

TI 2011-049/4
Tinbergen Institute Discussion Paper



Relating Stochastic Volatility Estimation Methods

Charles S. Bos

*Dept. of Econometrics and OR, VU University of Amsterdam, and Tinbergen
Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Relating Stochastic Volatility Estimation Methods

Charles S. Bos*

Tinbergen Institute and VU University Amsterdam

March 2, 2011

Abstract

Estimation of the volatility of time series has taken off since the introduction of the GARCH and stochastic volatility models. While variants of the GARCH model are applied in scores of articles, use of the stochastic volatility model is less widespread. In this article it is argued that one reason for this difference is the relative difficulty of estimating the unobserved stochastic volatility, and the varying approaches that have been taken for such estimation.

In order to simplify the comprehension of these estimation methods, the main methods for estimating stochastic volatility are discussed, with focus on their commonalities. In this manner, the advantages of each method are investigated, resulting in a comparison of the methods for their efficiency, difficulty-of-implementation, and precision.

Keywords: Stochastic volatility, estimation, methodology.

JEL classification: C13, C18, C51.

1 Introduction

Interest in the volatility of returns dates back further than our capabilities of estimating such volatility sequences in a sensible manner. E.g. Markowitz (1952), discussing the possibility of constructing a portfolio with an optimal expected return-variance trade-off, indicates that

One suggestion as to tentative μ_i, σ_{ij} is to use the observed μ_i, σ_{ij} for some period of the past. I believe that better methods, which take into account more information, can be found. I believe that what is needed is essentially a “probabilistic” reformulation of security analysis.

In his article he indeed explicitly claims to start from a set of beliefs on expected returns and volatility μ, σ , and does not consider the stage of forming the relevant beliefs on these moments on the basis of observations.

Part I of the present volume is largely directed at non-probabilistic measures of volatility like the (G)ARCH class of models, where volatility depends in a deterministic way on past returns. These models have taken a large flight after the seminal publications of Engle (1982)

*Corresponding author: C.S. Bos, Department of Econometrics & O.R., VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email cbos@feweb.vu.nl, tel/fax: +31 (0)20 598 60 23/+31 (0)20 598 60 20.

and Bollerslev (1986). The article of Clark (1973) on the other hand is generally recognised as the first to explicitly refer to a probabilistic approach to variance estimation. In the article, log-prices are driven by two independent processes, one for the levels and one for the volatility. In discrete time, Taylor (1982) described a version of the product process, which later became known as the stochastic volatility (SV) model.

Taylor (1982) extracts the volatility of this product process by computing an exponentially weighted moving average of scaled absolute returns, where the weights are calibrated. Only a decade later a series of more efficient estimation methods are devised. Even though Ghysels et al. (1996) state that these estimation methods have made the SV model an attractive alternative to models such as GARCH, one cannot escape the impression that this attraction has not come to full fruition. While there are literally thousands of applications of GARCH, for SV this number is far lower.

Two reasons for this relative lack of applied work using SV are apparent. First, there are as of yet no standard packages for estimating SV models, whereas for GARCH most statistical packages have a wealth of options for incorporating GARCH effects. A second difference seems to be that GARCH has many variants of the model (Bollerslev, 2008), with basically a single estimation method for all of them. For SV, there are few variants of the model, but a full series of estimation methods. No estimation method has come out as clearly the simplest or most efficient, and the relations between the methods are not immediately clear.

Several reviews of the SV literature are available (see Ghysels et al., 1996; Broto and Ruiz, 2004; Shephard, 2005, to name a few). Instead of giving another review of the literature, this chapter focuses instead on the estimation of the basic univariate SV model, interrelating in detail the various estimation methods. In order to get a clear comparison and better understanding, initially only the most basic SV model is investigated here. Not all methods are equally adequate for incorporating extensions, and these extensions would hamper the understanding of the basics of estimating SV models. The methods covered are chosen to give a broad overview of the possible approaches for estimation, while still being relevant in that the methods have not been clearly surpassed by alternative estimation procedures.

Table 1: Methods and main references

Acronym	Central reference	Approach	State
QML	Harvey et al. (1994)	Classical	f/s
GMS	Kim et al. (1998)	Bayesian	s
SMM	Gallant and Tauchen (1996)	Classical	-
IS	Durbin and Koopman (1997)	Classical	s
EIS	Richard and Zhang (2007)	Classical	s
IIS	Nguyen (2007)	Classical	s
SSS	Carter and Kohn (1994)	Bayesian	s
MMS	Shephard and Pitt (1997)	Bayesian	s

Section 2 first introduces the basic SV model. It continues, using common notation, with a description of a series of estimation methods for the above model. First, the quasi maximum likelihood method (Harvey et al., 1994, QML) is introduced in Section 2.1, as it uses a linear approximation which is in essence common to the other methods as well. This method is only a *quasi* maximum likelihood approach, as it does not adapt for the approximation error. The Gaussian mixture sampling (see Carter and Kohn, 1997; Kim et al., 1998, GMS), in Section

2.2, uses a Bayesian approach in which a mixture of normals decreases the approximation error. It is followed by the simulated method of moments (Section 2.3, see Gallant and Tauchen, 1996, among others, SMM), which provides a classical approach for extracting only the parameter estimates. Importance sampling for SV, as proposed by Durbin and Koopman (1997, IS), can be seen as an improvement on QML. It is however not fully efficient, and hence gave rise to both efficient importance sampling (Richard and Zhang, 2007, EIS) and improved importance sampling (Nguyen, 2007, IIS). These three methods are combined in Section 2.4. Lastly, the importance sampling approach can also be turned into a Bayesian sampling scheme as described in Section 2.5. Sampling either a single state at a time (Carter and Kohn, 1994, SSS) or using a multimove sampler (Shephard and Pitt, 1997, MMS) could provide more flexibility and/or efficiency. This set of methods is summarised in Table 1, indicating for each method the acronym used, one main reference, whether the method is classical or Bayesian in background, and whether the method allows for extracting either the filtered (f) or smoothed (s) state.

The theory can be interesting by itself, though ultimately the methods should be compared in their efficiency and ease-of-use in obtaining the estimates. Section 3 provides the results on a simulation exercise, where each method is used for estimating the volatility model on the same 1,000 simulated data series. Results are presented on different aspects of efficiency. It is followed by Section 4, which looks at practical results on two financial returns series. These two return series are estimated first univariately, with the basic SV model. Afterwards a simple extension to a bivariate volatility model with stochastic correlation is introduced and estimated. A discussion of the methods with concluding remarks is found in Section 5.

2 Theory and methodology

The notation used for the simple univariate stochastic volatility model is the following. Given a series of returns $Y \equiv \{y_1, \dots, y_T\}$ with expectation zero,¹ the model is given by

$$\begin{aligned}
 y_t &= \epsilon_t, & \epsilon_t &\sim \mathcal{N}(0, \sigma_\epsilon^2 \exp(h_t)), & (1) \\
 h_{t+1} &= \phi h_t + \xi_t, & \xi_t &\sim \mathcal{N}(0, \sigma_\xi^2), & (2) \\
 h_1 &\equiv 0, & t &= 1, \dots, T. & (3)
 \end{aligned}$$

Here $H = \{h_1, \dots, h_T\}$ is the underlying stochastic volatility, and $\theta = (\sigma_\epsilon, \phi, \sigma_\xi)$ the vector of parameters. Interest is in estimating the parameters in the model, and preferably also extracting the volatility sequence, as this is useful for evaluating the uncertainty pertaining to the returns.

The difficulty in estimating the SV model lies in the fact that the observation equation (1) depends non-linearly on the states h_t . These states evolve according to the state equation (2). As the dependence in (1) is non-linear, the Kalman filter (KF, see Harvey, 1989) equations cannot be used directly for obtaining the likelihood of the observations, or to extract filtered states. The estimation methods indicated in Table 1 each take their individual approach in solving this problem. However, those approaches are necessarily interrelated as the final model to estimate is the same. A commonality shared between most of the methods is the

¹The model is kept as simple as possible here on purpose, for ease of exposition. Extensions to include either autocorrelation in the returns, or non-Gaussianity, could be taken up into the specification as well.

use of an approximating linear Gaussian state space model, approximating (1) by

$$c_t = h_t + u_t, \quad u_t \sim \mathcal{N}(0, d_t), \quad (4)$$

for certain choices of c_t, d_t . In the remainder of this section, the relations between the methods, and their choice for the auxiliary data $y_t^* = (c_t, d_t)$ of the approximating linear model, are exposed.

2.1 Quasi maximum likelihood estimation

The most direct approach for estimating SV is found in Harvey et al. (1994). As the non-linearity of the observation equation is the main source of difficulty, the authors linearise the equation by writing

$$\ln y_t^2 = h_t + \ln \epsilon_t^2 = h_t + \ln \sigma_\epsilon^2 + \ln e_t^2, \quad e_t \sim \mathcal{N}(0, 1).$$

The density of $\ln e_t^2$ is approximated by a normal density with expectation $\mathbf{E}(\ln e_t^2) = \left(\frac{\Gamma'(\frac{1}{2})}{\Gamma(\frac{1}{2})} - \ln(\frac{1}{2}) \right) \approx -1.2704$ and variance $\mathbf{var}(\ln e_t^2) = \left. \frac{\partial^2 \ln \Gamma(\nu)}{\partial \nu^2} \right|_{\nu=\frac{1}{2}} = \frac{1}{2}\pi^2$ (see Abramowitz and Stegun, 1972, eq. 26.4.36). In terms of the linear approximating density (4) this gives parameters

$$\begin{aligned} c_t &\approx \ln y_t^2 - \ln \sigma_\epsilon^2 + 1.2704, \\ d_t &= \frac{1}{2}\pi^2. \end{aligned}$$

The resulting linear Gaussian state space model is amenable to evaluation using the Kalman filter equations, which deliver the likelihood of this model efficiently. No correction is made for approximating the density of $\ln e_t^2$ by a simple normal density, hence the resulting optimisation delivers no true but only quasi-maximum likelihood (QML) estimates.

