moments or Markov chain Monte Carlo (MCMC) techniques. Because of the computational costs, SV models are

still less popular in financial practice. Nevertheless, increasing computer power and the further development of efficient sampling techniques weaken this drawback noticeably. Furthermore, recent literature on the estimation of realized volatility confirms the idea of the MDH that log returns follow a normal - log normal mixture (see, e.g., Andersen, Bollerslev, Diebold and Labys (2003)) and thus strengthens the economic foundation of the SV model. Finally, SV models provide a natural framework to accommodate specific properties of financial return processes such as fat-tailedness, leverage effects and the occurrence of jumps.

The main objective of this chapter is to present the most important specifications of discrete-time SV models, to illustrate the major principles of Markov Chain Monte Carlo (MCMC) based statistical inference, and to show how to implement these techniques to estimate SV models. In this context, we provide a hands-on approach which is easily extended in various directions. Moreover, we will illustrate empirical results based on different SV specifications using returns on stock indices and foreign exchange rates.

In Section 12.1, we will introduce the standard SV model. Section 12.2 presents several extended SV models. MCMC based Bayesian inference is discussed in Section 12.3, whereas empirical illustrations are given in Section 12.4.

12.1 The Standard Stochastic Volatility Model

The standard stochastic volatility model as introduced by Taylor (1982) is given by

$$y_t = \exp(h_t/2)u_t, \qquad u_t \sim \mathrm{N}(0,1), \tag{12.1a}$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \tag{12.1b}$$

where y_t denotes the log return at time t, $t = 1, \ldots, T$, and h_t is the log volatility which is assumed to follow a stationary AR(1) process with persistence parameter $|\phi| < 1$. The error terms u_t and η_t are Gaussian white noise sequences. The unconditional distribution of h_t is given by

$$h_t \sim \mathrm{N}\left(\mu_h, \sigma_h^2\right), \quad \mu_h = \mu, \quad \sigma_h^2 = \frac{\sigma_\eta^2}{1-\phi^2}, \tag{12.2}$$

where μ_h and σ_h^2 denote the unconditional mean and variance of returns, respectively.

Under the assumption that $\mathsf{E}[y_t^4] < \infty$, the first two even moments of y_t are given by

$$\mathsf{E}[y_t^2] = \mathsf{E}[\exp(h_t)]\,\mathsf{E}[u_t^2] \;\; = \exp(\mu_h + \sigma_h^2/2), \tag{12.3}$$
$$\mathsf{E}[y_t^4] = \mathsf{E}[\exp(2h_t)]\,\mathsf{E}[u_t^4] = 3\exp(2\mu_h + 2\sigma_h^2). \tag{12.4}$$

Consequently, the kurtosis is

$$\mathrm{K}(y_t) \stackrel{\text{def}}{=} \frac{\mathsf{E}[y_t^4]}{\mathsf{E}[y_t^2]^2} = 3\exp(\sigma_h^2) = 3\exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right) \tag{12.5}$$

with $\mathrm{K}(y_t) > 3$ as long as $\sigma_\eta^2 > 0$. Hence, the kurtosis generated by SV processes increases with σ_η^2 and $|\phi|$ (given $|\phi| < 1$).

The autocorrelation function (ACF) of y_t^2 is computed as

$$\mathrm{Corr}(y_t^2, y_{t-\tau}^2) = \frac{\exp(\sigma_h^2\phi^\tau) - 1}{3\exp(\sigma_h^2) - 1}, \quad \tau = 1, 2, \ldots, \tag{12.6}$$

and thus decays exponentially in τ. Consequently, for $\phi \in (0, 1)$, squared returns are positively autocorrelated.

The estimation of SV models is not straightforward since the likelihood cannot be computed in closed form. Let θ denote the collection of all model parameters, e.g., $\theta = (\mu, \phi, \sigma_\eta^2)$ for the standard SV model. Then, the likelihood function is defined by

$$p(y|\theta) \stackrel{\text{def}}{=} \int_h p(y|h,\theta)p(h|\theta)dh, \tag{12.7}$$

where $y = (y_1, \ldots, y_T)$ and $h = (h_1, \ldots, h_T)$ are the vectors of returns and latent volatility states, respectively. The so-called full-information likelihood, corresponding to the conditional probability density function (p.d.f.), $p(y|h,\theta)$, is specified by (12.1a), whereas the conditional p.d.f. of the volatility states, $p(h|\theta)$, is given by (12.1b). The likelihood function (12.7) is an analytically intractable T-dimensional integral with respect to the unknown latent volatilities. In the econometric literature, several estimation methods have been proposed, including generalized method of moments (Melino and Turnbull, 1990), quasi-maximum likelihood estimation (Harvey, Ruiz, and Shephard, 1994), efficient method of moments (Gallant, Hsie, and Tauchen, 1997), simulated maximum likelihood (Danielsson, 1994) and efficient importance sampling (Liesenfeld and Richard, 2003). Markov Chain Monte Carlo (MCMC) techniques have been introduced by Jacquier, Polson, and Rossi (1994) and Kim, Shephard, and Chib (1998). More details on MCMC-based inference will be given in Section 12.3.

12.2 Extended SV Models

12.2.1 Fat Tails and Jumps

Though the standard SV model is able to capture volatility clustering typically exhibited by financial and economic time series, the model implied kurtosis is often far too small to match the sample kurtosis observed in most financial return series. See, for example, Liesenfeld and Jung (2000) and Chib, Nardari, and Shephard (2002). An obvious reason is that a normal - log normal mixture as implied by the standard SV model is not flexible enough to capture the fat-tailedness commonly observed in financial return distributions. A further reason is that the basic SV model cannot account for potential jumps in the return process.

In this section, we discuss two SV specifications taking into account both pitfalls. The first one is an extension of the standard SV model allowing the error term u_t to be Student-t distributed resulting in the so-called SVt model. In the second approach, a jump component is introduced in the measurement equation in (12.1). This will lead to the so-called SVJ model.

The SVt Model

The SVt model is specified by

$$y_t = \exp(h_t/2)u_t, \qquad u_t \sim t_v, \tag{12.8a}$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \tag{12.8b}$$

where u_t follows a standardized t-distribution with $v > 2$ degrees of freedom. The model can be alternatively represented by a scale mixture of normal distributions. Let λ_t denote an i.i.d. random variable following an inverse-gamma distribution. Then, the SVt model can be rewritten as

$$y_t = \exp(h_t/2)\sqrt{\lambda_t}u_t, \qquad u_t \sim \mathrm{N}(0,1), \tag{12.9a}$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \tag{12.9b}$$
$$\lambda_t \sim \text{Inv-Gamma}(v/2, v/2), \qquad v > 2, \tag{12.9c}$$

where λ_t itself is a latent variable. The representation of the SVt model in terms of a scale mixture is particularly useful in an MCMC context since it converts a non-log-concave sampling problem into a log-concave one. This allows for sampling algorithms which guarantee convergence in finite time, see ,e.g., Frieze, Kannan and Polson (1994).

Allowing log returns to be Student-t distributed naturally changes the behavior of the stochastic volatility process. In the standard SV model, large values of $|y_t|$ induce large values of h_t. In contrast, with an additional source of flexibility, λ_t, the SVt model can caputure large values of $|y_t|$ without necessarily increasing h_t. A tpyical consequence is that SVt models imply a higher persistence in volatility dynamics than the standard SV model.

