Bayesian Estimation of the Heston Stochastic Volatility Model

Article · April 2003

DOI: 10.1007/978-3-642-55537-4_78

2 authors:

Sylvia Frühwirth-Schnatter
Wirtschaftsuniversität Wien
102 PUBLICATIONS **3,229** CITATIONS
SEE PROFILE

Leopold Sögner
Institut für Höhere Studien
48 PUBLICATIONS **351** CITATIONS
SEE PROFILE

Bayesian Estimation of the Heston Stochastic Volatility Model

Sylvia Frühwirth-Schnatter[a] and Leopold Sögner[b]

[a] sylvia.fruehwirth-schnatter@jku.at, Department of Applied Statistics, Johannes Kepler University Linz, Altenberger Strasse 69, A-4040 Linz, Austria.

[b] soegner@ibab.tuwien.ac.at, Department of Managerial Economics and Industrial Organization, Vienna University of Technology, Theresianumgasse 27, A-1040 Vienna, Austria.

March 27, 2003

Abstract

The goal of this article is an exact Bayesian analysis of the Heston (1993) stochastic volatility model. We carefully study the affect different parameterizations of the latent volatility process and the parameters of the volatility process have on the convergence and the mixing behavior of the sampler. We apply the sampler to simulated data and to some DM/US$ exchange rate data.

1 Introduction

Modern option pricing theory often uses models where a closed form expression for the price of an European call option is available and some characteristic features of financial time series can be incorporated into the model. A first yardstick when considering option pricing theory is the Black and Scholes (1973) model, which models the log of an asset price as the solution to the stochastic differential equation

$$dy^\star(t) = (\mu + \beta\sigma(t)^2)dt + \sigma(t)dW(t), \tag{1}$$

where $W(t)$ is a standard Brownian motion and volatility is a constant, i.e. $\sigma(t) = \sigma$. The model implies that aggregate returns $y_n = y^\star(t_n) - y^\star(t_{n-1})$ are normal with mean $(\mu+\beta\sigma^2)\Delta$ and constant variance $\sigma^2\Delta$, where $\Delta = (t_n - t_{n-1})$. Despite the fact that a closed form solution for an European call option is well known, this model suffers from the fact that it does not incorporate important stylized features of financial time series, such as heavy tails, skewness and volatility clustering. With the goal to improve model (1), *stochastic volatility*

models have been introduced. These models share the property that the volatility term $\sigma(t)$ follows a stochastic process. Most of these extensions are based on diffusion type models, e.g. an *Gaussian Ornstein-Uhlenbeck-processes* (OU) for the log volatility (see Andersen and Lund (1997), Hull and White (1987)) or constant elasticity models (see Heston (1993), Meddahi and Renault (2000)), or alternatively, based on Ornstein-Uhlenbeck processes driven by a Lévy process (see Barndorff-Nielsen and Shephard (2001)).

Empirical estimation of a stochastic volatility model means estimating the unknown parameters θ of the underlying asset process from discrete observations. The measurement equation specifies the distribution of the aggregated returns $y_n := y^\star(t_n) - y^\star(t_{n-1})$ given the latent volatility process $\sigma^2(t)$. By integrating this volatility process over time t one obtains the integrated volatility $\sigma^{2\star}(t) := \int_0^t \sigma^2(s)ds$, and the conditional distribution of $y_n|h_n$ is normal, where

$$h_n := \sigma^{2\star}(t_n) - \sigma^{2\star}(t_{n-1}) \ . \tag{2}$$

Regarding to an estimation of the model parameters, the stochastic models mentioned above suffer from the fact that the likelihood $L(\theta)$, where θ are the model parameters, is not available. This does not allow to estimate the parameters by means of the maximum likelihood method. Alternative methods to investigate continuous time models documented in literature are methods based on simulations (e.g. Gallant and Tauchen (1996)), generalized methods of moments (e.g. Barndorff-Nielsen and Shephard (2002), Bollerslev and Zhou (2002), Hansen and Scheinkman (1995) and Hubalek and Tomkins (2001)), non-parametric techniques (e.g. Stanton (1997)), the approximation of transition densities (e.g. Ait-Sahalia (2002)) and Bayesian methods (e.g. Elerian *et al.* (2001), Eraker (2001), Papaspiliopoulos *et al.* (2001)).

The focus of the present paper is to discuss Bayesian estimation of an augmented version of the Heston (1993) stochastic volatility model. This paper is organized as follows: Section 2 provides a brief description of the Heston (1993) model and specifies the conditional distributions. In section 3 we describe how *Markov-chain Monte Carlo* (MCMC) methods can be applied to derive the model parameters. Section 4 presents an analysis with simulated data and empirical data of DM/US$ exchange rates. We show that the sampling properties can be improved substantially if (i) a strong prior is put on the mean volatility parameter in a one factor setup or to the sum of these mean volatility parameters in multi factor setup and (ii) a partially centered parameterization of the latent processes is used.

2 The Heston model

In this paper we investigate the following model, based on Heston (1993):

$$dy^\star(t) = (\mu + \beta\sigma(t)^2)dt + \sigma(t)dW_0(t) \ , \tag{3}$$

where actual volatility is given by a superposition of k independent square root processes:

$$\begin{aligned} \sigma^2(t) &= \sum_{i=1}^{k} \sigma_i^2(t) \ , \\ d\sigma_i^2(t) &= \lambda_i(\alpha_i - \sigma_i^2(t))dt + \tau_i \sigma_i(t) dW_i(t) \ , \end{aligned} \tag{4}$$

where $i = 1, \ldots, k$; λ_i is often interpreted as the speed of mean reversion, α_i is the mean of the instantaneous volatility process and τ_i is the volatility parameter; $W_j(t), j = 0, 1, \ldots, k$, are independent Brownian motions.