2.2 Gaussian mixture sampling

In Carter and Kohn (1997), Kim et al. (1998) and Omori et al. (2007) a similar linearisation of the observation equation is used, but now the density of $\ln e_t^2$ is approximated using a Gaussian mixture (GM) density, as in

$$\ln e_t^2 \approx u_t \sim \sum_{i=1}^C p_i \mathcal{N}(m_i, s_i^2).$$

In this notation, p_i are the weights of the mixture components, and m_i, s_i^2 the respective means and variances. The distinction between the three articles is the number C of mixture components that is used (5, 7 and 10 respectively), where obviously the mixture with the largest number of components delivers the closest match between the approximating and true density. The QML method would correspond to a 1-component mixture in this notation.

With this extension to a Gaussian mixture, the full model is no longer an unconditionally linear and Gaussian state space (StSp) model, as it depends on the discrete unobserved states i_t . Instead of optimising the likelihood directly, a Bayesian Markov chain Monte Carlo (MCMC) approach with data augmentation is used: If one assumes indices i_t indicating which

element i of the mixture is used at time t , then the model for $Y|I$ is *conditionally* linear and Gaussian, where $I = \{i_1, \dots, i_T\}$. In terms of the general approximating linear Gaussian observation equation (4), conditional choices of

$$c_t|i_t = \ln y_t^2 - \ln \sigma_\epsilon^2 - m_{i_t}, \quad d_t|i_t = s_{i_t}^2$$

are used.

The simulation smoother (De Jong and Shephard, 1995; Durbin and Koopman, 2002) delivers a sample from the states H conditional on the data Y , state indices I and parameters θ . Conditionally on the states H , on data Y and on the parameters θ , each mixture node has a certain probability $P(i_t = i)$ which can be evaluated. From the resulting multinomial density a sample of new indices i_t is drawn. Afterwards, conditional on states I, H and data Y , new parameters θ are drawn as in a standard Gibbs scheme (Smith and Gelfand, 1992). A final sample from the parameters and states describes the posterior density of these, conditional on the full set of data. The posterior mean of the states is an estimate of the smoothed state.

2.3 Simulated method of moments

Where QML uses a likelihood approach, and GMS a Bayesian MCMC, the simulated method of moments approach (SMM, Gallant and Tauchen, 1996; Andersen et al., 1999) takes a third route to parameter estimation. If the loglikelihood function $\ln P_{\text{SV}}(Y; \theta)$ of the SV model were available in closed form, then it could be optimised by finding those parameters which set the average score to zero, with score function

$$s_{\text{SV}}(Y; \theta) = \frac{1}{T} \frac{\partial \ln P_{\text{SV}}(Y; \theta)}{\partial \theta}.$$

This true score function cannot be evaluated directly, as the likelihood function of SV models is not available in closed form. SMM instead uses an auxiliary model which fits the most important characteristics of the data closely, and is easier to estimate. In this case, for the SV model, some GARCH model could be used as the auxiliary model. Suppose that this auxiliary model delivers auxiliary parameters $\hat{\theta}_{\text{Aux}}$. Clearly, at these estimates the score of the loglikelihood $\ln P_{\text{Aux}}(Y, \hat{\theta}_{\text{Aux}})$ of the auxiliary model for the current set of data equals zero. But if data were generated from the SV model, for the optimal set of parameters $\hat{\theta}$, then the same $\hat{\theta}_{\text{Aux}}$ should provide a good fit for the auxiliary model as well. Hence SMM evaluates

$$s_{\text{Aux}}(Y^*(\theta), \hat{\theta}_{\text{Aux}}) = \frac{1}{T_{\text{Aux}}} \frac{\partial \ln P_{\text{Aux}}(Y^*(\theta); \theta_{\text{Aux}})}{\partial \theta_{\text{Aux}}} \Big|_{\theta_{\text{Aux}} = \hat{\theta}_{\text{Aux}}},$$

which is the score of the auxiliary model, at parameters $\hat{\theta}_{\text{Aux}}$, for T_{Aux} observations $Y^*(\theta)$ simulated from the SV model using parameters θ . Searching for those parameters θ which equate these scores to zero delivers estimates of the parameters of the SV model, through the detour of the auxiliary model.

As an auxiliary model, the QML-SV model of Harvey et al. (1994) mentioned before could be used, though for this QML-SV model the score function is not available analytically either. As mentioned before, a GARCH model, with either Gaussian or Student- t disturbances, or an EGARCH (Nelson, 1991), provides an approximating model where the analytic scores can be derived. In those cases, the parameters of the auxiliary model are different in number

and type from the parameters in the true model. Therefore, the score vector of the auxiliary model should be scaled by the inverse information matrix of the model, minimising effectively

$$Q(\theta; \hat{\theta}_{\text{Aux}}) = s_{\text{Aux}}(Y^*(\theta), \hat{\theta}_{\text{Aux}})' \mathcal{I}^{-1}(\hat{\theta}_{\text{Aux}}) s_{\text{Aux}}(Y^*(\theta), \hat{\theta}_{\text{Aux}}),$$

$$\mathcal{I}(\hat{\theta}_{\text{Aux}}) = \frac{1}{T} \sum \left. \frac{\partial \ln P_{\text{Aux}}(y_t; \theta_{\text{Aux}})}{\partial \theta_{\text{Aux}}} \frac{\partial \ln P_{\text{Aux}}(y_t; \theta_{\text{Aux}})}{\partial \theta'_{\text{Aux}}} \right|_{\theta_{\text{Aux}} = \hat{\theta}_{\text{Aux}}}.$$

If the auxiliary model indeed covers the important characteristics of the data, and if the auxiliary sample size T_{Aux} is sufficiently large, then optimising $Q(\theta; \hat{\theta}_{\text{Aux}})$ provides an efficient estimate of the parameters θ of the SV model (see also Andersen et al., 1999). SMM however never optimises, or even calculates, a likelihood of the SV model, but arrives at a set of parameter estimates without even touching upon the underlying states. If only parameter estimation is requested, it can be a good alternative to the other methods in this chapter.

2.4 Methods based on importance sampling

The three aforementioned methods all have their drawbacks: QML is only approximative, GMS does not provide a classical maximum likelihood estimator, and SMM lacks an estimate for the state of the volatility. A set of methods based on the idea of importance sampling provides alternative estimators. In the following, three versions of importance sampling will be contrasted: basic Importance Sampling (IS, Durbin and Koopman, 1997), Efficient Importance Sampling (EIS, Liesenfeld and Richard, 2006; Richard and Zhang, 2007), and an improved version of importance sampling (IIS, Nguyen, 2007) which will be found to be as efficient as EIS.

For the basic method of importance sampling, the likelihood function $\mathcal{L}(Y; \theta)$ is approximated using simulation through

$$\mathcal{L}(Y; \theta) = \int \frac{P(Y, H)}{G(H|Y^*)} G(H|Y^*) \partial H$$

$$\approx \frac{1}{M} \sum \frac{P(Y, H^{(i)})}{G(H^{(i)}|Y^*)} = \frac{1}{M} \sum \omega_i \equiv \bar{\omega}, \quad (5)$$

where the $H^{(i)}$ are sampled volatilities from the approximating importance density $G(H|Y^*)$. The latter density depends on some auxiliary data Y^* , and is chosen such that it is a close approximation to $P(Y, H)$ (where dependence on parameters θ is skipped in the notation for simplicity). From this importance density, M series of $H^{(i)} = \{h_1^{(i)}, \dots, h_T^{(i)}\}$ are drawn and used for evaluating the likelihood.²

The approximation in (5) can be made as accurate as required by increasing the number M of sampled volatility sequences. If a common random seed is used for sampling $H^{(i)}$ when the likelihood is evaluated for different values of θ , the likelihood will be a smooth function of the parameters, and standard quasi-Newton optimisation methods can be used.

²In practice, it is often desired to estimate $\log \mathcal{L}(Y; \theta)$. Durbin and Koopman (1997) derive how an unbiased estimator for this quantity is

$$\log \mathcal{L}(Y; \theta) \approx \log \bar{\omega} + \frac{s_{\omega}^2}{2M\bar{\omega}^2},$$

with $\bar{\omega}$ the average weight and s_{ω}^2 the variance.

The three methods IS, EIS and IIS use the same setup (4) for the approximating linear observation equation. Where QML chose c_t, d_t only taking the $\ln \chi^2$ density into account, the IS methods also look at the values of h_t around the current observation, to obtain an optimal fit between $P(Y, H)$ and $G(H|Y^*)$. The methods IS, EIS and IIS differ in the way optimal $Y^* \equiv \{c_t, d_t, t = 1, \dots, T\}$ and the sample of $H|Y^*$ are obtained.

Approximating in the basic IS approach

In the method of Durbin and Koopman (1997) full advantage of the linear Gaussian unobserved component structure of the approximating model (2) and (4) is taken. Notice that $G(H|Y^*)$ is the density of the linear states, conditional on the auxiliary data as represented by Y^* . A sample from this density can be drawn using the simulation smoother. The weights ω_i are evaluated by rewriting

$$G(H^{(i)}|Y^*) = \frac{G(Y^*|H^{(i)})G(H^{(i)})}{G(Y^*)},$$

$$P(Y, H^{(i)}) = P(Y|H^{(i)})P(H^{(i)}),$$

such that

$$\begin{aligned} \omega^{(i)} &= \frac{P(Y, H^{(i)})}{G(H^{(i)}|Y^*)} = G(Y^*) \frac{P(Y|H^{(i)})P(H^{(i)})}{G(Y^*|H^{(i)})G(H^{(i)})} \\ &= G(Y^*) \frac{P(Y|H^{(i)})}{G(Y^*|H^{(i)})} = G(Y^*) \prod_t \frac{P(y_t|h_t^{(i)})}{G(y_t^*|h_t^{(i)})}. \end{aligned}$$

In these equations, $G(H^{(i)})$ represents the density of the states as defined by the transition equation (2), and $G(Y^*|H^{(i)})$ the density corresponding with the approximating observation equation (4). $G(Y^*)$ is the full unconditional likelihood of the auxiliary data according to the approximating model and can be evaluated using the KF equations. As the transition equations of the auxiliary and true model are the same, the factor $P(H^{(i)})/G(H^{(i)})$ cancels from the weight ω_i .