Employing simulated maximum likelihood methods Liesenfeld and Jung (2000) provide an estimate $\widehat{\nu} = 6.31$ for the USD/DM foreign exchange (FX) rate from 1980 to 1990, and a value of 6.30 for the USD/JPY FX rate over 5 years from 1981 to 1985. Chib et al. (2002) estimate the SVt model based on MCMC techniques and report an estimate $\widehat{\nu} = 12.53$ for daily S&P 500 returns between July 1962 and August 1997.

The SV Model with Jump Components

The question of to which extent asset return processes are driven by continuous and/or jump components is an ongoing topic in the current literature. Both (G)ARCH and standard SV models rest on the assumption of a continuous price process and thus are not able to accommodate jumps in returns. The latter is particularly important during periods of news arrivals when the market gets under stress and becomes less liquid. However, the SV framework allows for a natural inclusion of a jump component in the return process. This yields the SVJ model given by

$$y_t = k_t q_t + \exp(h_t/2)u_t, \qquad u_t \sim \mathrm{N}(0,1), \tag{12.10}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \tag{12.11}$$

$$k_t \sim \mathrm{N}(\alpha_k, \beta_k), \tag{12.12}$$

$$q_t \sim \mathbb{B}(\kappa), \tag{12.13}$$

where q_t is a Bernoulli random variable taking on the value one whenever a jump occurs with probability κ, and is zero otherwise. The jump size is represented by the time-varying random variable k_t which is assumed to follow a normal distribution with mean α_k and variance β_k. Both q_t and k_t are latent variables. Then, the model is based on three latent components, h_t, q_t, and k_t.

As in the SVt model, the inclusion of a jump component influences the properties of the stochastic volatility process. Large values of $|y_t|$ are now attributed rather to the the jump component than to the volatility process. As in the SVt model this typically induces a higher persistence in the volatility process.

Eraker, Johannes, and Polson (2003) estimate the number of jumps in returns

to be approximately 1.5 per year for daily S&P 500 returns from 1980 to 1999, and 4.4 per year for NASDAQ 100 index returns from 1985 to 1999. Chib et al. (2002) estimate 0.92 jumps per year for daily S&P 500 returns covering a period from 1962 to 1997.

Similarly, jump components can be also included in the volatility process in order to capture instantaneous movements in volatility. Bates (2000) and Duffie, Pan, and Singleton (2000) provide evidence that both jumps in returns and volatilities are important to appropriately capture the dynamics in financial return processes. For S&P 500 returns from 1980 to 1999, Eraker et al. (2003) estimate 1.4 volatility jumps per year.

12.2.2 The Relationship Between Volatility and Returns

Studying the relation between expected stock returns and expected variance is a fundamental topic in financial economics. Though a positive relationship between expected returns and expected variances is consistent with the notion of rational risk-averse investors requiring higher expected returns as a risk premium during volatile market periods, it is not consistently supported by empirical research. Whereas French, Schwert, and Stambaugh (1987) and Campbell and Hentschel (1992) find positive relationships between expected risk premia and conditional volatility, several other studies find converse dependencies. In fact, there is evidence that unexpected returns and innovations to the volatility process are negatively correlated. This can be explained either by the volatility feedback theory by French et al. (1987), or by the well-known leverage effect discussed by Black (1976).

In this section, we will discuss two types of SV models allowing the return and volatility process to be correlated, namely the SV-in-Mean (SVM) model and the Asymmetric SV (ASV) model. While the SVM model includes the volatility component directly in the mean equation, the ASV model allows for mutual correlations between return and volatility innovations.

The SV-in-Mean Model

The SV-in-Mean (SVM) model is given by

$$y_t = d \cdot h_t + \exp(h_t/2)u_t, \qquad u_t \sim \mathrm{N}(0,1), \tag{12.14a}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \tag{12.14b}$$

where the parameter d captures the relationship between returns and both expected as well as unexpected volatility components. This can be seen by

rewriting (12.14a) as

$$y_t = d \cdot h_{t|t-1} + d\left(h_t - h_{t|t-1}\right) + \exp(h_t/2)u_t, \tag{12.15}$$

where $h_{t|t-1}$ denotes the expected volatility defined by the conditional variance at time t given the information available at time $t-1$. Accordingly, the term $(h_t - h_{t|t-1})$ gives the innovation to the volatility process.

French et al. (1987) regress monthly excess returns of U.S. stock portfolios on both expected and unexpected volatility components stemming from ARMA models based on daily data. Excluding the unexpected volatility component results in a weakly positive relationship between excess returns and volatility. In contrast, including both volatility components does not only result in a significantly negative impact of the volatility innovation but also reverses the sign of the ex ante relationship. Hence, the negative relationship between unexpected returns and innovations to the volatility process seems to dominate the weaker, presumably positive, relation between the expected components.

The Asymmetric SV Model

Empirical evidence for 'good' and 'bad' news having different effects on the future volatility is typically referred to as the leverage or asymmetric effect. According to the leverage effect, an unexpected drop in prices ('bad' news) increases the expected volatility more than an unexpected increase ('good' news) of similar magnitude. According to Black (1976) this is due to asymmetric effects of changes of the firm's financial leverage ratio. In SV models, leverage effects are captured by allowing the observation error u_t and the future process error η_{t+1} to be correlated. Then, the ASV model is specified by

$$y_t = \exp(h_t/2)u_t, \tag{12.16a}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \tag{12.16b}$$

$$\begin{pmatrix} u_t \\ \eta_{t+1} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta \end{pmatrix} \right\}, \tag{12.16c}$$

where ρ denotes the correlation between u_t and η_{t+1}.

The ASV model has been extensively studied in the literature. Harvey and Shephard (1996) estimate the model using quasi-maximum likelihood providing $\widehat{\rho} = -0.66$ for daily U.S. stock returns ranging from 1962 to 1987. Based on the same data, Sandmann and Koopman (1998) and Jacquier, Polson, and Rossi (2004) estimate an ASV specification, where the *contemporaneous* return and volatility are correlated. Using simulated MLE methods and MCMC

based Bayesian inference, the two studies provide estimates of $\widehat{\rho} = -0.38$ and $\widehat{\rho} = -0.48$, respectively.

12.2.3 The Long Memory SV Model

In the previous sections, we have considered a first order autoregressive process for the log volatility h_t. This induces that the autocorrelations of h_t decay geometrically and volatility is said to exhibit short memory. However, empirical autocorrelations for absolute and squared returns typically decay more slowly and thus are not geometrically bounded. This implies so-called long range dependence or long memory effects. See, for example, Bollerslev and Mikkelsen (1996). One possibility to capture such effects is to allow for fractionally integrated processes, which have been developed and extensively studied over the last 25 years, see, e.g., Granger and Joyeux (1980), and Beran (1994), among others. Long memory SV models have been introduced by Breidt, Carto, and de Lima (1998), Harvey (1998), and Arteche (2004). Then, the log volatility process follows an ARFIMA(p, d, q) process given by

$$y_t = \exp(h_t/2)u_t, \qquad u_t \sim \mathrm{N}(0,1), \tag{12.17}$$

$$\phi(L)(1-L)^d(h_t - \mu) = \theta(L)\eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \tag{12.18}$$

where d denotes the fractional differencing parameter and L denotes the lag operator with

$$\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i, \quad \theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i, \tag{12.19}$$

and the roots of the polynomials $\phi(\cdot)$ and $\theta(\cdot)$ lying strictly outside the unit circle. If $d \in (-0.5, 0.5)$, the volatility process reveals long memory and is weakly stationary. The fractional differencing operator $(1-L)^d$ can be expressed in terms of the series expansion

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}(-1)^k L^k, \tag{12.20}$$

with $\Gamma(\cdot)$ denoting the gamma function (see, e.g., Beran (1994)).