For the underlying stochastic volatility model, the parameters μ, β, α, λ, and τ, where $\alpha = (\alpha_1, \ldots, \alpha_k)$, etc. are unknown. These parameters – abbreviated by θ – have to be inferred from the data available (for the parameters of component i we use $\theta_i := (\alpha_i, \lambda_i, \tau_i)$). For financial time series these are (equidistant) observations of asset returns (y_n), where the step-width will be denoted by Δ in the further analysis. The index n is used to abbreviate the points of time $\ldots, (n-1)\Delta, n\Delta, \ldots$, i.e. if N yields are observed, then $Y = (y_n)_{n=1}^N$ within the time span $[\Delta, T]$, where $T = N\Delta$. The same notation is also applied to the instantaneous volatilities $\sigma_i^2(t)$: $(\sigma_{i,n}^2)_{n=0}^N$ are $N+1$ instantaneous volatilities in the time span $[0, T]$ observed at equidistant intervals Δ; $\sigma_n^2 = \sum_{i=1}^k \sigma_{i,n}^2$ from (4). In this paper we abbreviate increases in Brownian motion by $\omega_j(s,t) = W_j(t) - W_j(s)$, $j = 0, 1, \ldots, k$. If a constant step width $\Delta = t - s$ is used, then simply $\omega_{j,n}(\Delta) = W_j(n\Delta) - W_j((n-1)\Delta)$ will be used. Note that $\omega_{j,n}(\Delta)$ has the same distribution as $\sqrt{\Delta}\omega_{j,n}(1)$.

The goal of the following paragraphs is the specification of the conditional distributions, such that an exact Bayesian analysis can be performed. Let us define integrated volatility:

$$\sigma^{2\star}(t) := \sum_{i=1}^{k} \sigma_i^{2\star}(t); \quad \sigma_i^{2\star}(t) := \int_0^t \sigma_i^2(u)du \ , \tag{5}$$

and the increases in integrated volatility h_n from $t = (n-1)\Delta$ to $t = n\Delta$,

$$h_n := \sum_{i=1}^{k} h_{i,n}; \quad h_{i,n} := \sigma_i^{2\star}(n\Delta) - \sigma_i^{2\star}((n-1)\Delta); \ \forall i = 1, \ldots, k \ . \tag{6}$$

Conditionally on these increases in integrated volatility, h_n, the asset returns are normally distributed, i.e.

$$y_n \sim \mathcal{N}(\mu\Delta + \beta h_n, h_n) \ . \tag{7}$$

The conditional distribution of the increases in integrated volatility fulfills

$$\pi(h_n | h_{n-1}, \sigma_{n-1}^2; \theta) = \Pi_{i=1}^k \pi(h_{i,n} | \sigma_{i,n-1}^2; \theta_i) \ . \tag{8}$$

In the ongoing analysis we abbreviate $(h_{1,n},\ldots,h_{i,n},\ldots,h_{k,n})$ and $(\sigma^2_{1,n},\ldots,\sigma^2_{i,n},\ldots,\sigma^2_{k,n})$ by H_n and S_n respectively. Although the conditional density $\pi(h_{i,n}|\sigma^2_{i,n-1};\theta_i)$ cannot be derived analytically, the Fourier transform is available as described in Lamberton and Lapeyre (1996)[pp. 132].
The conditional distribution of the instantaneous volatility process $\pi(\sigma^2_n|\sigma^2_{n-1};\theta)$, factorizes into

$$\pi(\sigma^2_n|\sigma^2_{n-1};\theta) = \Pi^k_{i=1}\pi(\sigma^2_{i,n}|\sigma^2_{i,n-1};\theta_i) \ , \tag{9}$$

where $\pi(\sigma^2_{i,n}|\sigma^2_{i,n-1};\theta_i)$ is a non-central χ^2 distribution (see Appendix B). Last but not least a stationary law of $(\sigma^2_i(t))$ exists if the parameters satisfy: $\lambda_i > 0$ and $2\lambda_i\alpha_i/\tau^2_i \geq 1$. The stationary law is a Gamma distribution. For $b_i := 2\lambda_i/\tau^2_i$ and $a_i := b_i\alpha_i$, the marginal densities of the instantaneous volatilities are given by:

$$\pi(\sigma^2_i(t)) = \frac{b_i}{\Gamma(a_i)}(\sigma^2_i(t))^{a_i-1}\exp(-b_i\sigma_i(t)^2) \ . \tag{10}$$

Note that by the Fubini theorem and the ergodic theorem, $\mathbb{E}(h_n) = \mathbb{E}(\sigma^2)\Delta = \alpha\Delta$ and $\mathrm{VAR}\,(h_n) = \mathrm{VAR}\,(\sigma^2)\Delta^2 = \alpha\tau^2\Delta^2/(2\lambda)$.

Remark 1 *Despite the fact that only the conditional distributions are required to apply MCMC, the restriction* $2\lambda_i\alpha_i/\tau^2_i \geq 1$ *– implying stationary instantaneous volatility processes with marginal gamma law – is plausible from an economic point of view. If* $2\lambda_i\alpha_i/\tau^2_i \geq 1$ *is met, the asset yields will remain stochastic since* $\sigma^2_i(t) > 0$ *with probability one.*

3 Bayesian estimation of the Heston model

Parameter estimation for stochastic volatility models is known to be a difficult problem (see e.g. Shephard (1996)). The problem stems from the fact that the conditional distribution $f(y_n|h_n,\mu,\beta)$ of the observed returns y_n depends on the unobservable integrated volatility $\sigma^{2\star}(t)$. To apply the method of maximum likelihood, where

$$L(\theta) = f(y_1,\ldots,y_N|\theta) = \prod_{n=2}^{N} f(y_n|y_1,\ldots,y_{n-1},\theta) \ , \tag{11}$$

the closed forms of the conditionals $f(y_n|.)$ are required. To derive $h_n|\sigma^2_{n-1}$ we have to invert the characteristic function. To derive $f(y_n|.)$, the integrated volatilities have to be integrated out in (7). Since this integration is practically impossible, the application of maximum likelihood becomes infeasible.
The goal of this section is to describe how the model parameters θ can be derived by using MCMC methods. Although $f(y_1,\ldots,y_N|\theta)$ is not available, we know from (7) that the

yields are conditionally normal. Furthermore, we know the conditionals $\pi(h_{i,n}|\sigma^2_{i,n-1};\theta)$ and $\pi(\sigma^2_{i,n}|\sigma^2_{i,n-1};\theta)$. To use these densities – e.g. to sample from these densities – we simply include $h_{i,n}$ and $\sigma^2_{i,n}$ in the set of unknowns (*data augmentation*). This set of additional unknown parameters will be abbreviated by X in the further analysis.