For finding the optimal $y_t^* = \{c_t, d_t\}$ of the approximating model, IS ensures that the densities relating to the observation equations (1) and (4) correspond in their first two moments. Define $p \equiv \ln P(y_t|h_t), g \equiv \ln P(y_t^*|h_t)$ with first and second derivatives with respect to h_t indicated as $\dot{p}, \dot{g}, \ddot{p}, \ddot{g}$, then

$$p = -\frac{1}{2} \left(\ln 2\pi + \ln \sigma_\epsilon^2 + h_t + \frac{y_t^2}{\sigma_\epsilon^2 \exp(h_t)} \right),$$

$$g = -\frac{1}{2} \left(\ln 2\pi + \ln d_t + \frac{(c_t - h_t)^2}{d_t} \right), \quad (6)$$

$$\dot{p} = -\frac{1}{2} \left(1 - \frac{y_t^2}{\sigma_\epsilon^2 \exp(h_t)} \right), \quad \dot{g} = \frac{c_t - h_t}{d_t}, \quad (7)$$

$$\ddot{p} = -\frac{y_t^2}{2\sigma_\epsilon^2 \exp(h_t)}, \quad \ddot{g} = -\frac{1}{d_t}. \quad (8)$$

For a given value of h_t , the corresponding optimal values of c_t, d_t can be solved from this system of equations.³ Ideally, the approximating model would fit $\ln P(y_t|h_t)$ well around a plausible value of h_t , for all $t = 1, \dots, T$. One such a plausible value of $H = (h_1, \dots, h_T)$ could be the smoothed state \tilde{H} from the auxiliary model. By iterating back and forth between improving estimates of Y^* and extracting a new \tilde{H} , an optimal solution for the parameters can be found in what is usually a small number of iterations (Sandmann and Koopman, 1998).

Improving on IS with IIS

The IS method of estimating a SV model is flexible, relatively easy to implement, and not too computationally intensive in practice either. A weak spot of the IS method, first recognised in the EIS method (Liesenfeld and Richard, 2006) described below, is that the likelihood evaluation (5) is based on a M samples $H^{(i)}, i = 1, \dots, M$ of the volatilities, whereas the auxiliary model is only fitted to the true stochastic volatility model using a second order Taylor expansion at the mode \tilde{H} .

The Taylor expansion however provides the closest possible approximation only locally, where the error in the approximation is increasing with the order $O((h_t - \tilde{h}_t)^3)$, for the second order Taylor expansion used here. Therefore, there is no guarantee that the current approximation is sufficiently close over the full range of possibly interesting values of H .

Nguyen (2007) proposes to take M samples from the approximating model, $H^{(i)}, i = 1, \dots, M$, and minimise the distance between the two observation densities $P(y_t|h_t^{(i)})$ and $G(y_t^*|h_t^{(i)})$, following the example of the EIS sampler of Richard and Zhang (2007). As all dependence through time is taken up by the simulation smoother and incorporated into the sample $H^{(i)}$, optimising the distance between the observation densities can be done at each time t separately from other time-points. A simple but effective approach is to minimise the variance of the log-weights $\ln \omega_t^{(i)}$, as in

$$\begin{aligned} Q_t(c_t, d_t) &= \sum_i (\ln \omega_t^{(i)} - \overline{\ln \omega_t})^2 = \sum_i \left(p(y_t|h_t^{(i)}) - g(y_t^*|h_t^{(i)}) + \text{const} \right)^2 \\ &= \sum_i \left(-\frac{1}{2}h_t^{(i)} - \frac{y_t^2}{2\sigma_\epsilon^2 \exp(h_t^{(i)})} + \frac{\ln d_t}{2} + \frac{(h_t^{(i)} - c_t)^2}{2d_t} + \text{const} \right)^2 \\ &\equiv \sum_i \left(z^{(i)} - a_0 - a_1 h_t^{(i)} - a_2 (h_t^{(i)})^2 \right). \end{aligned}$$

The optimal solution for c_t, d_t can be found using a linear regression of

$$z^{(i)} = -\frac{1}{2}h_t^{(i)} - \frac{y_t^2}{2\sigma_\epsilon^2 \exp(h_t^{(i)})} \tag{9}$$

on $h_t^{(i)}$ and $(h_t^{(i)})^2$, solving for c_t, d_t from

$$a_1 = \frac{c_t}{d_t}, \quad a_2 = -\frac{1}{2d_t}. \tag{10}$$

³There is no guarantee that the second derivative \ddot{p} is negative, such that some variances d_t could become non-positive. Jungbacker and Koopman (2007) note that the simulation smoothing equations of De Jong and Shephard (1995) can be adapted to allow for this case, without influencing the effectiveness of the algorithm.

Note that the value of a_0 is not of interest, as it takes up $\overline{\ln \omega_t}$ and the integrating constants of the densities.

Given a new set of estimates $Y^* = \{c_t, d_t, t = 1, \dots, T\}$, a more appropriate sample of H can be drawn, and the optimisation repeated until convergence. As with the basic IS method, convergence will usually take place within a few iterations, and the final weights ω_t are expected to display lower variability.

Alternative efficiency gains with EIS

As seen with the IIS method, the fit of the auxiliary density could be improved compared to IS when the true and auxiliary density are matched for a sample $H^{(i)}$ of volatility sequences. Where IIS uses the simulation smoother to sample from the density of $H|Y^*$, EIS chooses instead to sample from the predictive densities, i.e. to sample each $h_t|y_t^*, h_{t-1}$. The final auxiliary sampling density will be indicated by $G^{\text{EIS}}(H|Y^*)$, where the notation does explicitly indicate that the sampling density here is the predictive density only.

These predictive densities can be constructed from the combination of the information stemming from Gaussian transition equation (2) and observation equation (4), as

$$G(h_t|y_t^*, h_{t-1}) = \frac{G(h_t|y_t^*)G(h_t|h_{t-1})}{\chi_t(h_{t-1})} \propto G(h_t|y_t^*)G(h_t|h_{t-1}),$$

$$h_t|y_t^* \sim \mathcal{N}(c_t, d_t), \quad h_t|h_{t-1} \sim \mathcal{N}(\phi h_{t-1}, \sigma_\xi^2).$$

This implies that $h_t|y_t^*, h_{t-1} \sim \mathcal{N}(m_t, s_t^2)$ with

$$s_t^2 = \left(\frac{1}{\sigma_\xi^2} + \frac{1}{d_t} \right)^{-1}, \quad m_t = s_t^2 \left(\frac{\phi h_{t-1}}{\sigma_\xi^2} + \frac{c_t}{d_t} \right).$$

The integrating constant $\chi_t(h_{t-1})$ is

$$\chi_t(h_{t-1}) = \sqrt{\frac{2\pi d_t \sigma_\xi^2}{s_t^2}} \exp\left(\frac{c_t^2}{2d_t} + \frac{\phi^2 h_{t-1}^2}{2\sigma_\xi^2} - \frac{m_t^2}{2s_t^2} \right) \propto \exp\left(\frac{\phi^2 h_{t-1}^2}{2\sigma_\xi^2} - \frac{m_t^2}{2s_t^2} \right).$$

Note how $\chi_t(h_{t-1})$ depends on the previous value of the volatility, both directly through $\phi^2 h_{t-1}^2$ and also indirectly through the value m_t^2 . It is through the integrating constant that part of the information on the previous observation is handed over towards the present.

EIS, like IIS, looks at the variance of the log-weights

$$\begin{aligned} \ln \omega^{(i)} &= \sum_t \left(p(y_t|h_t^{(i)}) + p(h_t^{(i)}|h_{t-1}^{(i)}) - g(h_t^{(i)}|y_t^*) - g(h_t^{(i)}|h_{t-1}^{(i)}) + \ln \chi_t(h_{t-1}^{(i)}) \right) \\ &= \sum_t \left(p(y_t|h_t^{(i)}) - g(h_t^{(i)}|y_t^*) + \ln \chi_t(h_{t-1}^{(i)}) \right) \\ &\equiv \sum_t \left(p(y_t|h_t^{(i)}) - g(h_t^{(i)}|y_t^*) + \ln \chi_{t+1}(h_t^{(i)}) \right) = \sum_t w_t(h_t^{(i)}). \end{aligned}$$

In the third step, the integrating constant is moved forward in time, such as to combine all elements concerning $h_t^{(i)}$ together in $w_t(h_t^{(i)})$. Defining $\chi_{T+1}(h_T) = 1$ takes care of the endpoint condition.

At each time point t the sum-of-squares function to minimise becomes

$$\begin{aligned}
Q_t(c_t, d_t) &= \sum_i (w_t(h_t^{(i)}) - \bar{w}_t)^2 \\
&= \sum_i \left(-\frac{1}{2}h_t^{(i)} - \frac{y_t^2}{2\sigma_\epsilon^2 \exp(h_t^{(i)})} + \frac{\ln d_t}{2} + \frac{(h_t^{(i)} - c_t)^2}{2d_t} \right. \\
&\quad \left. + \frac{\phi^2 h_t^{(i)2}}{2\sigma_\xi^2} - \frac{m_{t+1}^{(i)2}}{2s_{t+1}^2} + \text{const} \right)^2 \\
&\equiv \sum_i \left(z_i - a_0 - a_1 h_t^{(i)} - a_2 h_t^{(i)2} \right)^2,
\end{aligned}$$

where now

$$z^{(i)} = -\frac{1}{2}h_t^{(i)} - \frac{y_t^2}{2\sigma_\epsilon^2 \exp(h_t^{(i)})} + \frac{\phi^2 h_t^{(i)2}}{2\sigma_\xi^2} - \frac{m_{t+1}^{(i)2}}{2s_{t+1}^2}. \quad (11)$$

Note how this specification for $z^{(i)}$ in (11) corresponds to the IIS regressor in (9), apart from the added terms relating to the integrating constant $\ln \chi_{t+1}(h_t)$. Minimising the sum of squares of $w_t(h_t^{(i)})$ then again corresponds to finding the solution of the linear regression model relating $z^{(i)}$ to $h_t^{(i)}$ and its square. From the regression output on a_1, a_2 the optimal values of (c_t, d_t) are extracted as in (10). This can be done recursively, starting at the end of the sample (as $\chi_{T+1}(h_t) \equiv 1$ is known), working backwards towards $t = 1$.

At this stage, a theoretical comparison between the setup for IS, IIS and EIS can be made. From the above description it is clear how IS uses the simulation smoother for a linear approximating Gaussian model as auxiliary sampler. The approximating model is fitted such that at the smoothed estimate \tilde{H} the true and approximating log-densities correspond in their first two moments. With IIS, the same approximating model is used, but the approximation is optimised not for an estimate \tilde{H} but for a full sample. EIS changes the auxiliary sampler to contain only past and present information; future information is taken into account by adapting for the integrating constant in the sum-of-squares function, and hence it uses the same amount of information as IIS. Below, in Section 3, it will indeed be seen that IIS and EIS are equally efficient, both surpassing the precision of the likelihood estimate of IS considerably.