The autocorrelation of $\log h_t^2$ is derived, e.g., by Baillie (1996), Breidt et al. (1998), or Harvey (1998). It is asymptotically proportional to π^{2d-1}, as long as $d \in (-0.5, 0.5)$. Similar asymptotic results are applicable to $|y_t|$ and y_t^2.

Breidt et al. (1998) estimate the Fractionally Integrated SV (FISV) model by maximizing the spectral quasi-likelihood and obtain estimates of $d = 0.44$ and

$\phi = 0.93$ for daily returns of a value-weighted market portfolio of U.S. stocks between 1962 and 1989. Gallant et al. (1997) use efficient method of moments techniques to provide estimates of d ranging between 0.48 and 0.55 for a series of daily returns from the S&P composite price index ranging from 1928 to 1987. Brockwell (2005) develops an MCMC sampling algorithm for the estimation of the FISV model and provides $d = 0.42$ for daily ASD-USD FX rates between 1999 and 2004.

12.3 MCMC-Based Bayesian Inference

In this section, we will give a brief review of MCMC-based Bayesian inference and will illustrate its application to estimate the standard SV model. For an introduction to Bayesian econometrics, see, for example, Koop (2006) and Greenberg (2008).

12.3.1 Bayes' Theorem and the MCMC Algorithm

Let θ denote a vector of model parameters including all latent variables, and let y collect the observed data. By considering θ to be a random vector, its inference is based on the posterior distribution, $p(\theta|y)$, which can be represented by Bayes' theorem

$$p(\theta|y) \propto p(y|\theta)p(\theta), \tag{12.21}$$

where $p(y|\theta)$ denotes the likelihood function depending on the model parameters and the data y. Correspondingly, $p(\theta)$ defines the prior distribution reflecting subjective prior beliefs on the distribution of θ. Consequently, the posterior distribution $p(\theta|y)$ can be viewed as a combination of objective and subjective information. If the prior is noninformative, Bayesian inference for the parameter vector θ is equivalent to likelihood-based inference.

The principle of MCMC-based Bayesian inference is to simulate $p(\theta|y)$ based on a Markov chain of random draws stemming from a family of candidate-generating densities from which it is easy to sample. Let $x \in \mathbb{R}^d$ denote a random variable (in the given context it corresponds to θ) following a Markov chain with transition kernel $p(x, y)$ corresponding to the conditional density of y given x. The invariant distribution is given by $\pi^*(y) = \int_{\mathbb{R}^d} p(x, y)\pi^*(x)dx$. An important result in Markov chain theory is that if $p(x, y)$ satisfies the reversibility condition

$$f(x)p(x, y) = f(y)p(y, x), \tag{12.22}$$

- For $g = 1, \ldots, G$:
  1. Generate Y from $q(x^{(j)}, y)$ and U from $\mathbb{U}[0,1]$.
  2. If $U \leq \alpha(x^{(j)}, Y) = \min\left\{\frac{f(Y)q(Y,x^{(j)})}{f(x^{(j)})q(x^{(j)},Y)}, 1\right\}$
     Set $x^{(j+1)} = Y$.
     Else
     Set $x^{(j+1)} = x^{(j)}$.
  3. Return $\{x^{(1)}, x^{(2)}, \ldots, x^{(G)}\}$.

Figure 12.1. The Metropolis-Hasings Sampling Algorithm

then, $f(\cdot)$ is the invariant density for the kernel $p(\cdot)$, i.e., $f(\cdot) = \pi^*(\cdot)$.

An important MCMC technique is the Metropolis-Hastings (M-H) algorithm as developed by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) and generalized by Hastings (1970). The major idea is to build on (12.22) and finding a reversible kernel whose invariant distribution equals the target distribution $f(\cdot)$. This is performed by starting with an irreversible kernel (proposal density) $q(y,x)$ for which $f(x)q(x,y) > f(y)q(y,x)$, i.e., loosely speaking, the process moves from x to y too often and from y to x too rarely. This can be corrected by introducing a probability $\alpha(x,y) < 1$ that the move is made. I.e., we choose $\alpha(x,y)$ such that

$$f(x)\alpha(x,y)q(x,y) = f(y)\alpha(y,x)q(y,x). \tag{12.23}$$

It is easily shown that this relationship is fulfilled for

$$\alpha(x,y) = \begin{cases} \min\left\{\frac{f(y)q(y,x)}{f(x)q(x,y)}, 1\right\}, & \text{if } f(x)q(x,y) \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{12.24}$$

This yields a transition kernel $q_{MH}(x,y)$ satisfying the reversibility condition and is defined by

$$q_{MH}(x,y) \stackrel{\text{def}}{=} q(x,y)\alpha(x,y), \quad x \neq y. \tag{12.25}$$

The resulting M-H sampling algorithm is summarized by Figure 12.1.

A crucial issue is an appropriate choice of the family of candidate-generating densities. Depending on the form and the complexity of the sampling problem, various techniques have been proposed in the literature. The probably most straightforward technique is proposed by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) suggesting a random walk chain, where $q(x,y) = q_0(y-x)$, and $q_0(\cdot)$ is a multivariate density. Then, y is drawn from $y = x + z$ with z following q_0. If q_0 is symmetric around zero, we

have $q(x,y) = q(y,x)$ and thus $\alpha(x,y) = f(y)/f/(x)$. A further simple choice of candidate-generating densities is proposed by Hastings (1970) and is given by $q(x,y) = q_0(y)$, i.e., y is sampled independently from x resulting in an independence chain. Then, $\alpha(x,y) = f(y)/f(x) \cdot q(x)/q(y)$. A popular and more efficient method is the acceptance-rejection (A-R) M-H sampling method which is available whenever the target density is bounded by a density from which it is easy to sample. If the target density is fully bounded, the M-H algorithm is straightforwardly combined with an acceptance-rejection step. This principle will be illustrated in more detail in the next section in order to sample the latent volatility states h_t. A more sophisticated M-H A-R algorithm which does not need a blanketing function but only a *pseudo*-dominating density is proposed by Tierney (1994).

If the dimension of x is high, the M-H algorithm is facilitated by applying it to blocks of parameters. For instance, if the target density can be expressed in terms of two blocks of variables, i.e., $f(x_1, x_2)$, the M-H algorithm allows to sample from each block x_i given the other block x_j, $j \neq i$. Then, the probability for moving from x_1 to the candidate value Y_1 given x_2 is

$$\alpha(x_1, Y_1|x_2) = \frac{f(Y_1, x_2)q_1(Y_1, x_1|x_2)}{f(x_1, x_2)q_1(x_1, Y_1|x_2)}. \tag{12.26}$$

If the kernel $q_1(x_1, Y_1|x_2)$ is the conditional distribution $f(x_1|x_2)$, then

$$\alpha(x_1, Y_1|x_2) = \frac{f(Y_1, x_2)f(x_1|x_2)}{f(x_1, x_2)f(Y_1|x_2)} = 1 \tag{12.27}$$

since $f(Y_1|x_2) = f(Y_1, x_2)/f(x_2)$ and $f(x_1|x_2) = f(x_1, x_2)/f(x_2)$. If $f(x_1|x_2)$ is available for direct sampling, the resulting algorithm is referred to as the *Gibbs sampler*, see (Geman and Geman, 1984).