3.1 Parameterization of the model

To sample from the joint posterior distribution

$$\pi(X,\theta|Y) \propto f(Y|X,\theta)\pi(X|\theta)\pi(\theta) \ , \tag{12}$$

the "complete data" likelihood $f(Y|X,\theta)$ as well as the "prior density" $\pi(X|\theta)$ of the joint distribution of the latent variables X under θ has to be known explicitly. $\pi(\theta)$ is the prior of θ. The "complete data" likelihood $f(Y|X,\theta)$ is easily obtained from (7) as the product of $(N-1)$ densities from a normal distribution,

$$f(Y|X,\theta) = \prod_{n=2}^{N} f(y_n|\mu\Delta + \beta h_n, h_n) \ . \tag{13}$$

From section 2 we already know that the densities $f(y_n|h_n,\mu,\beta)$, $\pi(h_n|h_{n-1},\sigma_{n-1};\theta)$ and $\pi(\sigma^2_n|\sigma^2_{n-1};\theta)$ are available, such that an exact Bayesian analysis can be applied to the parameter estimation problem.
Remember that $H_n = (h_{1,n},\ldots,h_{i,n},\ldots,h_{k,n})$ and $S_n = (\sigma^2_{1,n},\ldots,\sigma^2_{i,n},\ldots,\sigma^2_{k,n})$. Let us define $X_1 := (H_1,\ldots,H_n,\ldots,H_N)$ and $X_2 = (S_0,S_1,\ldots,S_n,\ldots,S_N)$. A well known fact with MCMC methods, is that the convergence behavior of the Markov chain depends on the parameterization of the latent processes (see e.g. Papaspiliopoulos *et al.* (2002)). With respect to our problem this implies that the rate convergence depends on the parameterization of $(h_{i,n})$ and $(\sigma^2_{i,n})$.

Parameterization of the latent volatility processes:
(i) Centered Parameterization [C]: The natural candidates to describe the latent processes are the integrated volatilities $X_1^C := (H_{1,C},\ldots,H_{n,C},\ldots,H_{N,C})$ and the instantaneous volatilities $X_2^C = (S_{0,C},S_{1,C},\ldots,S_{n,C},\ldots,S_{N,C})$. This version to express X_1 and X_2 will be called *centered version*, where

$$h_{i,n} \ = \ id(h_{i,n,C}) \ , \ \sigma^2_{i,n} = id(\sigma^2_{i,n,C}) \ , \ i = 1,\ldots,k \ . \tag{14}$$

(ii) Non-centered Parameterization [NC]: Following Papaspiliopoulos *et al.* (2001) and Roberts and Stramer (2001), another version – called the *non-centered* version X_1^{NC}, X_2^{NC} – uses a starting value $\sigma^2_{i,0,NC} = \sigma^2_{i,0}$ and increases in Brownian motion $\omega_{i,n}(1)$, to calculate $(h_{i,n,NC})$ and $(\sigma^2_{i,n,NC})$ by means of the model parameters θ_i. When simulating $(\sigma^2_{i,n,NC})$ by

means of the Euler scheme, the map $\omega_{i,n}(1) \mapsto \sigma^2_{i,n-1,NC}\exp(-\lambda_i\Delta) + (1-\exp(-\lambda_i\Delta))\alpha_i + \tau\sigma_{i,n-1,NC}(\Delta)^{0.5}\omega_{i,n}(1)$, where $\omega_{i,n}(1) \sim \mathcal{N}(0,1)$, provides us with an (approximation of the) instantaneous volatility process (for a proof the reader is referred to the Appendix A). Since this map is one-to-one it is indeed possible to derive the process $(\omega_{i,n}(1))$ from the instantaneous volatilities $(\sigma^2_{i,n,NC})$. An integration of $\sigma^2_{i,n,NC}$ over the time interval Δ provides us with $h_{i,n,NC}$. Thus, with the non-centered version:

$$\begin{aligned} h_{i,n,NC} &:= \alpha_i\Delta + (\sigma^2_{i,n-1,NC} - \alpha_i)\frac{1-\exp(-\lambda_i\Delta)}{\lambda_i} + \tau_i\sigma_{i,n-1,NC}(\Delta)^{1.5}\tilde{\omega}_{i,n}(1) \\ \sigma^2_{i,n,NC} &:= \sigma^2_{i,n-1,NC}\exp(-\lambda_i\Delta) + (1-\exp(-\lambda_i\Delta))\alpha_i + \tau_i\sigma_{i,n-1,NC}(\Delta)^{0.5}\omega_{i,n}(1)\ . \quad (15) \end{aligned}$$

The reader should note that we use $\omega_{i,n}(1)$ and $\tilde{\omega}_{i,n}(1)$ in the above expressions. If Δ approaches 0, then the equality $\omega_{i,n}(1) = \tilde{\omega}_{i,n}(1)$ has to hold. However, for observations on a discrete grid, $h_{i,n}$ remains a random variable given $\sigma^2_{i,n-1}$. To account for this fact we use $\omega_{i,n}(1)$ and $\tilde{\omega}_{i,n}(1)$. In the non-centered version the returns and (h_n) depend on the parameters θ. This is not the case when the centered version is applied.
Therefore, with the non-centered parameterization we use $h_{i,n} = h_{i,n,NC}$ and $\sigma^2_{i,n} = \sigma^2_{i,n,NC}$ to describe the increases in the integrated volatilities and the instantaneous volatilities. For completeness let $\tilde{O}_n = (\tilde{\omega}_{1,n}, \ldots, \tilde{\omega}_{k,n})$ and $O_n = (\omega_{1,n}, \ldots, \omega_{k,n})$, then $X_1^{NC} = (\tilde{O}_1, \ldots, \tilde{O}_N)$ and $X_2^{NC} = (S_{0,NC}, O_1, \ldots, O_N)$.
(iii) Partially-centered Parameterization [PC]: In the Section 4 we shall observe that neither the centered version, nor the non-centered version will provide us with acceptable convergence properties and reasonable acceptance rates. Regarding to this issue, Papaspiliopoulos *et al.* (2002) present and discuss the opportunity to construct a *partially centered* parameterization. This can be obtained by a convex combination of centered and non-centered terms. With respect to our problem a partially centered version (PC) can be constructed by "$X_1^{PC} = (1-\nu)X_1^C + \nu X_1^{NC}$" and "$X_2^{PC} = (1-\nu)X_2^C + \nu X_2^{NC}$", where $\nu \in [0,1]^k$. I.e.