2.5 Alternative sampling methods: SSS and MMS

The IS-based methods effectively provide a sampling-based alternative to evaluating the exact loglikelihood function for the stochastic volatility model (albeit by simulation), instead of the rough approximation of the QML method of Section 2.1. A different route discussed in Section 2.2 provided a Bayesian approach to a similar goal: Improve the approximation of the density of $\ln e_t^2$ through a mixture density, and sample in turn from all full conditional densities. This latter approach, however exact the mixture density can be, is still an approximation, as the mixture density is theoretically not equal to the target density.⁴

Two alternative flexible methods are available for obtaining a theoretically exact Bayesian sampler for the SV model. These methods, the Single Site Sampler (SSS) and the MultiMove

⁴In the original work of Kim et al. (1998) this is solved by performing an extra reweighting step. Here we will forgo this possibility for simplicity.

Sampler (MMS) can be described together, where the SSS is a special (and simpler) case of the MMS. Both methods provide a sampler of H from the true SV model. The methods follow the framework of GMS of Section 2.2 in that they try to sample H from the full conditional density. Instead of enlarging the parameter space with a set of mixture indices I , the problem of sampling $H = \{h_1, \dots, h_T\}$ is broken up into smaller blocks. Say we have block sizes $k_1, \dots, k_k, \sum k_j = T$, then successively each block of k_j volatilities $H_{k_j} \equiv \{h_{t_j}, \dots, h_{t_{j+1}-1}\}$ with $t_1 = 1, t_l = t_{l-1} + k_{l-1}, l > 1$ is sampled conditioning on Y and $(H \setminus H_{k_j}) = \{h_1, \dots, h_{t_j-1}, h_{t_{j+1}}, \dots, h_T\}$. With SSS the blocksize $k_j = 1$, whereas with MMS k_j is larger than 1.

To sample from the present block, a slightly enlarged approximating model is fitted to the data. Equations (2) and (4) are extended with end-point conditions for the auxiliary data outside the current block,

$$c_{t_j-1} \equiv h_{t_j}, \quad d_{t_j-1} \equiv 0, \quad (12)$$

$$c_{t_{j+1}} \equiv h_{t_{j+1}}, \quad d_{t_{j+1}} \equiv 0. \quad (13)$$

The knots outside the block are fixed at the current value of h_{t_j-1} and $h_{t_{j+1}}$. Within the block, again c_t, d_t can be chosen to provide a good fit for the approximating model to the true observation equation. As in Section 2.4, the first two derivatives of the log-densities are equated ($\dot{p} \equiv \dot{g}, \ddot{p} \equiv \ddot{g}$), for an estimate of the present set of volatilities. Again, one could use a smoothed estimate \tilde{H}_{k_j} , or even a full sample of $\tilde{H}_{k_j}^{(i)}$.

From the approximating model a candidate sample $H_{k_j}^*$ is drawn using the simulation smoother. The candidate draw $H_{k_j}^*$ is accepted as new value for $H_{k_j}^{(i)}$ with Metropolis-Hastings (MH) acceptance probability α , comparing the likelihood contributions of the true and approximating models:

$$\alpha \left(H_{k_j}^*, H_{k_j}^{(i-1)} \right) = \min \left[\frac{P(Y_{k_j} | H_{k_j}^*) G(Y_{k_j}^* | H_{k_j}^{(i-1)})}{P(Y_{k_j} | H_{k_j}^{(i-1)}) G(Y_{k_j}^* | H_{k_j}^*)}, 1 \right].$$

If the candidate draw is not accepted, one continues with $H_{k_j}^{(i)} \equiv H_{k_j}^{(i-1)}$, else $H_{k_j}^{(i)} \equiv H_{k_j}^*$. This approximating and sampling step is taken successively for all blocks, until a full new sample $H^{(i)}$ is obtained.

With SSS, constructing the optimal approximating density does not need to go through the full state space setup of the above equations, as the optimal Gaussian approximation sampler for $h_t | h_{t-1}, h_{t+1}, y_t$ can be found analytically. Notice that there are three sources of information on h_t , with

$$h_t | h_{t-1} \sim \mathcal{N}(\phi h_{t-1}, \sigma_\xi^2) \equiv \mathcal{N}(c^p, d^p), \quad \text{Past information,}$$

$$h_t | h_{t+1} \sim \mathcal{N}\left(\frac{1}{\phi} h_{t+1}, \frac{1}{\phi^2} \sigma_\xi^2\right) \equiv \mathcal{N}(c^f, d^f), \quad \text{Future information,}$$

$$h_t | y_t^* \sim \mathcal{N}(c_t, d_t) \equiv \mathcal{N}(-\ddot{p}^{-1} \dot{p} + \tilde{h}_t, -\ddot{p}^{-1}), \quad \text{Time } t \text{ information.}$$

The \tilde{h}_t is an estimate of the present h_t , for which the last draw could be filled in. These three normal densities can be convoluted into one sampling density, with

$$h_t | h_{t-1}, h_{t+1}, y_t^* \sim \mathcal{N}(c^c, d^c), \quad \text{Convoluted information,}$$

$$c^c = d^c \left(\frac{c^p}{d^p} + \frac{c^f}{d^f} + \frac{c_t}{d_t} \right), \quad d^c = \left(\frac{1}{d^p} + \frac{1}{d^f} + \frac{1}{d_t} \right).$$

This manner of sampling, conditioning on the present and immediate past and future information, may lead to large correlation in the Markov chain of sampled volatilities, and of sampled parameters. The MMS lowers the correlation by sampling larger blocks of volatilities at once. When the blocksize becomes too large, the acceptance rate of the MH step could become too low, again leading to an increase in sample correlation. Hence an intermediate choice for the average blocksize should be made.

As the elements h_t close to the knots are tied down more than elements in between two knots, it is advisable to change the location of the knots between iterations. The random procedure of Shephard and Pitt (1997) for choosing knots is used in this chapter.

3 Comparison of methods

3.1 Setup of data generating process and estimation procedures

From a theoretical point-of-view, the links between the estimation methods are described in Section 2. The practical implication of those differences for the estimation results is the topic of the present section. We consider a pure SV model (1)–(2), with parameters $\sigma_\epsilon, \phi, \sigma_\xi$ as specified in Table 2. A value of $\phi = 0.95$ allows for a considerable amount of persistence in the volatility, while $\sigma_\xi = 0.1$ indicates a long-run variance of volatility of $\sigma_\xi^2/(1 - \phi^2) \approx .1$. The σ_ϵ is a scaling parameter, and is set at 1. Note that the choices for these parameter are modelled to mimic results of Section 4.

Further columns in the table indicate the prior densities for the parameters. For ϕ , a conjugate normal prior is taken, with expectation 0.9 and standard deviation 0.1. For $\sigma_\epsilon, \sigma_\xi$, Inverted Gamma-1 priors are conjugate. The prior-parameters are chosen such that the prior mean corresponds roughly with the DGP value, and the prior standard deviation allows for sufficient flexibility for the data to decide on the location of the posterior.

Table 2: DGP and prior specifications

	DGP	$\pi(\theta)$	r	a	$\mathbf{E}(\theta)$	$\mathbf{s}(\theta)$
σ_ϵ	1	IG-1(r, a)	2	1.5	1.09	0.57
ϕ	0.95	$\mathcal{N}(\mu, \sigma^2)$			0.9	0.1
σ_ξ	0.1	IG-1(r, a)	1.3	0.006	0.1	0.1

With these settings, 1,000 data series of length $T = 1,000$ are simulated from the SV model. The parameters of the SV model are estimated using each of the methods, for each of these data series, using a set of programs written in Ox (Doornik, 2009) in combination with SsfPack (Koopman et al., 2008). Estimation is started at the values of the DGP, and continued until convergence for the QML, SMM, IS, EIS and IIS methods. For the Bayesian methods GMS, SSS, MMS a sample of 10,000 draws from the posterior density is collected, after a burn-in sample of 1,000 iterations for obtaining an initial estimate of H while fixing the parameters, followed by a further 1,000 iterations where both H and parameters are sampled, but not stored. From the posterior sample the mode is taken as the estimate of the parameters. The methods based on importance sampling take $M = 250$ draws from the auxiliary density using antithetic sampling, therefore effectively $M = 500$ weights are evaluated. For SMM, a GARCH(1, 1) auxiliary model with either normal (SMMn) or Student- t (SMMt) disturbances is used. From the estimated auxiliary model $N = 100,000$ observations are generated for each evaluation of the criterion function.

3.2 Parameter estimates for the simulation

Table 3 displays results on the parameter estimates for the simulation. It reports the average parameter estimates (in the first three columns) together with the standard deviation of the estimates (in columns 4-6), and the average time to completion in seconds. It can be seen that all methods estimate the standard deviation σ_ϵ close to the DGP value of 1. The overall level of volatility is relatively easy to extract from the data. For the persistence parameter ϕ , QML leads to a considerable downward bias, with SMM and the IS methods displaying a less severe bias. The Bayesian methods in the last three rows of the table seem to estimate ϕ closest to 0.95. In general, if ϕ is underestimated, then the parameter σ_ξ tends to be overestimated. The bias on the long run variance $\sigma_\xi^2/(1 - \phi^2)$ is generally smaller than the bias on the underlying parameters themselves.

