Simulated Maximum Likelihood Estimation of Continuous Time Stochastic Volatility Models*

Tore Selland Kleppe
University of Bergen

Jun Yu
Singapore Management University

Hans J. Skaug
University of Bergen

Abstract

In this paper we develop and implement a method for maximum simulated likelihood estimation of the continuous time stochastic volatility model with the constant elasticity of volatility. The approach does not require observations on option prices nor volatility. To integrate out latent volatility from the joint density of return and volatility, a modified efficient importance sampling technique is used after the continuous time model is approximated using the Euler-Maruyama scheme. The Monte Carlo studies show that the method works well and the empirical applications illustrate usefulness of the method. Empirical results provide strong evidence against the Heston model.

JEL classification: C11, C15, G12

Keywords: Efficient importance sampler; Constant elasticity of volatility

1 Introduction

Continuous time stochastic volatility (SV) models have been proven to be very useful from various aspects. For example, it has been found that SV models provide a valuable tool for pricing contingency-claims. Seminal contributions include Wiggins (1987), Hull and White (1987), Heston (1993), and Duffie et al. (2000). See Bakshi et al. (1997) for an empirical analysis of the SV models. For another example, SV models have proven successful to describe the time series behavior of financial variables. Important contributions include Andersen and Lund (1997) and Jones (2003). Unfortunately, maximum likelihood estimation (MLE) of continuous time stochastic volatility models poses substantial challenges. The first challenge lies in the fact that the joint transition density of price (or return) and volatility is typically unknown in closed form. This is the well known problem in the continuous time literature (see Aït-Sahalia (2002) and Phillips and

*Kleppe gratefully acknowledges the hospitality during his research visit to Sim Kee Boon Institute for Financial Economics at Singapore Management University. Yu gratefully acknowledges support from the Singapore Ministry of Education AcRF Tier 2 fund under Grant No. T206B4301-RS. We wish to thank two anonymous referees for their helpful comments

Yu (2009a)). The second challenge is that when only the time-series of spot prices is observed, volatility has to be integrated out from the joint transition density. Such integrals are analytically unknown and have to be calculated numerically. The dimension of integration is the same as the number of observations. When the number of observations is large, which is typical in practical applications, the numerical integration is difficult.

In recent years, solutions have been provided to navigate such challenges. To deal with the second challenge, for example, Jones (2003) and Aït-Sahalia and Kimmel (2007) proposed to estimate the model using data from both underlying spot and options markets. Option price data are used to extract volatility, making the integration of volatility out of the joint transition density unnecessary. To deal with the first challenge, Jones (2003) suggested using in-filled Euler-Maruyama approximations which enables a Gaussian approximation to the joint transition density while Aït-Sahalia and Kimmel (2007) advocated using a closed form polynomial approximation that can approximate the true the joint transition density arbitrarily well. With the two problems circumvented, the full likelihood based inference is possible. For example, the method of Aït-Sahalia and Kimmel (2007) is frequentistic while the method of Jones (2003) is Bayesian.

It is well known that option prices are derived from the risk-neutral measure. Consequently, a benefit of using data from both spot and options markets jointly is that one can learn about the physical as well as risk-neutral measures. However, this benefit comes at expense. To connect the physical and risk-neutral measures, the functional form of the market price of risk has to be specified.

In this paper, we develop and implement a method for maximum simulated likelihood estimation of the continuous time SV model with the constant elasticity of volatility (CEV-SV). The approach does not require observations of option prices or volatility and hence there is no need to specify the functional form of the market price of risk. As a result, we only learn about the physical measure. The CEV-SV model was first proposed by Jones (2003), as a simple way to nest some standard continuous time SV models, such as the square-root SV model of Heston (1993) and the GARCH diffusion model of Nelson (1990). To our knowledge, the present paper constitutes the first time ML is used to estimate the CEV-SV model using the spot price only.

To deal with the second challenge, we propose to use a modified Efficient Importance Sampler (EIS) algorithm, originally developed in Richard and Zhang (2007), to integrate out a latent volatility process. To deal with the first challenge, we consider the Euler-Maruyama approximation. We examine the performance of the proposed maximum simulated likelihood using both simulated data and real data. Based on simulated results, we find that the algorithm performs well. Empirical illustration suggests that the Heston square-root SV model does not fit data well. This empirical finding reinforces those of Andersen et al. (2002), Jones (2003), Aït-Sahalia and Kimmel (2007) and others.