$$\begin{aligned} h_{i,n,PC} &= (1-\nu_i)h_{i,n,C} + \nu_i h_{i,n,NC} \quad (16) \\ \sigma^2_{i,n,PC} &= (1-\nu_i)\sigma^2_{i,n,C} + \nu_i\sigma^2_{i,n,NC}\ . \quad (17) \end{aligned}$$

Here we get the increases in the integrated volatilities and the instantaneous volatilities by $h_{i,n} = h_{i,n,PC}$ and $\sigma^2_{i,n} = \sigma^2_{i,n,PC}$. $X_1^{PC} = X_1^C \cup X_1^{NC}$ and $X_2^{PC} = X_2^C \cup X_2^{NC} \backslash S_{0,NC}$.[1]
Note that with $\nu_i = 1$ we approach at the non-centered parameterization for component i, while with $\nu_i = 0$ the parameterization degenerates to the centered version. After we have discussed how volatilities can be represented in our analysis, let us investigate the "prior"

[1]Since one common starting value for the instantaneous volatility process will be used, $S_{0,NC}$ is subtracted in X_2^{PC}.

$\pi(X|\theta)$, where $X = (X_1^j, X_2^j)$ with $j \in \{C, NC, PC\}$. From the above definitions $\pi(X|\theta)$ is nothing more than the product of N conditional densities

$$\pi(H_n, S_n^2|H_{n-1}, S_{n-1}^2; \theta) = \Pi_{i=1}^k \pi(h_{i,n}, \sigma_{i,n}^2|h_{i,n-1}, \sigma_{i,n-1}^2; \theta_i) \ , \tag{18}$$

and the prior

$$\pi(S_0|\theta) = \Pi_{i=1}^k \pi(\sigma_{0,i}^2|\theta_i) \ . \tag{19}$$

Note that $h_{i,n}$ and $\sigma_{i,n}^2$ are conditionally independent, i.e.

$$\pi(h_{i,n}, \sigma_{i,n}^2|h_{i,n-1}, \sigma_{i,n-1}^2; \theta_i) = \pi(h_{i,n}|\sigma_{i,n-1}^2; \theta_i)\pi(\sigma_{i,n}^2|\sigma_{i,n-1}^2; \theta_i) \ . \tag{20}$$

From section 2 we already known that in the case of the centered parameterization ($\nu_i = 0$) the first density in (20) can be reconstructed by means of Fourier inversion while the second density is given by a non-central χ^2 law (see Appendix B [equations (31) and (32)]). For the non-centered parameterization ($\nu_i = 1$), $h_{i,n,NC} \sim \mathcal{N}(a_{i,n,H}, A_{i,n,H}^2)$ and $\sigma_{i,n,NC}^2 \sim \mathcal{N}(a_{i,n,S}, A_{i,n,S}^2)$, where

$$\begin{aligned}
a_{i,n,H} &:= \alpha_i\Delta + (\sigma_{i,n-1,NC}^2 - \alpha_i)\frac{1 - \exp(-\lambda_i\Delta)}{\lambda_i} \ , & (21)\\
A_{i,n,H} &:= \tau_i\sigma_{i,n-1,NC}(\Delta)^{1.5} \ , & (22)\\
a_{i,n,S} &:= \sigma_{i,n-1,NC}^2 \exp(-\lambda_i\Delta) + (1 - \exp(-\lambda_i\Delta))\alpha_i \ , & (23)\\
A_{i,n,S} &:= \tau_i\sigma_{i,n-1,NC}(\Delta)^{0.5} \ . & (24)
\end{aligned}$$

For the partially-centered parameterization the Fourier transforms for both conditionals $\pi(h_{i,n,PC}|\sigma_{i,n-1,PC}^2; \theta_i)$ and $\pi(\sigma_{i,n,PC}^2|\sigma_{i,n-1,PC}^2; \theta_i)$ are available and described in Appendix B [equations (35) and (36)].

Parameterization θ: Working with the reference parameterization [P1] $\theta_R = (\alpha, \lambda, \tau)$ results in very strong correlation of λ and τ, if λ_i and τ_i are not blocked. The parameterizations $\alpha, \tau, \lambda/\tau$ [P2], $\alpha, \lambda, \tau^2/\lambda$ [P3] $\alpha, \lambda, (\alpha\tau^2)/(2\lambda)$ [P4] do a better job, for the factor where λ_i is large. The best performance was attained with a mixture of these two parameterizations. For the component where λ_i is small the reference prior perform better. Therefore, we update θ as follows: (i) Update of $\alpha = (\alpha_1, \ldots, \alpha_k)'$: We use a random walk proposal on $\sum_i \alpha_i$, i.e. a candidate for this sum is generated from $(\sum_i \alpha_i) \cdot \exp(c_\alpha\zeta)$; $c_\alpha = 0.2$, $\zeta \sim \mathcal{N}(0, 1)$. Then, for the remaining $k - 1$ parameters of α a random walk proposal is applied to α_i. α is updated in one block. Update of λ,τ: Finally, we update λ and τ factor-wise, where a mixture of the reference parameterization [P1] and [P4] is used for the low λ_i factor and a mixture of [P3] and [P4] is applied to the high λ_i component. For the update we use random walk proposals where the updating step for λ_i and τ_i is blocked.[2]

[2]Replacing [P3] or [P4] by [P2] in the above setups keeps the performance of the sampler unchanged.