Table 3: Parameter estimates for SV simulation

	$\bar{\sigma}_\epsilon$	$\bar{\phi}$	$\bar{\sigma}_\xi$	$s(\sigma_\epsilon)$	$s(\phi)$	$s(\sigma_\xi)$	Time
DGP	1	0.95	0.1				
QML	1.003	0.809	0.174	0.052	0.248	0.218	0.1
SMMn	0.996	0.934	0.113	0.049	0.065	0.114	38.9
SMMt	0.990	0.868	0.147	0.056	0.181	0.120	43.8
IS	0.998	0.914	0.116	0.041	0.115	0.063	63.0
EIS	0.998	0.913	0.117	0.041	0.116	0.064	133.0
IIS	0.998	0.914	0.117	0.041	0.112	0.064	300.7
GMS	0.997	0.944	0.086	0.040	0.030	0.027	14.1
SSS	0.999	0.945	0.088	0.041	0.032	0.031	13.1
MMS	1.000	0.944	0.088	0.038	0.030	0.030	253.9

The second set of columns displays the standard deviation of the parameter estimates. The Bayesian methods deliver the lowest uncertainty in the estimates. This effect can also be seen from Figure 1, which displays the density of the 1,000 parameter estimates for each of the methods. The top panels display results for σ_ϵ , ϕ and σ_ξ for the classical QML, SMM, IS, EIS and IIS methods, whereas the bottom panels plot the density of the parameters for the Bayesian GMS, SSS and MMS methods, together with the prior. From the figure it is seen how the mode of the estimates of σ_ξ in panel (i-a) corresponds closely with the DGP value of $\sigma_\xi = 0.1$, but the large right-hand tail leads to an upward bias for the estimates on average. The IS, EIS and IIS results are plot separately, but are not distinguishable from each other in Figure 1 as the estimates are virtually equal. QML delivers only quasi-maximum likelihood estimates, which are clearly distinct from the IS estimates. The SMM estimates theoretically should deliver the same efficient estimates as the maximum likelihood-based methods; the distinction in results in this case appears to be due to neither of the auxiliary models fitting the true model sufficiently well at the present sample size.⁵

With the Bayesian methods, estimates of σ_ξ in the bottom right panel are seen to be spread around 0.1. The distribution of these estimates is close to the distribution of the prior for σ_ξ , indicating that the present DGP with 1,000 observations provides little information on this parameter measuring volatility of volatility. The prior influences the location of the

⁵As additional auxiliary model a semi-nonparametric GARCH model (Gallant and Nychka, 1987) was tried. With $T = 1,000$ observations the data was not informative enough to obtain good convergence for the auxiliary parameters.

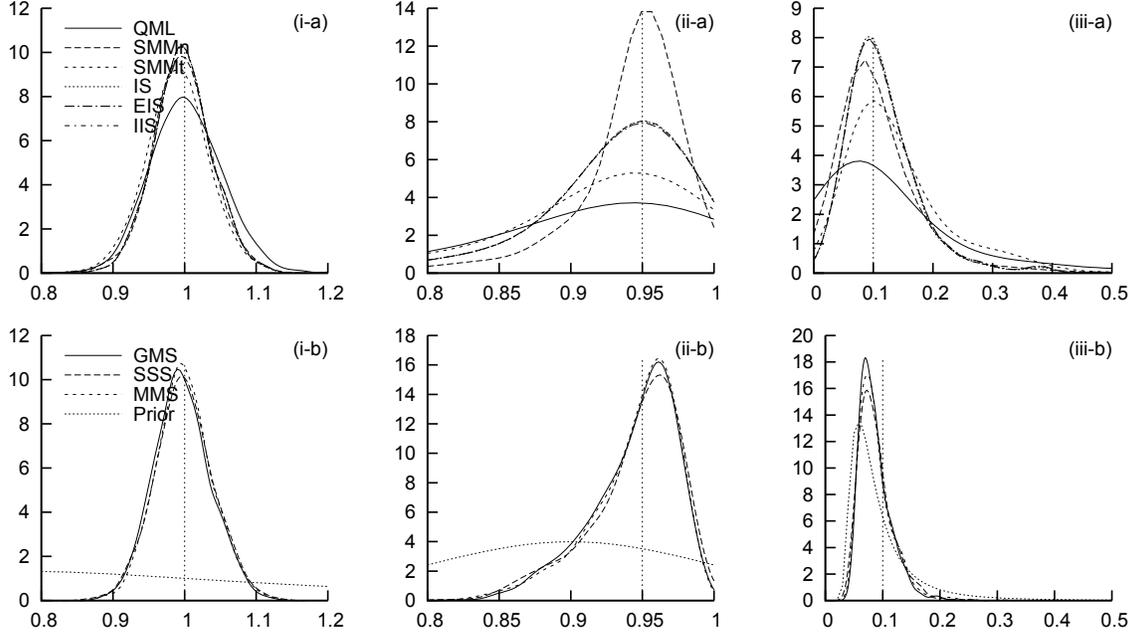


Figure 1: Density of parameter estimates σ_ϵ (panels (i-a/b)), ϕ (panels (ii-a/b)) and σ_ξ (panels (iii-a/b)) for SV simulation exercise, for classical (panels (a)) and Bayesian (panels (b)) estimation methods.

posterior mode estimates relatively strongly in such a case.

The last column of Table 3 reports the average optimisation times in seconds. QML is clearly the fastest, as it is only a simple and direct approximation. The speed of SMM depends fully on the length of the auxiliary sampler. For the IS methods, the basic IS is moderately quick as the approximating model needs only a single run from the volatility sequence, whereas EIS needs a full sample of runs for the approximating algorithm. EIS and IIS theoretically make a similar computational effort, though in practice it is simpler to generate a sample from EIS's predictive density than from IIS's simulation smoother. For the Bayesian samplers, clearly MMS is computationally most intensive. For each block of observations, an approximating model has to be found, and a separate sample is drawn. For the SSS, the approximating model can be obtained using analytic closed form expressions for the smoothed states, which tends to be quicker. GMS is comparable to SSS, at least in effort that is needed for obtaining a similarly sized sample from the posterior density. Note how these computational times are roughly in line with the theoretical properties of the estimators: The procedures which are theoretically nicer, providing either a direct link with the likelihood or a full sample from the posterior density take more time. While EIS may be slightly harder than IIS to set up, this extra effort is rewarded with a smaller computational effort.

3.3 Precision of IS

Among the methods, IS, IIS and EIS are based on the idea of importance sampling, and the final likelihood estimate is calculated in (5) as the average of a set of weights, $L(Y; \theta) \approx \frac{1}{M} \sum \omega_i$. This estimator of the likelihood only converges if a central limit theorem holds for

the average of weights ω_i (Koopman et al., 2009). Most importantly, the variance of the weights should converge to a finite constant. A lower variance then corresponds to a more precise likelihood estimate.

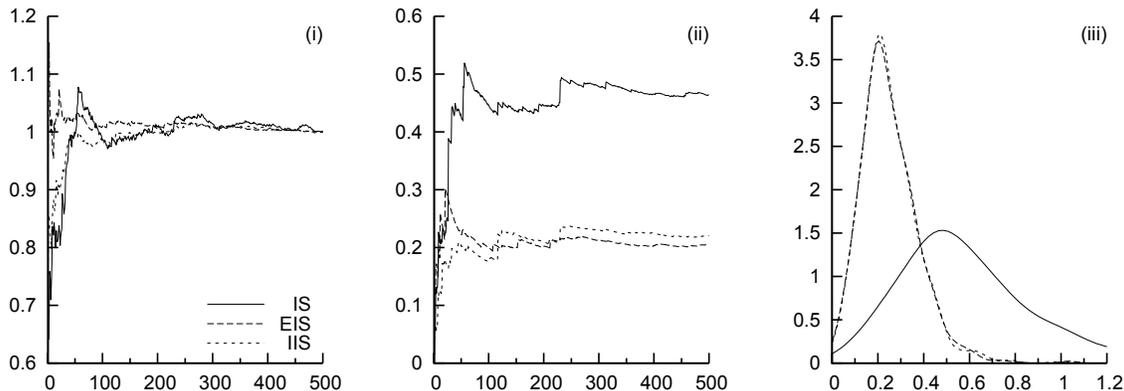


Figure 2: Cumulative average of standardised weights (panel (i)) and related standard deviation (ii) for one iteration of IS, EIS and IIS, and density of 1,000 final weight standard deviations of full simulation (iii).

For one specific sample of the DGP, the three methods were used for estimation, and at the final estimates the weights ω_i were collected. A sample of size $M = 250$ with antithetic sampling was used, resulting in a total of 500 weights for each method. The weights were standardised to have mean 1. Figure 2 plots the cumulative estimates of the average of the weights (panel (i)) and of the standard deviation (panel (ii)).

From panel (i) it is seen that for IS the cumulative average converges slower than for the EIS and IIS approaches. In the second panel indeed the final standard deviation of the weights for IS is found to be larger than for EIS and IIS. The latter two methods appear to have virtually equal standard deviations for their weights. Panel (iii) of Figure 2 compares the final standard deviations of IS, EIS and IIS over the $S = 1,000$ simulations. Indeed the IS is less efficient in general, and the EIS and IIS are indistinguishable. As EIS and IIS use exactly the same information for constructing the auxiliary density, they lead to similar draws of the weights with similar standard deviations.

3.4 Precision of Bayesian methods

With the Bayesian methods GMS, SSS and MMS a final sample of size $S = 10,000$ of draw from the posterior was collected. For MMS the average blocksize had to be chosen. A small blocksize leads to behaviour similar to the SSS, sampling a single volatility at a time. Larger block sizes could lower correlation in the chain, as a larger number of volatilities is updated in a single sweep. When the block size becomes too large, the Metropolis-Hastings acceptance probability drops, and correlation again increases when a candidate draw is rejected too often (see also the discussion in Shephard and Pitt (1997) and Liesenfeld and Richard (2008)).

In the above simulation, the volatility series of H was divided into 7 blocks, resulting in a blocksize of around $k = T/7 \approx 143$ volatilities h_t on average when using MMS. To judge the effect of the block size, panel (i) of Figure 3 displays the relative numerical inefficiency (Shephard and Pitt, 1997, $R_I(B)$), for different values of the average blocksize. The $R_I(B)$

compares the variance of the posterior sample when correcting for B lags of possible autocorrelation, to the variance when no correction for autocorrelation is made. A bandwidth of $B = 1,000$ lags is chosen.