Applying the M-H (or Gibbs) algorithm to sub-blocks of the vector x is a common proceeding in Bayesian statistics if the posterior distribution is of high dimension. This is particularly true for SV models where θ also includes the unobservable volatility states. In this context, the posterior distribution $p(\theta|y)$ is broken up into its complete conditional distributions $p(\theta_i|\theta_{-i}, y), i = 1, \ldots, N$, where N is the number of conditional distributions, θ_i denotes the i-th block of parameters and θ_{-i} denotes all elements of θ excluding θ_i. The theoretical justification for this proceeding is given by the theorem by Hammersley and Clifford (71) which is proven by Besag (1974). The intuition behind this theorem is that the knowledge of the complete set

of conditional posterior distributions,

$$\begin{aligned} &p(\theta_1|\theta_2,\theta_3,\ldots,\theta_k,y),\\ &p(\theta_2|\theta_1,\theta_3,\ldots,\theta_k,y),\\ &\vdots\\ &p(\theta_k|\theta_1,\theta_2,\ldots,\theta_{k-1},y), \end{aligned}$$

up to a constant of proportionality, is equivalent to the knowledge of the posterior distribution $p(\theta_1,\ldots,\theta_k|y)$. This allows applying the M-H algorithm to sub-blocks of θ leading to the Gibbs sampler if the individual conditional posterior distributions $p(\theta_i|\theta_{-i},y)$ are directly available for sampling. In practice, Gibbs and M-H algorithms are often combined resulting in "hybrid" MCMC procedures as also illustrated in the next section.

The implementation of MCMC algorithms involves two steps. In the first step, M-H algorithms generate a sequence of random variables, $\{\theta^{(i)}\}_{i=1}^{G}$, converging to the posterior distribution $p(\theta|y)$. The algorithm is applied until convergence is achieved. In practice, the convergence of the Markov chain can be checked based on trace plots, autocorrelation plots or convergence tests, such as Geweke's Z-score test, Heidelberg-Welch's stationarity test and the half-width test, see, e.g., Cowles and Carlin (1996). In the second step, Monte Carlo methods are employed to compute the posterior mean of the parameters. In particular, given the generated Markov chain, $\{\theta^{(g)}\}_{g=1}^{G}$, the population mean $\mathsf{E}[f(\theta)|y] = \int f(\theta)p(\theta|y)d\theta$ can be consistently estimated by the sample mean

$$\frac{1}{G-g_1}\sum_{g=g_1+1}^{G} f(\theta^{(g)}), \tag{12.28}$$

where g_1 is the number of burn-in periods which are discarded to reduce the influence of initial values $(\theta^{(0)})$. The length of the burn-in period typically consists of $10\% - 15\%$ of all MCMC iterations.

Consequently, the implementation of MCMC techniques requires both the convergence of the Markov chain and the convergence of the sample average. If the Markov chain is irreducible, aperiodic and positive recurrent, the Markov chain $\{\Theta^{(g)}\}_{g=1}^{G}$ generated from the MCMC algorithm converges to its invariant distribution, i.e.

$$\theta^{(g)} \xrightarrow{\mathcal{L}} \theta \quad \text{for } g \to \infty, \tag{12.29}$$

where $\theta \sim p(\theta|y)$. For more details, see, e.g.,Tierney (1994) or Greenberg (2008).

The convergence of the sample average of a function $m(\cdot)$ of $\{\Theta^{(g)}\}_{g=1}^{G}$ to its population counterpart,

$$\frac{1}{G}\sum_{g=1}^{G} m(\theta^{(g)}) \xrightarrow{\text{a.s.}} E[m(\theta)|y] \quad \text{for } G \to \infty \tag{12.30}$$

is ensured by the ergodicity of the Markov chain. As shown by Tierney (1994), the latter property is sufficient to ensure also the convergence of the Markov chain to its invariant distribution.

12.3.2 MCMC-Based Estimation of the Standard SV Model

In this section, we will illustrate the estimation of the standard SV model using the M-H algorithm. For convenience, we restate model (12.1) as given by

$$y_t = \exp(h_t/2)u_t, \qquad u_t \sim \mathrm{N}(0,1), \tag{12.31a}$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2) \tag{12.31b}$$

with $\theta = (\mu, \phi, \sigma_\eta^2)$ and $h = (h_1, \cdots, h_T)$. Applying Bayes' theorem we have

$$p(\theta, h|y) \propto p(y|\theta, h)p(h|\theta)p(\theta). \tag{12.32}$$

Bayesian inference for the model parameters θ and the volatility states h is based on the posterior distribution $p(\theta, h|y)$ which is proportional to the product of the likelihood function $p(y|\theta, h)$ specified by (12.31a), the conditional distribution of the volatility states $p(h|\theta)$ given by (12.31b), and the prior distribution $p(\theta)$.

The model is completed by specifying the prior distributions for θ. We assume that the model parameters are a priori independently distributed as follows:

$$p(\mu) = \mathrm{N}(\alpha_\mu, \beta_\mu^2), \tag{12.33a}$$
$$p(\phi) = \mathrm{N}(\alpha_\phi, \beta_\phi^2)\mathbf{1}(-1, +1)(\phi), \tag{12.33b}$$
$$p(\sigma_\eta^2) = \mathbb{IG}(\alpha_\sigma, \beta_\sigma), \tag{12.33c}$$

where $\mathbb{IG}(\cdot, \cdot)$ denotes an inverse-gamma distribution and $\mathrm{N}(a, b)\mathbf{1}(-1, +1)(x)$ defines a normal distribution with mean a, variance b, which is truncated between -1 and 1. This rules out near unit-root behavior of ϕ. The parameters $\alpha_{(\cdot)}$ and $\beta_{(\cdot)}$, characterizing the prior distributions, are called hyperparameters, which are specified by the researcher.

- Initialize $h^{(0)}, \mu^{(0)}, \phi^{(0)}$ and $\sigma_\eta^{2(0)}$.
- For $g = 1, \ldots, G$:
  1. For $t = 1, \ldots, T$:
     Sample $h_t^{(g)}$ from $p(h_t|y, h_{<t}^{(g)}, h_{>t}^{(g-1)}, \mu^{(g-1)}, \phi^{(g-1)}, \sigma_\eta^{2(g-1)})$.
  2. Sample $\sigma_\eta^{2(g)}$ from $p(\sigma_\eta^2|y, h^{(g)}, \mu^{(g-1)}, \phi^{(g-1)})$.
  3. Sample $\phi^{(g)}$ from $p(\phi|y, h^{(g)}, \sigma_\eta^{2(g)}, \mu^{(g-1)})$.
  4. Sample $\mu^{(g)}$ from $p(\mu|y, h^{(g)}, \phi^{(g)}, \sigma_\eta^{2(g)})$.

Figure 12.2. Single-move Gibbs sampler for the standard SV model

Given the prior distributions, the conditional posteriors for the model parameters are derived as

$$p(\mu|y, h, \phi, \sigma_\eta^2) \propto p(y|h, \mu, \phi, \sigma_\eta^2)p(h|\mu, \phi, \sigma_\eta^2)p(\mu), \tag{12.34a}$$
$$p(\phi|y, h, \sigma_\eta^2, \mu) \propto p(y|h, \mu, \phi, \sigma_\eta^2)p(h|\mu, \phi, \sigma_\eta^2)p(\phi), \tag{12.34b}$$
$$p(\sigma_\eta^2|y, h, \mu, \phi) \propto p(y|h, \mu, \phi, \sigma_\eta^2)p(h|\mu, \phi, \sigma_\eta^2)p(\sigma_\eta^2). \tag{12.34c}$$

Since the volatility states h subsume all information about $(\mu, \phi, \sigma_\eta^2)$, the full information likelihood function $p(y|h, \mu, \phi, \sigma_\eta^2)$ is a constant with respect to the model parameters, and thus can be omitted.