In this paper we restrict to the estimation of $(h_{i,n},\ \sigma^2_{i,n})$, and the parameters α, λ and τ. β and μ are fixed at 0. Thus we can construct an ergodic Markov chain to obtain samples from the posterior $\pi(X, \theta|Y)$.

3.2 Priors

In the ongoing analysis we put the following priors on the model parameters: $\sum_{i=1}^k \alpha_i \sim \mathcal{N}(a_0, A_0)$, for λ_i and τ_i we use a gamma prior $\mathcal{G}(l_0, L_0)$ and $\mathcal{G}(d_0, D_0)$. We use $l_{0,i} = 1$, $L_{0,i} = 1/50$, $d_{0,i} = 0.1$ and $D_{0,i} = 1/5$. For the starting value of the instantaneous volatility process we set $\sigma^2_{0,i} \sim \mathcal{G}(s_0, S_0)$, where $s_0 = 0.001$ and $S_0 = s_0$. $\mathcal{N}$ denotes the normal law and $\mathcal{G}$ stands for the gamma distribution. The parameters μ and β will not estimated and set to zero.

For $\sum_{i=1}^k \alpha_i$ we use an informative prior. The importance of using an informative prior with the mean volatility parameters can be motivated by the fact that (i) $\sum \alpha_i$ equals to the expected sum of instantaneous volatilities and (ii) also the properties of a maximum likelihood estimation (with a fixed process (σ^2_n) and $k = 1$) improve if α is a-priori fixed at the sample mean. To incorporate this 'prior' knowledge into a Bayesian model, we use strong priors on $\sum_i \alpha_i$, more precisely we set $a_0 = y'y/T/\Delta$ and $A_0 = 0.001\ a_0$.

Argument (ii) should be strengthened by a small simulation experiment. Assume that $\alpha^{tr} = 0.012$, $\lambda^{tr} = 200$ and $\tau^{tr} = 2$ and generate 1000 paths each with 500 data points; $\Delta = 1/252$, $k = 1$. For these paths maximum likelihood estimation can be applied. The initial values for the maximization are taken at random from uniform distributions on the intervals: $\alpha^{start} \in [\alpha^{tr} - 0.005, \alpha^{tr} + 0.005]$, $\lambda^{start} \in [\lambda^{tr} - 100, \lambda^{tr} + 100]$ and $\tau^{start} \in [\tau^{tr} - 1, \tau^{tr} + 1]$. The mean and the standard deviations of the estimates are provided in Table 1. The reader should note that the standard deviations are high and that the mean of the estimate of α is approximately four times the true parameter value α^{tr}. However, we know from square root processes that the expectation is α^{tr}. Therefore, let us fix the estimates of α at the corresponding sample means. Then we estimate (for the same paths generated above) the remaining two parameters by maximizing the likelihood function given $\hat{\alpha}$ by the sample mean. Also these estimates are provided in Table 1. The reader could easily observe that this second approach improves the estimation of the parameters substantially. To take use of these properties observed with the ML estimation in a Bayesian setting we integrated this "prior "knowledge into the priors on $\sum_{i=1}^k \alpha_i$.

$\hat{\alpha}$ by ML			$\hat{\alpha}$ from sample mean		
$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\tau}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\tau}$
0.0576	203.0085	1.5150	0.0123	253.8385	2.1849
(0.004)	(59.696)	(0.285)	(0.001)	(37.746)	(0.119)

Table 1: Means and standard deviations (in parentheses) of ML estimates from square root processes.

4 Results

First we tested our sampler in simulated data. For the one factor setting ($k = 1$), we generated paths with $N = 500$ data points. The performance of the sampler for different parameters and ν is investigated in subsection 4.1. $\beta = 0$ and $\mu = 0$. Throughout this paper, we work with a step-width $\Delta = 1/252$. For the two factor setting ($k = 2$), we test the sampler with $\alpha = (0.006\ 0.007)'$, $\lambda = (50\ 150)'$, $\tau = (0.5\ 1)'$, $\beta = 0$ and $\mu = 0$. The parameters are updated by means of the Metropolis-Hastings (MH) algorithm. In the MH algorithm we use smoother densities

$$\pi(H_n, S_n^2 | H_{j,j\neq n}, S_{j,j\neq n}^2, Y, \theta) = \Pi_{i=1}^k \pi(h_{i,n}, \sigma_{i,n}^2 | h_{i,j,j\neq n}, \sigma_{i,j,j\neq n}^2, Y, \theta_i) \tag{25}$$

as described in Jong and Sheppard (1995) to update the parameters of the latent process $X = (X_1^j, X_2^j)$ with $j \in \{C, NC, PC\}$. In all update steps random-walk proposals are used. The most important point of this analysis is the *parameterization of the latent process* and *blocking*. For a centered parameterization we observe that λ and τ are positively correlated, their posterior modes are much too large. This can neither be improved by blocking λ_i and τ_i nor by blocking α_i, λ_i and τ_i. For the non-centered parameterizations the acceptance rates in the updates of θ are very small and the posterior distributions of λ, τ have modes much smaller than the true parameter values. This motivated us to construct the partially centered parameterization. Subsection 4.1 is concerned with the tuning of ν_i.

4.1 Performance in simulated data

We generated data from a one factor model, $k = 1$, with $N = 500$ observations.[3] For this experiment we generated paths from four different processes M_j:

[3]Since $k = 1$ we skip the component index i in this subsection.