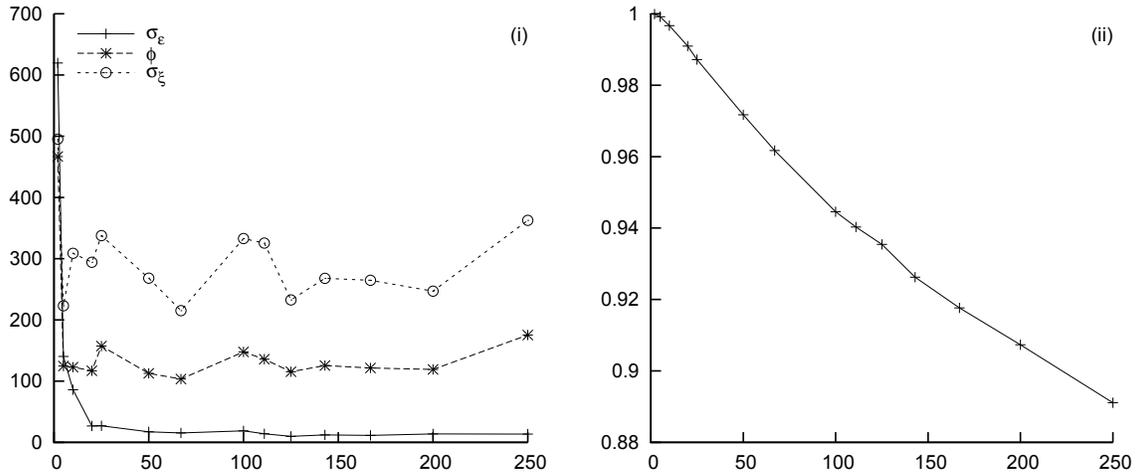


Figure 3: Inefficiency versus blocksize (panel (i)) and acceptance rate (ii)

In panel (i) it is seen how for very small block sizes, inefficiency is high. Starting at a blocksize of at least 10, inefficiency stabilises, until it increases again when the blocksize reaches 250. These results correspond with the findings of Shephard and Pitt (1997), that the efficiency gains are fairly robust against the blocksize.

The panel (ii) of Figure 3 displays the related acceptance rate of the Metropolis-Hastings step. This acceptance rate is negatively related with the blocksize, and in this case if the acceptance rate drops below around 0.9, the efficiency of the sampler starts to deteriorate with blocksize. Note that the insights of EIS/IIS could be combined with MMS as well: Using the improved approximating density of EIS/IIS, the acceptance rate of MMS could be increased, allowing larger blocks to be sampled with lower inefficiencies as a result. See Liesenfeld and Richard (2008). This possibility is not pursued in this chapter.

Table 4: Comparing inefficiencies
Average relative numerical inefficiency Acceptance rate

	σ_ϵ	ϕ	σ_ξ	rate
GMS	27.72	143.42	332.10	
SSS	339.45	436.82	616.80	0.997
MMS	17.78	146.17	342.86	0.921

Given a blocksize for the MMS sampler, the inefficiencies of the samplers can be compared. It must be kept in mind that the mixture sampler, GMS, theoretically only provides a sample for the model with the approximating Gaussian mixture as observation density (instead of using the true observation density). Table 4 compares the average inefficiencies of the GMS to that of the SSS (effectively the MMS but with $k = 1$), and to the MMS with $k \approx 143$ in the last line of the table. In the present situation the single site sampler delivers clearly higher correlation in the chain, and the MMS is roughly on par with GMS. Notice however

that the behaviour of the samplers can also be affected by the sampling procedure for the parameters θ , by their parametrisation, and by the choice of those parameters in the DGP (cf. Bos and Shephard, 2006). In practical situations the SSS sampler can still be of interest, as it is simpler to implement than the MMS, considerably quicker (such that a longer sample could be drawn, subsampling the draws from the posterior to lower the correlation), and more flexible in combination with extended models, see also Section 4.4 for an example.

4 Estimating volatility models in practice

4.1 Describing return data of Goldman Sachs and IBM stock

With simulated data, the DGP is fully under the control of the researcher, and estimation of the volatility models can be performed with few surprises. This section however compares the estimation results for the alternative estimation approaches for two series of financial stock returns, to judge their performance in a practical setting.

The data considered here are the stock returns on the adjusted closing prices of Goldman Sachs (GS) and IBM, measured over the period 2005/1/3–2009/12/31, for a total of 1259 daily observations. These closing prices are available through Yahoo Finance. The period covers the financial crisis which has been affecting the economy since October 2007. For Goldman Sachs, this crisis led to large turmoil and great changes in the stock price. While the initial crisis delivered some large profits for Goldman Sachs (as they had invested in shorts on the sub-prime market before the major collapse, delivering them large gains), later the firm was hit by large losses and lawsuits. The uncertainty around the banking sector and the aftermath of the collapse of Lehman Brothers in September 2008 forced the firm to convert into a traditional bank, instead of the investment bank they used to be.

The second stock, IBM, is chosen for its entirely different field of operation. As a computer firm, the financial crisis will have had far less of an effect on its stock price. Indeed, Figure 4 displays in panel (i) the stock price of GS and IBM, with strong gains and equally strong losses for GS, whereas the IBM stock has moved comparatively little over this 5 year period.

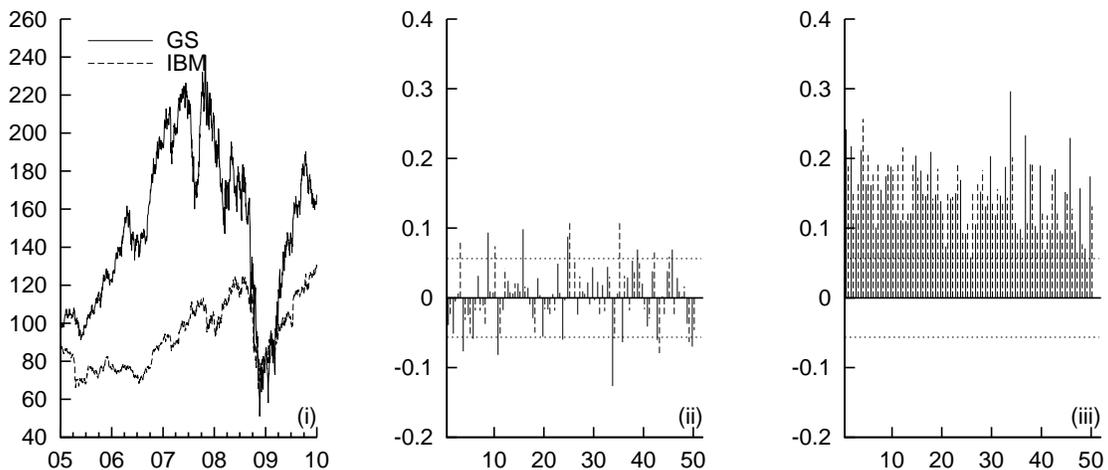


Figure 4: Daily Goldman Sachs and IBM stock price (panel (i)), ACF of returns (ii) and ACF of squared returns (iii), over the period 2005–2009.

Table 5: Moments of the percentage returns on Goldman Sachs and IBM stocks, over the period 2005-2009

	Goldman Sachs	IBM
n	1259	1259
Mean	0.04	0.03
Standard deviation	3.06	1.54
Skewness	0.44	0.06
Kurtosis	15.33	8.57

The center panel of Figure 4 displays the autocorrelation function (ACF) of the returns. Some of these autocorrelations are larger (in absolute value) than the 95% error bands. However, it should be kept in mind that these error bands are based on the assumption of underlying white noise data, whereas the variance of the returns is time-varying. These ACFs indicate that disregarding the autocorrelation in the return series is not at great odds with the data.

Panel (iii) of the figure displays the autocorrelation of the squared returns. As usual, the squares of the returns are strongly correlated through time, and a model for volatility could help in fitting the return series correctly. Table 5 displays the moments of the returns. The kurtosis of the returns is around 15.3 for GS vs. 8.6 for IBM. Though this could indicate either heavy tails in the return distribution or time-varying variance, for simplicity we limit ourselves to the pure Gaussian stochastic volatility model.

4.2 Estimating SV models

Table 6 displays the estimated parameters for the selection of methods. For each method, the optimal parameter is reported with its standard deviation. The results indicate a close correspondence between the group of IS, the group of Bayesian methods, and (mostly) with the QML approach. The only estimates standing out are the SMM estimates. The choice of auxiliary density here seems to matter for the outcomes, with the GARCH model with normal disturbances delivering results which seem to be closer to other estimates. These results are in line with findings of Andersen et al. (1999). Trials to incorporate a semi-nonparametric density (Gallant et al., 1997) were not successful, as such an auxiliary model could not be optimised in a stable manner for the current sample size.

In Section 3 the methods based on importance sampling were compared on their sampling efficiency through the standard deviations of the importance sampling weights. These standard deviations are reported in Table 7. IS is found to be less efficient in its likelihood estimation than EIS and IIS. These latter methods deliver virtually the same efficiency, where the difference seems due to simulation error, as it was in Section 3.3.

Table 8 compares the inefficiencies in the posterior samples of the parameters from the three Bayesian estimation methods. Note that for these estimates a sample of size 1,000,000 draws was collected after a burn-in period. The GMS method, through its approximation of the density of $\ln e_t^2$ by a mixture, delivers a well-mixing chain, in which a new draw from the full vector of volatilities is obtained through the simulation smoother. The SSS does not make the approximation, but instead draws a single $h_t|h_{t-1}, h_{t+1}, y_t, \theta$, incurring far higher correlation in the sampling chain. The MMS, on the other hand, by sampling blocks of $k \approx 150$ volatilities at once, delivers efficiency values at least as good as those of GMS, at acceptance rates dropping towards 74-82%.