By successively conditioning we get

$$p(h|\mu, \phi, \sigma_\eta^2) = p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2), \tag{12.35}$$

where $p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2)$ is specified according to (12.31b). Moreover, inserting $p(\sigma_\eta^2)$, $p(\phi)$, $p(\mu)$, given by (12.33), and $p(h|\mu, \phi, \sigma_\eta^2)$, given by (12.35), into (12.34), the full conditional posteriors can be reformulated, after eliminating constant terms, as (for details, see Appendix 12.5.1)

$$p(\sigma_\eta^2|y, h, \mu, \phi) \propto \mathbb{IG}(\widehat{\alpha}_\sigma, \widehat{\beta}_\sigma), \tag{12.36}$$
$$p(\phi|y, h, \sigma_\eta^2, \mu) \propto \mathrm{N}(\widehat{\alpha}_\phi, \widehat{\beta}_\phi^2)\mathbf{1}(-1, +1)(\phi), \tag{12.37}$$
$$p(\mu|y, h, \phi, \sigma_\eta^2) \propto \mathrm{N}(\widehat{\alpha}_\mu, \widehat{\beta}_\mu^2), \tag{12.38}$$

where the hyper-parameters are estimated by

$$\widehat{\alpha}_\sigma = \alpha_\sigma + \frac{T}{2}, \tag{12.39}$$

$$\widehat{\beta}_\sigma = \beta_\sigma + \frac{1}{2}\left\{\sum_{t=1}^{T-1}(h_{t+1} - \mu - \phi(h_t - \mu))^2 + (h_1 - \mu)^2(1-\phi^2)\right\}, \tag{12.40}$$

$$\widehat{\alpha}_\phi = \widehat{\beta}_\phi^2 \left\{\frac{\sum_{t=1}^{T-1}(h_{t+1} - \mu)(h_t - \mu)}{\sigma_\eta^2} + \frac{\alpha_\phi}{\beta_\phi^2}\right\}, \tag{12.41}$$

$$\widehat{\beta}_\phi^2 = \left\{\frac{\sum_{t=1}^{T-1}(h_t - \mu)^2 - (h_1 - \mu)^2}{\sigma_\eta^2} + \frac{1}{\beta_\phi^2}\right\}^{-1}, \tag{12.42}$$

$$\widehat{\alpha}_\mu = \widehat{\beta}_\mu^2 \left\{\frac{h_1(1-\phi^2) + (1-\phi)\sum_{t=1}^{T-1}(h_{t+1} - \phi h_t)}{\sigma_\eta^2} + \frac{\alpha_\mu}{\beta_\mu^2}\right\}, \tag{12.43}$$

$$\widehat{\beta}_\mu^2 = \left\{\frac{1 - \phi^2 + (T-1)(1-\phi)^2}{\sigma_\eta^2} + \frac{1}{\beta_\mu^2}\right\}^{-1}. \tag{12.44}$$

Since it is possible to directly sample from the conditional posteriors, we obtain a straightforward (single-move) Gibbs sampler which breaks the joint posterior $p(\theta, h, y)$ into $T+3$ univariate conditional posteriors. The resulting Gibbs algorithm is summarized in Figure 12.2, where the subscripts of $h_{<t}^{(\cdot)}$ and $h_{>t}^{(\cdot)}$ denote the periods before and after t respectively.

The most difficult part of the estimation of SV models is to effectively sample the latent states h_t from their full conditional posterior. In this context, an M-H A-R algorithm can be applied. Below we briefly illustrate a sampling procedure which is also used by Kim et al. (1998). In this context, Bayes' theorem implies

$$p(h_t|y, h_{-t}, \theta) \propto p(y_t|h_t, \theta)p(h_t|h_{-t}, \theta), \tag{12.45}$$

$$= \frac{1}{\sqrt{2\pi\exp(h_t)}}\exp\left\{-\frac{y_t^2}{2\exp(h_t)}\right\}p(h_t|h_{-t}, \theta), \tag{12.46}$$

$$= f^*(y_t, h_t, \theta)p(h_t|h_{-t}, \theta), \tag{12.47}$$

where, h_{-t} denotes all elements of $h = (h_1, \cdots, h_T)$ excluding h_t. Exploiting the Markovian structure of the SV model we can derive

$$p(h_t|h_{-t}, \theta) = p(h_t|h_{t-1}, h_{t+1}, \theta) = p_N(h_t|\alpha_t, \beta^2), \tag{12.48}$$

where, $p_N(x|a, b)$ denotes the normal density function with mean a and vari-

ance b, and

$$\alpha_t = \mu + \frac{\phi\{(h_{t-1} - \mu) + (h_{t+1} - \mu)\}}{(1 + \phi^2)}, \quad \beta^2 = \frac{\sigma_\eta^2}{1 + \phi^2}. \tag{12.49}$$

An acceptance-rejection step is implemented exploiting the fact that $\exp(-h_t)$ is bounded by a linear function in h_t. By applying a Taylor expansion for $\exp(-h_t)$ around α_t we obtain

$$\log f^*(y_t, h_t, \theta) \leq -\frac{1}{2}\log(2\pi) - \frac{1}{2}h_t - \frac{y_t^2}{2}[\exp(-\alpha_t)\{1 + \alpha_t - h_t \exp(-\alpha_t)\}] \tag{12.50}$$

$$\stackrel{\text{def}}{=} \log g^*(y_t, h_t, \theta). \tag{12.51}$$

Since $p(h_t|h_{-t}, \theta) = p_N(h_t|\alpha_t, \beta^2)$, we have

$$p(h_t|h_{-t}, \theta) f^*(y_t, h_t, \theta) \leq p_N(h_t|\alpha_t, \beta^2) g^*(y_t, h_t, \theta). \tag{12.52}$$

Then, the right-hand side of (12.52), after eliminating constant terms, can be represented by

$$p_N(h_t|\alpha_t, \beta^2) g^*(y_t, h_t, \theta) = k \cdot p_N(h_t|\alpha_t^*, \beta^2), \tag{12.53}$$

where k is a real valued constant, and $p_N(h_t|\alpha_t^*, \beta^2)$ denotes a normal density with mean $\alpha_t^* = \alpha_t + \frac{\beta^2}{2}(y_t^2 \exp\{-\alpha_t\} - 1)$ and variance β^2.

Hence, since the target distribution, $p(h_t|h_{-t}, \theta) f^*(y_t, h_t, \theta)$, is bounded by $p_N(h_t|\alpha_t^*, \beta^2)$ up to a constant k, the acceptance-rejection method can be applied to sample h_t from $p(h_t|y, h_{-t}, \theta)$ with acceptance probability

$$\mathrm{P}\left\{U \leq \frac{f^{**}(y_t, h_t, \theta) p(h_t|h_{-t}, \theta)}{k p_N(h_t|\alpha_t^*, \beta^2)}\right\} = \frac{f^{**}(y_t, h_t, \theta)}{g^{**}(y_t, h_t, \theta)}$$

where $U \sim \mathbb{U}[0, 1]$. Figure 12.3 summarizes the A-R algorithm to sample the latent volatility states h_t.