Parameters			$\mathbb{E}(\sigma^2)$	VAR (σ^2)	M_j
α	λ	τ			
0.01	50	0.25	0.01	6.250e-6	M_1
0.01	100	0.25	0.01	3.125e-6	M_2
0.01	100	1.00	0.01	2.500e-5	M_3
0.01	200	1.00	0.01	5.000e-5	M_4

Since the variance in the observed yields is $3.4225e-5$ and the expected value of σ^2 is $3.4225e-5/\Delta$, with $\Delta = 0.004$, for the empirical data, we chose $\alpha = 0.01$. For τ, we choose two processes with a small 'volatility of volatility' term, $\tau = 0.25$, and two with a larger volatility in the latent process. Last but not least, as claimed in Frühwirth-Schnatter (2002) for discrete time models, the speed of decay has an influence on the speed of convergence. The objective of this simulation study is to evaluate the impact of ν on the sampling properties for different values of λ and τ. In contrast to Papaspiliopoulos *et al.* (2002) and Frühwirth-Schnatter (2002) where Gibbs sampling can be applied, our simulations have to take care on the acceptance rate of the Metropolis-Hastings algorithm.

The results of our simulations are the following: Before going into the details we have to remark that the models with the higher 'volatility of volatility' tend to behave less sensitive in the choice of ν. Especially, M_3 and M_4 perform well for a relatively large range of parameters $\nu \leq 0.15$. The best performance with M_3 and M_4 was observed with $\nu = 0.05$. However, with small ν, the sampling properties deteriorate with models M_1 and M_2, respectively. Acceptable, sampling properties are observed with $0.05 \leq \nu \leq 0.1$, i.e. the choice of ν is much more sensitive in M_1 and M_2 compared to M_3 or M_4. For M_1 we observe the best performance with $\nu = 0.1$, while with M_2, $\nu = 0.05$ and $\nu = 0.1$ result in approximately the same sampling qualities. Unfortunately, due to limited computing power these simulation results are based only a view runs. A more rigorous treatment of this problem has to be postponed to the future.

Let us compare our results to the seminal works of Papaspiliopoulos *et al.* (2002) and Frühwirth-Schnatter (2002): First we can verify the higher the volatility of the latent process, measured by VAR (σ^2), the better the performance of a more centered parameterization; i.e. ν is small. Despite the fact that only a limited number of simulation runs can be performed, this effected is clearly observed with the for different processes. Furthermore, we investigate the claim that given τ, the higher λ the more centering is required. For this claim the results are mixed. For M_3 and M_4 we cannot find out significant difference, while for the $\tau = 0.25$ examples, $nu = 0.05$ outperforms $\nu = 0.1$ for M_2.

$k=1$			$k=2$					
α	λ	τ	α_1	λ_1	τ_1	α_2	λ_2	τ_2
0.0110	99.7766	0.0604	0.0062	64.9817	0.0.0719	0.0037	225.9881	0.0875
(8e-4)	(28.851)	(0.011)	(0.008)	(38.826)	(0.022)	(0.004)	(27.816)	(0.060)

Table 2: DM/US$ exchange rates form January 1998 to December 1999, $N = 504$. Posterior means, standard deviations and parentheses. 10000 simulation steps, 1000 burn-in steps.

4.2 Parameter estimation with DM/US$ exchange rate data

After we have tuned our sampler with simulated data, we applied this algorithm to daily DM/US$ exchange rate data from January 1998 to December 1999, where $N = 505$. Note that the squared variation $y'y/T$ was equal to $3.4224e-5$, such that a posterior mean for $\sum_i \alpha_i$ of $y'y/T/\Delta = 0.0086$ can be expected. For the empirical data a ν of 0.1 was applied for the one factor model, while $\nu = (0.05\ 0.1)'$ was used in the two factor setting such that the larger $\nu_2 = 0.1$ is applied to the component with the lower variance. To ensure this ordering we apply a permutation sampler with a restriction on λ_i. [4] Table 2 presents the mean and the standard deviation of the posterior for the corresponding model parameters for the one factor and the two factor setting. Comparing our results to the estimates of Bollerslev and Zhou (2002) we observe that our one factor estimates are very close to the estimates presented in Bollerslev and Zhou (2002) while for the two factor model, the both persistence parameter λ_i are larger than in Bollerslev and Zhou (2002). Nevertheless, we have to remark that, in contrast to the method of moments approach, MCMC output automatically provides us with extra information. The posterior estimates of the instantaneous volatilities and the increases in integrated volatility are a byproduct from data augmentation.

5 Conclusions

This article has implemented an exact Bayesian analysis of the Heston (1993) stochastic volatility model. We observe that different parameterizations of the latent volatility process and the parameters of the volatility process result in very different convergence behavior of the MCMC sampler; the best performance has been observed with a partially centered version of the latent process. Although the model estimated in this paper, is non-linear and in continuous time some properties observed with linear discrete time models continue to

[4] A restriction on the variance $\mathrm{VAR}(\sigma_i^2) = (\alpha_i \tau_i^2)/(2\lambda_i)$ exhibits similar performance. Furthermore, we stick to the simpler restriction on λ_i.

hold. First, the higher the variance of the latent volatility process – which implies a high volatility of returns – the closer we get to the centered parameterization. Secondly, a stronger mean reversion shifts the parameterization in the direction of the centered version. We apply the sampler to simulated data and to DM/US$ exchange rate data. Comparisons to GMM estimates in Bollerslev and Zhou (2002) shows that MCMC estimates deliver similar results.

References

Y. Ait-Sahalia (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70:223–264.

T. Andersen, J. Lund (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77:343–377.

O.E. Barndorff-Nielsen, N. Shephard (2001). Non-gaussian OU based models and some of their uses in financial economics. *Journal of the Royal Statistical Society B*, 63:167–241.

O.E. Barndorff-Nielsen, N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B*, 64:253–280.

F. Black, M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 8:1637–1659.

T. Bollerslev, H. Zhou (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109:33–65.

O. Elerian, S. S. Chib, N. Shephard (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, 69:959–993.

B. Eraker (2001). Markov chain Monte Carlo analysis of diffusion models with applications to finance. *Journal of Business, Economics and Statistics*, 19:177–191.

S. Frühwirth-Schnatter (2002). Efficient bayesian estimation for state space models based on reparameterizations. *mimeo, Vienna University of Economics and Business Administration.*

A. Gallant, G. Tauchen (1996). Which moments to match? *Econometric Theory*, 12:657–681.

L. Hansen, J. Scheinkman (1995). Back to the future: Generating moment implications for continuous time markov processes. *Econometrica*, 63:767–804.