The paper is organized as follows. Section 2 discusses the model and introduces the estimation method. Section 3 tests the accuracy of the method by performing Monte Carlo simulations for the square root SV model of Heston (1993), the GARCH diffusion model of Nelson (1990) and the CEV-SV model. In Section 4, we apply this estimation method to real data for the three stochastic volatility models, and analyze and compare the empirical results. A comparison with the log-normal model is also performed. Section 5 concludes.

2 Model and Methodology

This section first presents the constant elasticity of volatility (CEV)-SV model under consideration, and then outlines the Monte Carlo (MC) procedure used to do likelihood analysis when only the price process is observed.

2.1 SV model with constant elasticity of volatility

The continuous time CEV model was recently proposed to model stochastic volatility (see e.g. Jones (2003); Aït-Sahalia and Kimmel (2007)). While we mainly focus on the CEV-SV model in this paper, the proposed approach is applicable more generally.

Let $\bar{s}_t$ and $\bar{v}_t$ denote the log-price of some asset and the volatility respectively at some time t. Then the CEV model is specified in terms of the Itô stochastic differential equation:

$$d\begin{bmatrix} \bar{s}_t \\ \bar{v}_t \end{bmatrix} = \begin{bmatrix} a + b\bar{v}_t \\ \alpha + \beta\bar{v}_t \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\bar{v}_t} & \rho\sqrt{\bar{v}_t} \\ 0 & \sigma\bar{v}_t^{\gamma} \end{bmatrix} \begin{bmatrix} dB_{t,1} \\ dB_{t,2} \end{bmatrix}. \tag{1}$$

Here $B_{t,1}$ and $B_{t,2}$ denotes a pair of independent canonical Brownian motions. The parameters $\theta = [\alpha, \beta, \sigma, \rho, \gamma, a, b]$ have the restrictions $\alpha > 0$, $\rho \in (-1, 1)$, $\gamma \geq 1/2$ and $\beta < 0$ whenever $\gamma \leq 1$ (see Jones (2003) for a treatment of the volatility process for $\gamma > 1$). In addition, for $\gamma = 1/2$ we have the restriction $2\alpha > \sigma^2$ to ensure that $\bar{v}_t$ stays strictly positive (Cox et al. (1985)). The CEV model nests the affine SV model of Heston (1993) ($\gamma = 1/2$) and the GARCH diffusion model of Nelson (1990) ($\gamma = 1$), and we will treat these special cases separately in addition to the full CEV model. Parameters α and β characterize the linear drift structure of the volatility, in which $-\alpha/\beta$ is the mean and $-\beta$ captures the mean reversion rate. Parameter σ is the volatility-of-volatility and ρ represents the leverage effect. Parameter γ is the CEV elasticity. Parameters a and b represent, respectively, the long run drift and risk premium of the price process.

2.2 A change of variable and time discreteization

Under the parameter constraints described above, the volatility process $\bar{v}_t$ is strictly positive with probability one. The importance sampling procedure proposed here uses (locally) Gaussian importance densities for $\bar{v}_t$, and thus the supports of $\bar{v}_t$ and the importance density are inherently conflicting. To remove this boundary restriction, we shall work with the logarithm of the volatility process. As the latent process gets integrated out, the actual representation of the volatility (or the log-volatility) is irrelevant theoretically, but is very important for the construction of efficient importance sampling procedures as will be clear below. In addition, this change of variable will influence the properties of the time-discreteization scheme that will be discussed shortly.

Define $\bar{z}_t = \log(\bar{v}_t)$. Then, by Ito's lemma, we have:

$$d\begin{bmatrix} \bar{s}_t \\ \bar{z}_t \end{bmatrix} = \begin{bmatrix} a + b\exp(\bar{z}_t) \\ M(\bar{z}_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)}\exp(\bar{z}_t/2) & \rho\exp(\bar{z}_t/2) \\ 0 & \sigma\exp(\bar{z}_t(\gamma-1)) \end{bmatrix} \begin{bmatrix} dB_{t,1} \\ dB_{t,2} \end{bmatrix} \tag{2}$$

where $M(\bar{z}_t) = \beta + \alpha\exp(-\bar{z}_t) - \sigma^2\exp(2\bar{z}_t(\gamma-1))/2$. Clearly, the law of $\bar{s}_t$ is unaltered, but the latent process $\bar{z}_t$ now has support over the real line. As the transition probability density (TPD) in the general case is not known under either representation (1) or (2), an approximation is needed. This is achieved by defining a discrete time model that acts as an approximation to (2) based on