Table 6: Parameter estimates for Goldman Sachs and IBM, 2005–2009

	$\hat{\sigma}_\epsilon$	$\mathbf{s}(\sigma_\epsilon)$	$\hat{\phi}$	$\mathbf{s}(\phi)$	$\hat{\sigma}_\xi$	$\mathbf{s}(\sigma_\xi)$
Goldman Sachs						
QML	1.1346	(0.304)	0.9975	(0.003)	0.1122	(0.023)
SMMn	1.6949	(0.353)	0.9861	(0.018)	0.1073	(0.058)
SMMt	1.2354	(0.071)	0.8257	(0.110)	0.3317	(0.123)
IS	0.8794	(0.210)	0.9976	(0.002)	0.1424	(0.022)
EIS	0.8878	(0.224)	0.9976	(0.002)	0.1427	(0.023)
IIS	0.8814	(0.211)	0.9976	(0.002)	0.1429	(0.023)
GMS	1.1460	(0.186)	0.9966	(0.003)	0.1406	(0.023)
SSS	0.7964	(0.223)	0.9982	(0.002)	0.1411	(0.022)
MMS	0.8829	(0.215)	0.9979	(0.003)	0.1378	(0.022)
IBM						
QML	0.8523	(0.231)	0.9961	(0.004)	0.0994	(0.026)
SMMn	1.2016	(0.116)	0.9535	(0.030)	0.2279	(0.084)
SMMt	0.8636	(0.040)	0.7714	(0.116)	0.3925	(0.124)
IS	1.1005	(0.128)	0.9730	(0.011)	0.2176	(0.034)
EIS	1.1087	(0.127)	0.9725	(0.011)	0.2203	(0.035)
IIS	1.1012	(0.127)	0.9727	(0.011)	0.2200	(0.035)
GMS	0.8840	(0.129)	0.9903	(0.008)	0.1771	(0.031)
SSS	1.0808	(0.136)	0.9792	(0.010)	0.2004	(0.035)
MMS	1.0662	(0.148)	0.9797	(0.010)	0.1936	(0.033)

Table 7: Standard deviations $\sigma(\omega)$ of importance sampling weights

	IS	EIS	IIS
Goldman Sachs	2.152	0.709	0.652
IBM	3.585	1.049	1.119

Table 8: Comparing inefficiencies and acceptance rates

	Inefficiencies			Acceptance
	σ_ϵ	ϕ	σ_ξ	rate
Goldman Sachs				
GMS	364.65	101.90	138.76	
SSS	1675.29	535.67	353.56	0.996
MMS	268.338	100.843	137.282	0.82
IBM				
GMS	342.38	191.44	167.48	
SSS	1336.78	429.93	384.73	0.993
MMS	162.59	123.68	164.21	0.74

4.3 Extracting underlying volatility

All estimation methods, except for the simulated method of moments, deliver an estimate of the smoothed volatility together with the parameter estimates. These smoothed states of H are plotted in Figure 5. The line for QML-based volatility is clearly distinct from the others,

running smoother than the other state estimates. The approximation of the non-linear SV model by the linear Gaussian state space model influences the extracted state considerably. Close to the QML estimate of the state runs the GMS volatility estimate, both for GS and IBM data. Remember that the GMS uses a 7-component mixture density to approximate the $\ln e_t^2$ density, where we did not perform any resampling to get rid of the remaining approximation error. The extracted state for GMS seems to run in-between the QML and the other estimates. Notice that the remaining methods were all theoretically exact, therefore it is no surprise that their state estimates are similar.

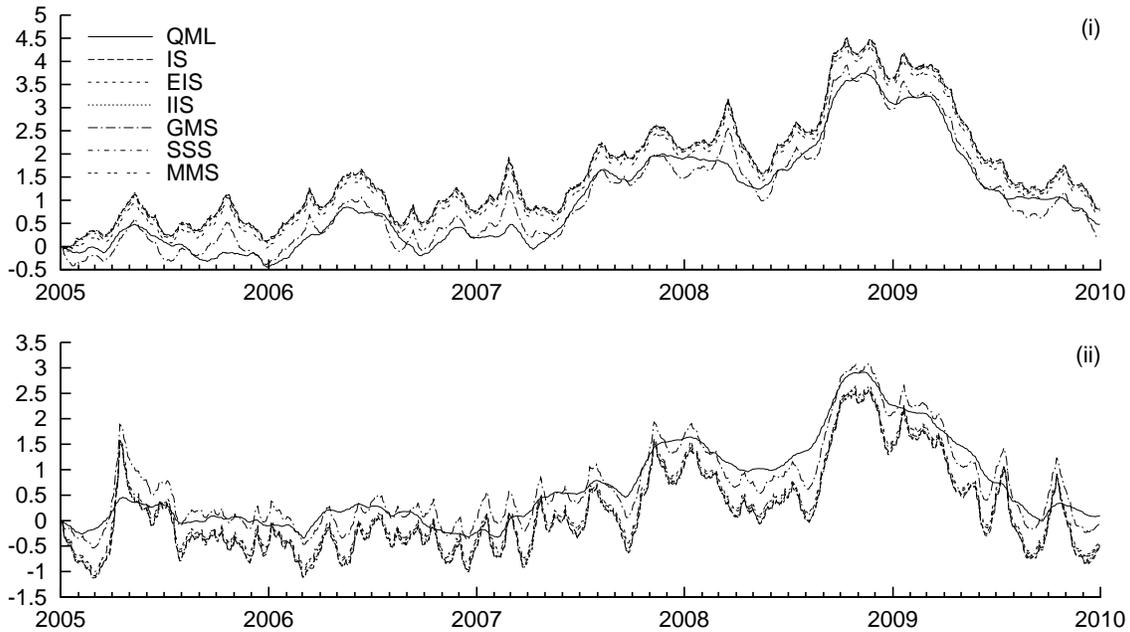


Figure 5: Extracted smoothed volatility sequences h_t for Goldman Sachs (panel (i)) and IBM (ii) returns.

4.4 Relating the returns in a bivariate model

Both Goldman Sachs and IBM operate in the same market, and their stocks are traded in the same country on the same stock exchange. Therefore one could expect that the returns on both stocks are related, as they both may react to the general atmosphere on the stock market. As an extension to the standard stochastic volatility model, this section looks into the correlation between the stock returns. For this purpose, the vector of bivariate returns at

time t can be modelled as

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim \mathcal{N}(\mathbf{0}, \Sigma_t), \\
\Sigma_t &= D_t R_t D_t', \\
D_t &= \begin{pmatrix} \sigma_{\epsilon_1} \exp(\frac{1}{2}h_{1t}) & 0 \\ 0 & \sigma_{\epsilon_2} \exp(\frac{1}{2}h_{2t}) \end{pmatrix}, & R_t &= \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}, \\
\rho_t &= \frac{\exp q_t - 1}{\exp q_t + 1}, \\
q_{t+1} &= q_t + v_t, & v_t &\sim \mathcal{N}(0, \sigma_q^2).
\end{aligned}$$

This dynamic correlation stochastic volatility setup (DCSV, Yu and Meyer, 2006; Bos and Gould, 2007) allows the returns to be correlated, with the correlation ρ_t changing stochastically over time in a manner similar to the behaviour of the stochastic volatility component. As correlation is not easily estimated from the data, especially when it is time-varying, a simple specification like the present transformed random walk is preferred.

With the likelihood or moments-based methods introduced above, estimating this extended model would require obtaining an approximating density for this full bivariate returns model, depending on the three underlying unobserved states. The Bayesian sampling methods however perform all sampling steps conditionally on other results. Adapting the model to include the correlation implies adding two steps to the sampling scheme, and adapting a third. Firstly, the parameter σ_q is sampled conditionally on all the states and other parameters, which is comparable to drawing parameters σ_ξ or σ_ϵ . Secondly, $Q = \{q_1, \dots, q_T\}$ is sampled conditionally on the parameters and the volatilities, in a manner fully comparable to the sampling of the H sequence. And lastly, the sampling of H has to be adapted to incorporate the information stemming from the correlation. With the GMS this is not easily performed, as a fully new mixture sampling procedure has to be devised (cf. Omori et al., 2007). With SSS sampling the volatilities $H_i|Y, H_j, \rho, \theta$, from the univariate volatility sequences conditioning on the parameters, correlation, second volatility and the data implies using slightly altered conditional densities instead of Equations (6)–(8). If one uses $p = \ln P(y_{1t}, y_{2t}; H_t, \rho_t, \theta) \equiv \ln P(y_1, y_2; h_1, h_2, \rho, \theta)$ (skipping the dependence on time from the notation), then

$$\begin{aligned}
u_j &= \frac{y_j}{\sigma_{\epsilon_j}} \exp\left(-\frac{1}{2}h_j\right), \\
p &= -\frac{1}{2} \left(h_1 + h_2 + \log \sigma_{\epsilon_1}^2 + \log \sigma_{\epsilon_2}^2 + \frac{u_1^2 - 2\rho u_1 u_2 + u_2^2}{1 - \rho^2} \right), \\
\dot{p}_1 &= -\frac{1}{2} \left(1 - \frac{u_1^2 - \rho u_1 u_2}{1 - \rho^2} \right), \\
\ddot{p}_1 &= -\frac{1}{2} \frac{u_1^2 - \frac{1}{2}\rho u_1 u_2}{1 - \rho^2},
\end{aligned}$$

where \dot{p}_1 and \ddot{p}_1 are the first and second derivatives of p with respect to the first volatility h_1 . Derivatives with respect to h_2 are similar.

With these changes to the derivatives, the SSS is implemented on this extended model without further problems. The sampler can be started using the parameter estimates from the univariate samplers as initial values, initialising the correlation sequence at the fixed full-sample correlation. The prior for parameter σ_q was taken as IG-1(r, a), with expectation and

standard deviation equal to 0.05. Results on the parameter estimates are reported in Table 9, with extracted volatilities and the correlation in Figure 6.

Table 9: Bivariate Goldman Sachs and IBM volatility/correlation results

Series	Parameter	$\hat{\theta}$	$\mathbf{s}(\theta)$	Inefficiency
GS	σ_ϵ	0.9993	(0.027)	15.23
	ϕ	0.9977	(0.002)	29.39
	σ_ξ	0.1236	(0.019)	297.51
IBM	σ_ϵ	1.0028	(0.027)	46.35
	ϕ	0.9790	(0.009)	259.38
	σ_ξ	0.1953	(0.031)	469.18
Correlation	σ_q	0.0241	(0.006)	3551.96