S.L. Heston (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343.

F. Hubalek, R. Tomkins (2001). Does it matter where jumps come from? Underlying versus volatility processes. *working paper, University of Technology, Vienna.*

J. Hull, A. White (1987). The pricing of options on assets with stochastic volatility. *Journal of Finance*, 42:281–300.

P. Jong, N. Sheppard (1995). The simulation smoother for time series models. *Biometrika*, 82:339–350.

D. Lamberton, B. Lapeyre (1996). *Introduction to Stochastic Calculus – Applied to Finance.* Chapman & Hall, London, 1^{st} edition.

N. Meddahi, E. Renault (2000). Temporal aggregation of volatility models. *Working paper CIRANO, Montreal.*

O. Papaspiliopoulos, P. Dellaportas, G. Roberts (2001). Bayesian inference for non-gaussian ornstein-uhlenbeck stochastic volatility processes. *mimio, Lancaster.*

O. Papaspiliopoulos, G. Roberts, M. Sköld (2002). Non-centered parameterisations for hierarchical models and data augmentation. *mimeo, Lancaster.*

G. Roberts, O. Stramer (2001). On inference for partially observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88:603–621.

N. Shephard (1996). Statistical aspects of ARCH and stochastic volatility. *In Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, D.V. (Eds.): Time Series Models in econometrics, finance and other fields. London: Chapman & Hall.*

R. Stanton (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, 52:1973–2002.

A Euler approximation

Consider a square root process $(X(t))$:

$$dX(t) \quad = \quad \lambda(\alpha - X(t))dt + \tau\sqrt{X}(t)dW(t) \ , \tag{26}$$

and define a process $Y(t) = X(t)\exp(-\lambda t)$. Applying the Ito lemma results in:

$$\begin{aligned} Y(t) &= X(0)\exp(\lambda 0) + \int_0^t X\lambda\exp(\lambda s)ds + \int_0^t \exp(\lambda s)dX(s) \\ &= X(0) + \int_0^t X\lambda\exp(\lambda s)ds \\ &+ \int_0^t \exp(\lambda s)\lambda(\alpha - X)ds + \int_0^t exp(\lambda s)\tau\sqrt{X}dW(s) \\ &= X(0) + (\exp(\lambda t) - 1)\alpha + \int_0^t exp(\lambda s)\tau\sqrt{X}(s)dW(s) \end{aligned} \tag{27}$$

where by the definition of $Y(t)$, the solution for $X(t)$ is given by:

$$\begin{aligned} X(t) &= X(0)\exp(-\lambda t) + (1 - \exp(-\lambda t))\alpha \\ &+ \exp(-\lambda t)\int_0^t exp(\lambda s)\tau\sqrt{X}(s)dW(s) \qquad (28) \\ &\approx X(0)\exp(-\lambda t) + (1 - \exp(-\lambda t))\alpha + \tau\sqrt{X}\sqrt{\Delta}\omega(1)\ . \qquad (29) \end{aligned}$$

Equation (29) follows from (28) by a first order Taylor approximation. Integration of $(X(t))$ on an interval $[0, \Delta]$, and approximating the term with the Brownian motion, results in

$$\int_0^\Delta X(t)dt \approx = \alpha\Delta + (X(0) - \alpha)\frac{1 - \exp(-\lambda\Delta)}{\lambda} + \tau\sqrt{X}(\Delta)^{1.5}\tilde{\omega}(1)\ . \tag{30}$$

B Conditional densities

B.1 Centered Parameterization

This section presents the Fourier transform of $\pi(h_{i,n,C}|\sigma^2_{i,n-1,C};\theta_i)$ and the conditional distribution of the square root process is a non-central χ^2 law $\pi(\sigma^2_{i,n,C}|\sigma^2_{i,n-1,C};\theta_i)$, where $\theta_i = (\alpha_i, \lambda_i, \tau_i)$. $\pi(h_{i,n,C}|\sigma^2_{i,n-1};\theta_i)$ can be derived by means of Fourier inversion. The characteristic function $\phi(u, t, h_{i,n-1,C};\lambda_i, \alpha_i, \tau_i)$ is given by (see Lamberton and Lapeyre (1996)):

$$\begin{aligned} \phi(u, \Delta, h_{i,n,C}, \sigma^2_{i,n-1,C};\theta_i) &= E[\exp(\imath u h_{i,n,C})] \\ &= \exp(B_{C1}(\Delta, u, \theta_i))\exp(B_{C2}(\Delta, u, \theta_i)\sigma^2_{i,n-1,C}) \\ B_{C1}(\Delta, u, \theta_i) &= \frac{2\alpha_i\lambda_i}{\tau_i^2}\log\left(\frac{2\gamma_i\exp(0.5\Delta(\gamma_i + \lambda_i)}{\gamma - \lambda_i + \exp(\gamma_i\Delta)(\gamma_i + \lambda_i)}\right) \\ B_{C2}(\Delta, u, \theta_i) &= \frac{2\imath u(\exp(\gamma_i\Delta) - 1)}{\gamma_i - \lambda_i + \exp(\gamma_i\Delta)(\gamma_i + \lambda_i)} \\ \gamma_i &= \sqrt{\lambda_i^2 - 2\lambda_i^2\imath u} \end{aligned} \tag{31}$$

where $\imath = \sqrt{-1}$ and $u \in (-\infty, \infty)$ is the convolution parameter.