12.4 Empirical Illustrations

12.4.1 The Data

Below we will illustrate estimations of the standard SV model, the SVt model and the SVJ model based on time series of the DAX index, the Dow Jones

- For $t = 1, \cdots, T$:
  1. Draw h_t^* from $p_N(h_t|\alpha_t^*, \beta^2)$.
  2. Draw U from $\mathbb{U}[0, 1]$.
  3. If $U \leq f^*(y_t, h_t^*, \theta)/g^*(y_t, h_t^*, \theta)$
     set $h_t = h_t^*$.
     Else
     go to step 1.

Figure 12.3. A-R method to sample the volatility states h_t

	Mean	SD	Median	0.1-q	0.9-q	Skewness	Kurtosis
DAX	3.7e-04	0.013	5.0e-4	-0.021	0.006	-0.295	7.455
Dow Jones	3.6e-04	0.009	3.0e-4	-0.009	0.008	-0.230	8.276
GBP/USD	3.6e-06	0.005	<1.0e-9	-0.006	0.009	-0.126	5.559

Table 12.1. Summary statistics for daily returns of the DAX index, the Dow Jones index, and the GBP/USD exchange rate from 01/01/1991 to 21/03/2007. `XFGsummary`

index and the GBP/USD FX rate. All time series cover the period from 1 January, 1991 to 21 March, 2007. We use daily continuously compounded returns yielding 4,231 observations. Table 12.1 reports the mean, standard deviation, median, 10%- and 90%-quantiles, and the empirical skewness as well as kurtosis of the three series. All series reveal negative skewness and overkurtosis which is a common finding for financial returns.

12.4.2 Estimation of SV Models

The standard SV model is estimated by running the Gibbs and A-R M-H algorithm based on 25,000 MCMC iterations, where $5,000$ iterations are used as burn-in period. Table 12.2 displays the choice of the prior distributions and the hyper-parameters as well as the resulting prior mean and standard deviation.

Table 12.3 shows the sample mean (MEAN), the sample standard deviation (SD), the time-series standard errors (ts-SE), and the 95%-credibility interval (CI) based on $G = 20,000$ MCMC replications. The time-series standard errors give an estimate of the variation that is expected in computing the

Prior Distribution	Hyper-Parameters		Mean	S.D.
$p(\mu) = \mathrm{N}(\alpha_\mu, \beta_\mu^2)$	$\alpha_\mu = 0$	$\beta_\mu = 100$	0	10
$p(\phi) = \mathrm{N}(\alpha_\phi, \beta_\phi^2) I_{(-1,+1)}(\phi)$	$\alpha_\mu = 0$	$\beta_\mu = 100$	0	1
$p(\sigma_\eta^2) = \mathcal{IG}(\alpha_\sigma, \beta_\sigma)$	$\alpha_\sigma = 2.5$	$\beta_\sigma = 0.025$	0.167	0.024

Table 12.2. Prior distributions, hyper-parameters, and implied prior means as well as standard deviations for the standard SV model. XFGprior

Parameter	Mean	SD	ts-SE	95% CI
DAX				
μ	-8.942	0.192	1.5e-3	(-9.327,-8.565)
ϕ	0.989	0.002	2.0e-4	(0.983, 0.994)
σ_η	0.115	0.009	1.0e-3	(0.096, 0.137)
Dow Jones				
μ	-9.471	0.171	1.3e-3	(-9.810,-9.142)
ϕ	0.990	0.003	2.0e-4	(0.984, 0.995)
σ_η	0.087	0.010	1.1e-3	(0.069, 0.108)
GBP/USD				
μ	-10.238	0.649	4.3e-3	(-10.519,-9.997)
ϕ	0.993	0.002	2.0e-4	(0.988, 0.997)
σ_η	0.041	0.006	8.0e-4	(0.029, 0.054)

Table 12.3. Estimation results for the standard SV model. XFGparameter

mean of the MC replications and is computed as $SD/\sqrt{n}$. As a rule of thumb, Geweke (1992) suggests to choose G such that the time series standard error is less than approximately 5% of the sample standard deviation.

Since the three time series reveal similar properties, we concentrate on the results for DAX index returns. The volatility process is highly persistent as indicated by an estimate of ϕ of 0.989. This near-to-unit-root behavior is a quite typical finding for financial return series and is consistent with the commonly observed volatility clustering. The estimated (smoothed) volatility

states are computed by

$$\widehat{h}_t = \frac{1}{G - g_1} \sum_{g=g_1+1}^{G} \exp(h_t^{(g)}/2), \tag{12.54}$$

where $h_t^{(g)}$ denotes the realizations of the Markov chain stemming from the M-H A-R algorithm illustrated in the previous section, and g_1 is the burn-in period. The resulting plots of the smoothed volatilities are shown in Figure 12.4. It is nicely illustrated that the estimated latent volatility closely mimics the movements of $|y_t|$ supporting the idea of using absolute or squared returns as (noisy) proxies for h_t.

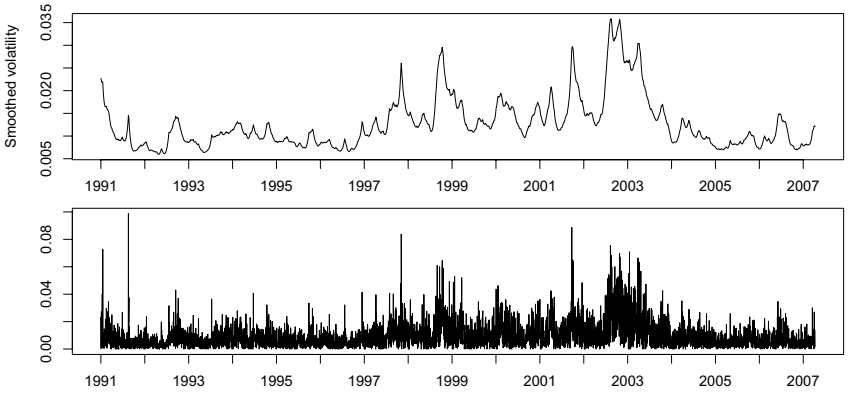

Figure 12.4. Top: Smoothed estimates of h_t. Bottom: Absolute returns, $|y_t|$. XFGvolabs

Misspecification tests are implemented based on the standardized innovations, $y_t \exp(-\widehat{h}_t/2)$ which should be i.i.d. Applying Ljung-Box tests and ARCH tests (Engle, 1982) shown in Figure 12.5 yield p-values of 0.094 and 0.023, respectively. For the BDS independence test we find a p-value of 0.011. The corresponding plot of the standardized innovations as well as ACF plots of standardized innovations and squared standardized innovations are given by graphs (a), (c) and (d), respectively, in Figure 12.5. The standardized innovations reveal a big outlier on 19/08/1991 where the DAX index dropped from 1653.33 to 1497.93. Such a behavior is not easily captured by a continuous distribution for h_t and requires accounting for jumps. Nevertheless,

though it is evident that the model is obviously not flexible enough to completely explain the volatility dynamics, the diagnostics indicate a quite satisfying dynamic performance. This is particularly true when the parameter parsimony of the model is taken into account.