the Euler-Maruyama (EM) scheme using a time-step equal to Δ time units. This discrete time process is given as the non-linear and heteroskedastic auto-regression

$$\begin{bmatrix} s_{i+1} \\ z_{i+1} \end{bmatrix} = \begin{bmatrix} s_i + \Delta(a + b\exp(z_i)) \\ z_i + \Delta M(z_i) \end{bmatrix} + \sqrt{\Delta} \begin{bmatrix} \sqrt{(1-\rho^2)}\exp(z_i/2) & \rho\exp(z_i/2) \\ 0 & \sigma\exp(z_i(\gamma-1)) \end{bmatrix} \begin{bmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \end{bmatrix},$$

where $[\varepsilon_{i,1},\ \varepsilon_{i,2}]$ are temporarily independent bi-variate standard normal shocks. It is convenient to work with the log-returns of the price, so we define $x_i = s_i - s_{i-1}$ as this process is stationary. Hence the discrete time dynamics for x_i are given by:

$$\begin{bmatrix} x_{i+1} \\ z_{i+1} \end{bmatrix} = \begin{bmatrix} \Delta(a + b\exp(z_i)) \\ z_i + \Delta M(z_i) \end{bmatrix} + \sqrt{\Delta} \begin{bmatrix} \sqrt{(1-\rho^2)}\exp(z_i/2) & \rho\exp(z_i/2) \\ 0 & \sigma\exp(z_i(\gamma-1)) \end{bmatrix} \begin{bmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \end{bmatrix}. \quad (3)$$

Throughout the rest of this paper, (3) will be the model that we shall work with.

Several authors (see e.g. Aït-Sahalia (2002), Durham and Gallant (2002), Durham (2006)) have argued that one should transform the latent process (instead of the log-transform applied here) in such a manner that it becomes a (non-linear) Ornstein-Uhlenbeck process – i.e. with a homoskedastic error term in the latent process. This variance stabilisation transform is given (see e.g. Rao (1999) p. 210), up to an affine transformation, by:

$$Z(v) = \begin{cases} \log(v) & \text{if } \gamma = 1 \\ \frac{v^{1-\gamma}-1}{1-\gamma} & \text{otherwise.} \end{cases}$$

However, excluding $\gamma = 1$, the variance stabilising does not completely solve the finite boundary problem on the domain of the transformed volatility $Z(\bar{v}_t)$. Thus, the Gaussian approximation obtained from the EM scheme will have a support that conflicts with the continuous time model. In Section 3, some MC experiments are conducted to verify that an approximate likelihood function based on the EM-scheme (3) with observed log-volatility does not lead to unacceptable biases for sample sizes and time steps that are relevant in practice.

Another reason for using the variance stabilisation procedure would be to bring the posterior density (i.e. the density of the latent given the observed price returns and parameters) closer to a multivariate Gaussian, as in Durham (2006) in the context of an EM discreteization of the Heston model. This should in theory pave the way for using a Laplace approximation-based importance density (see e.g. Shephard and Pitt (1997)), i.e. a multivariate Gaussian, to calculate the marginal likelihood of the data. By using the EIS procedure outlined below, there is no need to bring the posterior density closer to normality globally, as our importance density is only locally Gaussian. Thus, the argument against using the logarithm for all γ does not apply here. We therefore conclude this discussion and use the logarithm throughout the rest of the paper.

2.3 TPDs and joint densities

Assume that we have n observations of x_i, i.e $\mathbf{x} = [x_1, \ldots, x_n]$[1], sampled discretely over a regular time grid with Δ time units between the time-points. More general deterministic time grids are possible. Further, denote the unobserved vector of z_is at the corresponding times as $\mathbf{z} = [z_1, \ldots, z_n]$. For simplicity, we assume for now that z_0 is a known constant. The marginal distribution of z_0 is

[1] This is a slight abuse of notation, as the data are from the continuous time process (1) and not the discrete time approximation.

not known in closed form in the general case, and in practice we will estimate z_0 along with the parameter vector θ by maximum likelihood.