The conditional distribution of the square root process is a non-central χ^2 distribution. The density is given by:

$$\begin{aligned}
\pi(\sigma^2_{i,n,C}|\sigma^2_{i,n-1};\theta_i) &= c_i \exp\left(-c_i(\sigma_{i,n,C} + \exp(-\lambda_i\Delta)\sigma^2_{i,n-1,C})\right) \\
&\quad \cdot \left(\frac{\sigma^2_{i,n,C}}{\exp(-\lambda_i\Delta)\sigma^2_{i,n-1,C}}\right)^{0.5q_i} \cdot I_{q_i}[2c_i\sqrt{\sigma}_{i,n,C}\exp(-\lambda_i\Delta)\sigma^2_{i,n-1}] \ , \\
&\quad c_i := \frac{2\lambda_i}{\tau_i(1-\exp(-\lambda_i\Delta))} \\
&\quad q_i := \frac{2\lambda\alpha_i}{\tau_i^2} - 1 \ ,
\end{aligned}$$

where $I_{q_i}[.]$ is the modified Bessel function of the first kind of order q_i. The corresponding characteristic function for $\sigma^2_{i,n,C}$ is given by (see Lamberton and Lapeyre (1996)):

$$\begin{aligned}
\phi(u, \Delta, \sigma^2_{i,n,C}; \sigma^2_{i,n-1,C}, \theta_i) &= E[\exp(\imath u\sigma^2_{i,n,C})] \\
&= \exp(B_{C3}(\Delta, u, \theta_i))\exp(B_{C3}(\Delta, u, \theta_i)\sigma^2_{i,n-1,C}) \\
B_{C3}(\Delta, u, \theta_i) &= \frac{2\alpha_i\lambda_i}{\tau_i^2}\log\left(\frac{2\lambda_i\exp(\Delta\lambda_i)}{-\imath u\tau_i^2(\exp(-\lambda_i\Delta)-1)+2\lambda_i\exp(\lambda_i\Delta)}\right) \\
B_{C4}(\Delta, u, \theta_i) &= \frac{2\imath u\lambda_i}{-\imath u\tau_i^2(\exp(-\lambda_i\Delta)-1)+2\lambda_i\exp(\lambda_i\Delta)} \ . \qquad (32)
\end{aligned}$$

B.2 Partially-centered Parameterization

Section 2 defines $h_{i,n,PC} = (1-\nu_i)h_{i,n,C} + \nu_i h_{i,n,NC}$ and $\sigma^2_{i,n,PC} = (1-\nu_i)\sigma^2_{i,n,C} + \nu_i\sigma^2_{i,n,NC}$ as sums of independent random variables. From subsection B.1 we already know the Fourier transforms of $h_{i,n,C}$ and $\sigma_{i,n,C}$ (see Equations (31) and (32)).

For the non-centered parameterization ($\nu_i = 1$), $h_{i,n,NC} \sim \mathcal{N}(a_{i,n,H}, A^2_{i,n,H})$ and $\sigma^2_{i,n,NC} \sim \mathcal{N}(a_{i,n,S}, A^2_{i,n,S})$, where

$$\begin{aligned}
a_{i,n,H} &:= \alpha_i\Delta + (\sigma^2_{i,n-1,NC} - \alpha_i)\frac{1-\exp(-\lambda_i\Delta)}{\lambda_i} \ , \\
A_{i,n,H} &:= \tau_i\sigma_{i,n-1,NC}(\Delta)^{1.5} \ , \\
a_{i,n,S} &:= \sigma^2_{i,n-1,NC}\exp(-\lambda_i\Delta) + (1-\exp(-\lambda_i\Delta))\alpha_i \ , \\
A_{i,n,S} &:= \tau_i\sigma_{i,n-1,NC}(\Delta)^{0.5} \ .
\end{aligned}$$

To derive the Fourier transforms of $h_{i,n,PC}$ and $\sigma_{i,n,PC}$ we simply have to use the facts that (*F1*) the Fourier transform of a sum of independent random variables is the product of the

Fourier transforms and (*F2*) given the characteristic function $\phi(t)$ of a random variable X, the Fourier transform of an affine transformation $Y = aX + b$ is given by $\exp(\imath ub)\phi(au)$. Since $h_{i,n,NC} = a_{i,n,H} + A_{i,n,H}\tilde{\omega}_{i,n}(1)$, $\sigma^2_{i,n,NC} = a_{i,n,S} + A_{i,n,S}\omega_{i,n}(1)$, and the characteristic function of a standard normal variable is $\exp(-u^2/2)$, we derive by means of *F2*:

$$\phi(u, \Delta, h_{i,n,NC}; \sigma^2_{i,n-1,NC}, \theta_i) = exp(\imath u a_{i,n,H}) \exp(-0.5(A_{i,n,H}u)^2) \tag{33}$$

$$\phi(u, \Delta, \sigma^2_{i,n,NC}, \sigma^2_{i,n-1,NC}; \theta_i) = exp(\imath u a_{i,n,S}) \exp(-0.5(A_{i,n,S}u)^2) \tag{34}$$

By *F1* and *F2*, the Fourier transforms of $h_{i,n,PC}$ and $\sigma^2_{i,n,PC}$ are given by:

$$\begin{aligned}\phi(u, \Delta, h_{i,n,PC}; \sigma^2_{i,n-1,NC}, \theta_i) &= \phi(u(1-\nu_i), \Delta, h_{i,n,C}, \sigma^2_{i,n-1,C}; \theta_i)\phi(u\nu_i, \Delta, h_{i,n,NC}; \sigma_{i,n-1,NC}, \theta_i) \\ &= \exp(B_{C1}(\Delta, u(1-\nu_i))) \exp(B_{C2}(\Delta, u(1-\nu_i))\sigma^2_{i,n-1,C}) exp(\imath u\nu a_{i,n,H}) \exp(-0.5(A_{i,n,H}\nu u)^2) \quad (35) \\ \phi(u, \Delta, \sigma^2_{i,n,NC}; \sigma^2_{i,n-1,NC}, \theta_i) &= \phi(u(1-\nu_i), \Delta, \sigma^2_{i,n,C}, \sigma^2_{i,n-1,C}; \theta_i)\phi(u\nu_i, \Delta, \sigma_{i,n,NC}; \sigma_{i,n-1,NC}, \theta_i) \\ &= \exp(B_{C3}(\Delta, u(1-\nu_i))) \exp(B_{C4}(\Delta, u(1-\nu_i))\sigma^2_{i,n-1,C}) exp(\imath u\nu a_{i,n,S}) \exp(-0.5(A_{i,n,S}\nu u)^2) \quad (36)\end{aligned}$$