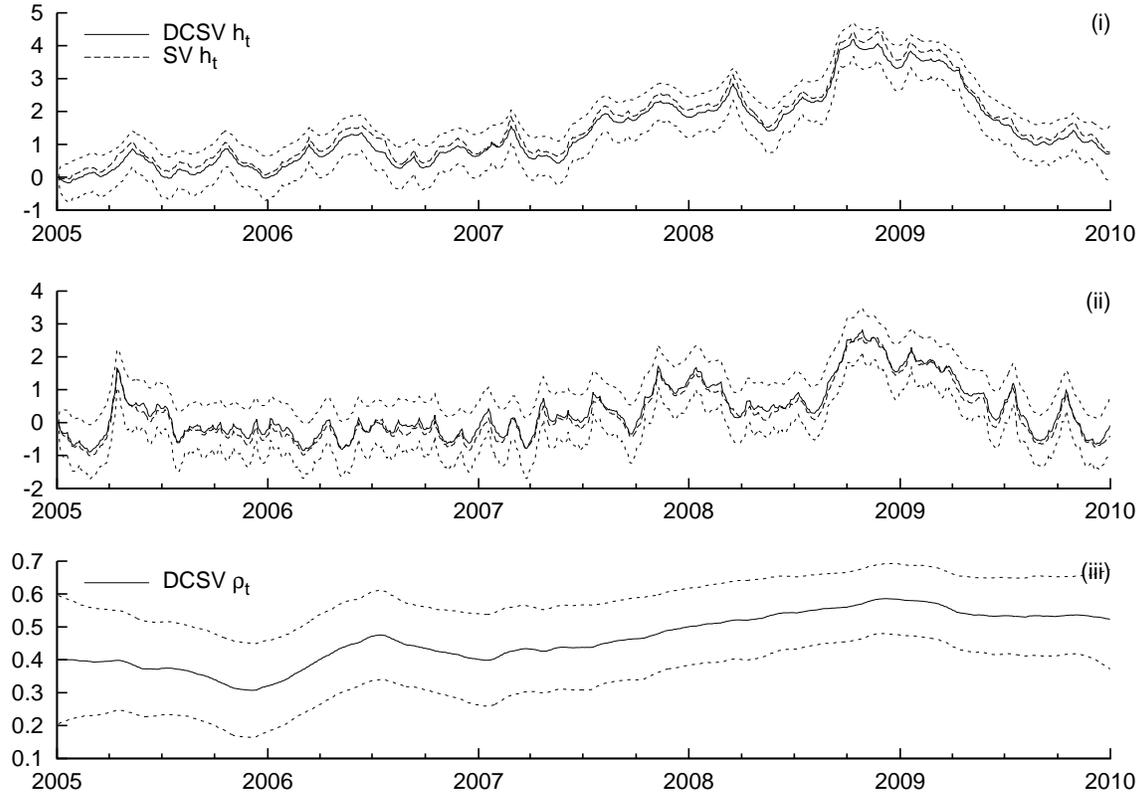


Figure 6: Extracted smoothed volatility sequences h_t for Goldman Sachs (panel (i)) and IBM stock returns (ii) together with correlation ρ_t (iii), with 95% confidence bands, using the bivariate DCSV model, and comparing with the univariate SV model.

The parameter estimates of Table 9 are very similar to the estimates found in for SSS in Table 6. The new parameter σ_q is estimated at 0.024, clearly below the prior mean of 0.05. In a plot of the posterior density of σ_q (not reported here) it is found that the mass of the posterior indeed is concentrated at the lower end of the range of the prior density, and hence the data is informative on the volatility of the correlation. The inefficiency measures are in

general better than corresponding values for the SSS sampler in Table 8. It is commonly found that an alteration in the specification of the model can indeed alter the efficiency of the SSS strongly.

Figure 6 compares the volatilities as extracted from the univariate model with the results from the bivariate DCSV model, in panels (i) and (ii) for Goldman Sachs and IBM respectively. Taking the second return series into the model does not alter the estimate of volatility, as the univariate return series themselves are sufficiently informative on the matter. Panel (iii) displays the extracted correlation. Here it is seen how correlation between stock returns for this bank and this information technology company are low, at around 0.4 at the start of the sample, to increase with the onset of the crisis reaching a maximum of $\rho_t = 0.59$ in December 2008, within a month's distance from the moment when volatility peaked. Even though these two companies are very dissimilar in their field-of-operation, they both suffered from the financial crisis. In bad times the co-movement of their stocks increases slightly, which is valuable information for investors who plan on building a portfolio out of these assets.

5 Discussion and concluding remarks

This chapter set out as an investigation into the alternative estimation methods for basic stochastic volatility models. A range of five classical and three Bayesian methods was described, using a common notation. While the simulated method of moments approach was found to stand apart, the others were found to share the basic idea of approximating the (non-linear) observation equation (1) through a linear model, fitting the first two derivatives of the log-densities using (with most methods) a single parameter for the mean and a second for the variance at each time period t .

Both QML and GMS made the approximation without taking the most likely volatility sequence into account, and without correcting for the approximation either. For QML this was seen to lead to a bias in the parameter estimates. For GMS, the approximation was far more precise using a mixture of several normal densities, and the problem of the bias was mostly solved for parameter estimation. The extracted state however was seen to diverge from estimates of other, theoretically exact methods.

The basic IS method improves on QML by extracting the smoothed state of the QML approximation, and adapting the approximating model iteratively to take this most probable state of the volatility into account. EIS and IIS continue in this same direction by taking a full sample of probable volatilities instead of only a single smoothed state. This results in a better approximating model and less variance of the importance weights in the evaluation of the likelihood. EIS depended on deriving analytically the integrating constant of the auxiliary density, taking information on past volatilities into account when optimising the parameters of the auxiliary model. For IIS this integrating constant was not needed, as the information was taken along in the sampling of the volatilities, applying the simulation smoother for this purpose.

The remaining Bayesian approaches can be seen as also taking the IS approach as a starting point. Either for a single state, or for a block of volatilities, the approximating model is formed as in the IS method, and a single sample is drawn. Instead of using the importance weights of IS, now a Metropolis-Hastings-within-Gibbs acceptance probability is calculated. After deciding upon acceptance, attention shifts to the next block of volatilities.

Simulation results were found to be in line with expectation, in that indeed QML can

lead to quick but biased estimates, and that the IS, EIS and IIS end up at virtually the same estimators. The advantage of highest efficiency is shared between EIS and the novel IIS, with the latter model requiring less input from the researcher in setting up the importance sampler. The SMM approach depends on the specification of the approximating model, though the mode of the estimates seems to be closely in line with the values of the DGP. It could be expected that SMM works better for larger samples, where a richer auxiliary model can be estimated with high precision. For the Bayesian approaches, SSS is found to be quick, and theoretically exact (in contrast to GMS if no resampling is applied), but the correlation in the resulting posterior sample is high. MMS, taking larger blocks of volatility for sampling at once, leads to less inefficiency in the posterior sample.

Extracting volatility for Goldman Sachs and IBM returns over the period 2005–2009 indicates that the estimates of parameters and volatility of both return series correspond largely between the methods. Mid 2007, when the first signs of the sub-prime crisis emerged, volatility started to creep up slowly. In September 2008 the big crisis hit the financial markets, and volatilities of both series went up further.

In an extension, the two stocks were modelled jointly, allowing for time-varying correlation between the returns. This was easily accommodated for in the SSS. For the volatilities, little difference was found compared to the univariate results. The extracted correlation indicated that these rather distinct companies display stock returns which are correlated at a level of around 0.4-0.6, where the amount of correlation is positively related to the uncertainty in the markets.

The GARCH model has the advantage that it can be estimated using a single, well-defined and simple, estimation method. There are however many variants of the GARCH model, which can make it cumbersome to choose the best specification for the data at hand. For the SV model, there is far less discussion about the preferred specification, and only the difficulty of estimation seems to hamper a larger uptake of the model in practical research. In this chapter it was found that the estimation methods have many commonalities. These commonalities can help both in understanding in detail the algorithms, and also in the practical implementation and application of the stochastic volatility model.

References

- Abramowitz, M. and I. A. Stegun (1972). *Handbook of Mathematical Functions*. New York: Dover Publications.
- Andersen, T. G., H.-J. Chung, and B. E. Sorensen (1999). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics* 91(1), 61–87.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.
- Bollerslev, T. (2008). Glossary to ARCH (GARCH). Research paper 2008-49, CREATES.
- Bos, C. S. and P. Gould (2007). Dynamic correlations and optimal hedge ratios. Discussion Paper TI 07-025/4, Tinbergen Institute.

- Bos, C. S. and N. Shephard (2006). Inference for adaptive time series models: Stochastic volatility and conditionally Gaussian state space form. *Econometric Reviews* 25(2–3), 219–244.
- Broto, C. and E. Ruiz (2004). Estimation methods for stochastic volatility models: A survey. *Journal of Economic Surveys* 18(5), 613–649.
- Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika* 81(3), 541–553.
- Carter, C. K. and R. Kohn (1997). Semiparametric Bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society, Series B* 59(1), 255–268.
- Clark, P. K. (1973). A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica* 41(1), 135–156. Reprinted in Shephard (2005).
- De Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika* 82(2), 339–350.
- Doornik, J. A. (2009). *Ox6: An Object-Oriented Matrix Programming Language*. London: Timberlake Consultants Ltd.
- Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* 84(3), 669–684.
- Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89(3), 603–615.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1008.
- Gallant, A. R., D. Hsieh, and G. Tauchen (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* 81(1), 159–192.
- Gallant, A. R. and D. W. Nychka (1987). Semiparametric maximum likelihood estimation. *Econometrica* 55(2), 363–390.
- Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory* 12, 657–681.
- Ghysels, E., A. C. Harvey, and É. Renault (1996). Stochastic volatility. In *Handbook of Statistics*, Volume 14 of *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies* 61(2), 247–264.
- Jungbacker, B. and S. J. Koopman (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika* 94(4), 827–839.

- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65(3), 361–393.
- Koopman, S. J., N. Shephard, and D. Creal (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics* 149(1), 2–11.
- Koopman, S. J., N. Shephard, and J. A. Doornik (2008). *SsfPack 3.0: Statistical Algorithms for Models in State Space Models*. London: Timberlake Consultants Ltd.
- Liesenfeld, R. and J.-F. Richard (2006). Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews* 25(2–3), 335–360.
- Liesenfeld, R. and J.-F. Richard (2008). Improving MCMC, using efficient importance sampling. *Computational Statistics and Data Analysis* 53(2), 272–288.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance* 7(1), 77–91.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59(2), 347–370.
- Nguyen, T. M. (2007). A new efficient algorithm for the analysis with non-linear and non-Gaussian state space models. Master’s thesis, Dept. of Econometrics & O.R., FEWEB, VU University Amsterdam.
- Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics* 140, 425–449.
- Richard, J.-F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics* 141(2), 1385–1411.
- Sandmann, G. and S. J. Koopman (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics* 87, 271–301.
- Shephard, N. (Ed.) (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84(3), 653–667.
- Smith, A. F. M. and A. E. Gelfand (1992, May). Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician* 46(2), 84–88.
- 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*, pp. 203–226. Amsterdam: North-Holland. Reprinted in Shephard (2005).
- Yu, J. and R. Meyer (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews* 25(2–3), 361–384.