It is not surprising that the model is unable to capture the distributional properties of the returns. We observe that the standard SV model with a model implied kurtosis of 5.74 is not able to fully explain the over-kurtosis in the data. This is confirmed by the Jarque-Bera normality test and the QQ plot revealing departures from normality mainly stemming from extreme innovations.

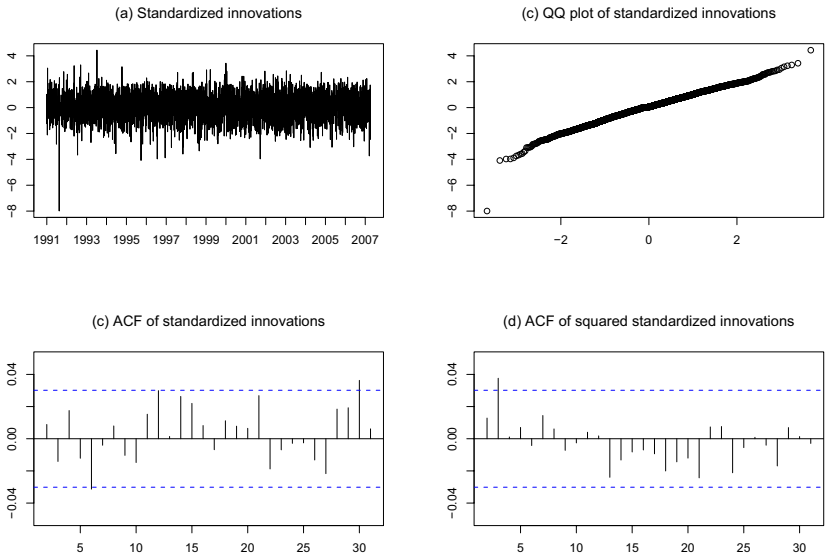

Figure 12.5. Time series plot, QQ plot and autocorrelogram of (squared) standardized innovations. XFGstdinnov

Finally, the results of convergence diagnostics are reported in Table 12.4. All parameters pass both the Geweke's z-scores test and the Heidelberg-Welch's stationarity and half-width tests indicating a proper convergence of the Markov chain to its invariant distribution.

Table 12.5 shows the estimation results based on the SVt and SVJ model. For the sake of brevity and given that we have qualitatively similar findings

Parameter	Z-score Test		Stationarity and Half-Width Test			
	z-score	p-value	p-value	Mean	Half-width	Ratio
μ^{SV}	0.199	0.843	0.645	-8.895	0.003	-0.001
ϕ^{SV}	0.032	0.972	0.897	0.928	0.001	0.001
σ_u^{SV}	-0.413	0.686	0.979	0.329	0.003	0.009

Table 12.4. Convergence Diagnostics. The half-width test is passed if the corresponding ratio is less than 0.01. XFGconvergence

Parameter	Mean	SD	ts-SE	95% CI
The SVt model:				
μ	-9.201	0.230	2.3e-3	(-9.663,-8.752)
ϕ	0.991	0.002	1.0e-4	(0.985, 0.995)
σ_η	0.117	0.012	1.1e-3	(0.095, 0.145)
ν	12.443	1.812	2.3e-1	(9.600,16.923)
The SVJ model:				
μ	-9.107	2.3e-01	1.8e-03	(-9.568,-8.663)
ϕ	0.991	2.7e-03	2.1e-04	(0.984, 0.995)
σ_η	0.124	1.3e-02	1.4e-03	(0.101, 0.153)
α_k	-0.005	2.9e-05	1.8e-07	(-0.005,-0.004)
$\sqrt{\beta_k}$	0.029	6.5e-03	8.7e-04	(0.020, 0.045)
κ	0.010	3.9e-03	3.1e-04	(0.003, 0.019)

Table 12.5. Estimation results for the SVt and SVJ model based on DAX index returns. XFGsvtjparameter

for the other return series, we focus only on DAX index returns. We obtain an estimate of the degrees of freedom in the SVt model of about $\widehat{\nu} = 12.44$ indicating the presence of fat-tailedness in the data and a clear misspecification of the standard (Gaussian) SV model. The estimates for the SVJ model reveal a daily average jump size of about $\widehat{\alpha}_k = 0.005\%$ with estimated standard deviation $\sqrt{\widehat{\beta}_k} = 0.029$. Estimates of κ reveal an average probability of observing a jump of about 1% on a daily basis. This implies that on average a jump in returns may occur on average every 100 trading days.

Figure 12.6 depicts the QQ plots of the normalized innovations based on the standard SV model (left), the SVt model (middle), and the SVJ model (right).

It is shown that the inclusion of Student-t errors improves the distributional properties of the model only slightly. Actually, we observe that both the basic SV and the SVt model are not able to capture extreme observations in the tails of the distribution. In contrast, the SVJ model turns out to be more appropriate to accommodate outliers. This result indicates the importance of allowing returns to be driven by a jump component.

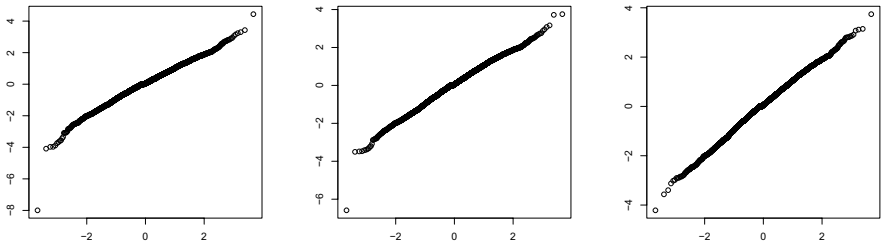

Figure 12.6. QQ plots of normalized innovations based on the standard SV model (left), the SVt model (middle), and the SVJ model (right). XFGsvtsvjqq

12.5 Appendix

12.5.1 Derivation of the Conditional Posterior Distributions

Using Bayes' theorem, the conditional posterior distribution of σ_η^2 is given by

$$p(\sigma_\eta^2|y,h,\mu,\phi) \propto p(y|h,\mu,\phi,\sigma_\eta^2)p(h|\mu,\phi,\sigma_\eta^2)p(\sigma_\eta^2).$$

By assuming σ_η^2 to follow an inverse-gamma distribution and successively conditioning on $p(h|\mu,\phi,\sigma_\eta^2)$, we obtain

$$p(\sigma_\eta^2|y,h,\mu,\phi) \propto p(h_1|\mu,\phi,\sigma_\eta^2)\prod_{t=1}^{T-1} p(h_{t+1}|h_t,\mu,\phi,\sigma_\eta^2)\mathbb{IG}(\sigma_\eta^2|\alpha_\sigma,\beta_\sigma),$$

where the density function $p(h_{t+1}|h_t,\mu,\phi,\sigma_\eta^2)$ is given by (12.1b).

After eliminating all constant terms with respect to σ_η^2, we obtain

$$
\begin{aligned}
&p(\sigma_\eta^2|y,h,\mu,\phi) \\
&\propto \exp\left[-\frac{(h_1-\mu)^2(1-\phi^2)}{2\sigma_\eta^2} - \frac{\sum_{t=1}^{T-1}\{h_{t+1}-\mu-\phi(h_t-\mu)\}^2}{2\sigma_\eta^2}\right] \\
&\quad\times \left(\frac{1}{\sigma_\eta^2}\right)^{\frac{T}{2}} \frac{(\beta_\sigma)^{\alpha_\sigma} e^{-\beta_\sigma/\sigma_\eta^2}}{\Gamma(\alpha_\sigma)(\sigma_\eta^2)^{\alpha_\sigma+1}} \\
&\propto \exp\left[-\frac{\beta_\sigma + \frac{1}{2}(h_0-\mu)^2(1-\phi^2) + \frac{1}{2}\sum_{t=1}^{T-1}\{h_{t+1}-\mu-\phi(h_t-\mu)\}^2}{\sigma_\eta^2}\right] \\
&\quad\times \left(\frac{1}{\sigma_\eta^2}\right)^{(\alpha_\sigma+\frac{T}{2})+1}.
\end{aligned}
$$

It is easy to see that the posterior density $p(\sigma_\eta^2|y,h,\mu,\phi)$ is proportional to an inverse-gamma density. Consequently, we have

$$p(\sigma_\eta^2|y,h,\mu,\phi) \propto \mathbb{IG}(\widehat{\alpha}_\sigma, \widehat{\beta}_\sigma),$$

where,

$$
\begin{aligned}
\widehat{\alpha}_\sigma &= \alpha_\sigma + \frac{T}{2}, \\
\widehat{\beta}_\sigma &= \beta_\sigma + \frac{1}{2}(h_1-\mu)^2(1-\phi^2) + \frac{1}{2}\sum_{t=1}^{T-1}\{h_{t+1}-\mu-\phi(h_t-\mu)\}^2.
\end{aligned}
$$

Mimicking the proceeding for σ_η^2 we can derive the conditional posteriors for μ and ϕ in a similar way. Then, we obtain

$$
\begin{aligned}
p(\mu|y,h,\phi,\sigma_\eta^2) &\propto p(h|\mu,\phi,\sigma_\eta^2)p(\mu), \\
&\propto p(h_1|\mu,\phi,\sigma_\eta^2)\prod_{t=1}^{T-1} p(h_{t+1}|h_t,\mu,\phi,\sigma_\eta^2)\mathrm{N}(\alpha_\mu,\beta_\mu), \\
&\propto \exp\Bigg(-\frac{1}{2}\Bigg[\mu^2\Bigg\{\underbrace{\frac{1-\phi^2+(T-1)(1-\phi)^2}{\sigma_\eta^2}+\frac{1}{\beta_\mu^2}}_{A}\Bigg\} \\
&\quad - 2\mu\Bigg\{\underbrace{\frac{h_1(1-\phi^2)+(1-\phi)\sum_{t=1}^{T-1}(h_{t+1}-\phi h_t)}{\sigma_\eta^2}+\frac{\alpha_\mu}{\beta_\mu^2}}_{B}\Bigg\}\Bigg]\Bigg), \\
&\propto \mathrm{N}\left(\frac{B}{A},\frac{1}{A}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
p(\phi|y,h,\sigma_\eta^2,\mu) &\propto p(h|\mu,\phi,\sigma_\eta^2)p(\phi),\\
&\propto p(h_1|\mu,\phi,\sigma_\eta^2)\prod_{t=1}^{T-1}p(h_{t+1}|h_t,\mu,\phi,\sigma_\eta^2)\mathrm{N}(\alpha_\phi,\beta_\phi^2)\infty_{(-1,+1)}(\phi),\\
&\propto \exp\Bigg(-\frac{1}{2}\Bigg[\phi^2\Bigg\{\underbrace{\frac{-(h_1-\mu)^2+\sum_{t=1}^{T-1}(h_t-\mu)^2}{\sigma_\eta^2}+\frac{1}{\beta_\phi^2}}_{C}\Bigg\}\\
&\qquad -2\phi\Bigg\{\underbrace{\frac{\sum_{t=1}^{T-1}(h_{t+1}-\mu)(h_t-\mu)}{\sigma_\eta^2}+\frac{\alpha_\phi}{\beta_\phi^2}}_{D}\Bigg\}\Bigg]\Bigg)\infty_{(-1,+1)}(\phi),\\
&\propto \mathrm{N}\left(\frac{D}{C},\frac{1}{C}\right)\infty_{(-1,+1)}(\phi).
\end{aligned}
$$

Bibliography

Andersen, T. G., Bollerslev, T., Diebold, F.X., and Labys, P. (2003). Modeling and Forecasting Realized Volatility, *Econometrica* **71**: 579–625.

Arteche, J. (2004). Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models, *Journal of Econometrics* **119**: 131–154.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics, *Journal of Econometrics* **73**: 5–59.

Bates, D. (2000). Post-'87 Crash fears in S&P 500 futures options, *Journal of Econometrics* **9**: 69–107.

Beran, J. (1994). *Statistics for Long-Memory Processes*, Chapman and Hall, Boca Raton.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society* **Series B 36**: 192–236.

Black, F. (1976). Studies of stock price volatility changes, *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section* 177–181.

Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* **73**: 151–184.

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

Brockwell, A. E. (2005). Likelihood-based analysis of a class generalized long-memory time series models, *Journal of Time Series Analysis* **28**: 386–407.

Bauwens, L., Lubrano, M., and Richard, J. F. (1999). *Bayesan Inference in Dynamic Econometric Models*, Oxford University Press & Sons, New York.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics* **31**: 281–318.

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

Cowles, M. K. and Carlin, B. P. (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review, *Journal of the American Statistical Association* **91**: 883–905.

Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative prices,", *Econometrica* **41**: 135–156.

Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood, *Journal of Econometrics* **64**: 375–400.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* **68**: 1343–1376.

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

Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns, *Journal of Finance* **58**: 1269–1300.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility, *Journal of Financial Economics* **19**: 3–29.

Frieze, A., Kannan, R., and Polson N. (1994). Sampling from log-concave distributions, *Annals of Applied Probability* **4**: 812–834.

Gallant, A.R., Hsie, D., and Tauchen G. (1997). Estimation of stochastic volatility models with diagnostics, *Journal of Econometrics* **81**: 159–192.

Geman, S. and Geman D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (ed.), *Bayesian Statistics 4*, Oxford University Press, Amsterdam.

Ghysels, E., Harvey, A. C., and Renault E. (1996). Stochastic volatility, *in* G. S. Maddala and C. R. Rao (ed.), *Handbook of Statistics 14, Statistical Methods in Finance*, North-Holland, Amsterdam.

Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* **1**: 15–29.

Greenberg, E. (2008). *Introduction to Bayesian Econometrics*, Cambridge University Press, New York.

Hammersley, J. and Clifford, P. (1971). Markov fields on finite graphs and lattices, *Unpublished Manuscipt*.

Harvey, A. C. (1998). Long-memory in stochastic volatility, *in* J. Knight and S. E. Satchell (ed.), *Forecasting Volatility in Financial Markets*, Butterworth-Heinemann, London, pp. 203–226.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models, *Review of Economic Studies* **61**: 247–264.

Harvey, A. C. and Shephard, N. (1996). The estimation of an asymmetric stochastic volatility model for asset returns, *Journal of Business and Economics Statistics* **14**: 429–434.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**: 97–109.

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

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.

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

Koop, G. (2006). *Bayesian Econometrics*, Wiley-Interscience, London.

Liesenfeld, R. and Richard, J. F. (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics, *Journal of Empirical Finance* **10**: 505–531.

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

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1092.

Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* **45**: 239–265.

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

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices 1961-79, *in* O. D. Anderson (ed.), *Time Series Analysis: Theory and Practice 1*, North-Holland, Amsterdam, pp. 203–226.

Tierney, L. (1994). Markov Chains for exploring posterior distributions, *The Annals of Statistics* **22**: 1701–1762.