Let $f_i = f_i(z_i, x_i|z_{i-1}, \theta, \Delta)$ denote the Gaussian TPD of the discrete time process (3). From the specification, it is evident that f_i is a bi-variate Gaussian density with mean vector and covariance matrix

$$\begin{bmatrix} \Delta(a + b\exp(z_{i-1})) \\ z_{i-1} + \Delta M(z_{i-1}) \end{bmatrix} \text{ and } \Delta \begin{bmatrix} \exp(z_{i-1}) & \sigma\rho\exp\left(\frac{z_{i-1}}{2}(2\gamma - 1)\right) \\ \sigma\rho\exp\left(\frac{z_{i-1}}{2}(2\gamma - 1)\right) & \sigma^2\exp(2z_{i-1}(\gamma - 1)) \end{bmatrix}$$

respectively. Exploiting the Markov structure of the discretized model, the joint density of $(\mathbf{z}, \mathbf{x})$ is given by:

$$p(\mathbf{z}, \mathbf{x}|\theta, z_0, \Delta) = \prod_{i=1}^{n} f_i(z_i, x_i|z_{i-1}, \theta, \Delta). \tag{4}$$

Clearly, this expression should also be regarded as an approximation to the continuous time joint density obtained when the f_is are exchanged with the (unknown) exact transition densities. The approximation is known to converge strongly as $\Delta \to 0$ (Kloeden and Platen, 1999).

2.4 Monte Carlo evaluation of the marginal likelihood

Since the log-volatility $\mathbf{z}$ is unobserved, approximate evaluation (based on the EM-discreteization) of the likelihood function for given values of θ and z_0 involves an integral over $\mathbf{z}$, say

$$l(\theta, z_0|\mathbf{x}) = \int p(\mathbf{z}, \mathbf{x}|\theta, z_0)d\mathbf{z}.$$

Due to the non-linear structure of the discrete time model (3), no closed form expression for this integral is at hand, and hence numerical methods generally are needed. Since the dimension of the integral is typically of the order of 1000-10000, quadrature rules are of little use here. Instead, we apply an importance sampling technique where the importance density is constructed using the Efficient Importance Sampling (EIS) algorithm of Richard and Zhang (2007).

The EIS algorithm (approximately) minimises the MC variance within a parametric class of auxiliary importance densities, say $m(\mathbf{z}|\mathbf{a}, \mathbf{x}, z_0)$, being indexed by the $n \times 2$-dimensional parameter $\mathbf{a}$. We denote the optimal choice of $\mathbf{a}$ by $\hat{\mathbf{a}}$.[2] Further, we refer to $m(\mathbf{z}|\mathbf{0}, \mathbf{x}, z_0)$ as the base-line importance density where $\mathbf{a} = \mathbf{0}$ denotes $\mathbf{a}$ with all elements equal to zero.

In this work, we restrict the importance densities to have the form:

$$m(\mathbf{z}|\mathbf{a}, \mathbf{x}, z_0) = \prod_{i=1}^{n} m_i(z_i|z_{i-1}, x_i, \mathbf{a}_i).$$

Notice that we allow the importance density to depend explicitly on the observed vector $\mathbf{x}$. The weak law of large numbers for $S \to \infty$ suggests that $l(\theta, z_0|\mathbf{x})$ may be approximated by

$$\tilde{l}(\theta, z_0|\mathbf{x}, \mathbf{a}) = \frac{1}{S}\sum_{j=1}^{S} \frac{p(\tilde{\mathbf{z}}^{(j)}, \mathbf{x}|\theta, z_0)}{m(\tilde{\mathbf{z}}^{(j)}|\mathbf{a}, \mathbf{x}, z_0)}, \tag{5}$$

[2] In general, both m and $\hat{\mathbf{a}}$ depend on θ and Δ, but we suppress this dependence in our notation.

where $\tilde{\mathbf{z}}^{(j)},\ j = 1, \ldots, S$ are drawn from $m(\mathbf{z}|\mathbf{a}, \mathbf{x}, z_0)$. This law of large numbers applies in particular to $\hat{\mathbf{a}}$ that is obtained using the EIS algorithm so the variance of $\tilde{l}$ is approximately minimized. Thus, the approximate MLE estimator will have the form

$$(\hat{\theta}, \hat{z}_0) = \arg\max_{(\theta, z_0)} \log \tilde{l}(\theta, z_0|\mathbf{x}, \hat{\mathbf{a}}), \tag{6}$$

where the logarithm is taken for numerical convenience.

2.4.1 The Base-line importance density

Typically $m(\mathbf{z}|\mathbf{a}, \mathbf{x}, z_0)$ is taken to be a parametric extension to the so-called natural sampler (i.e. $p(\mathbf{z}|\theta, z_0)$) (see, for example, Liesenfeld and Richard (2003), Liesenfeld and Richard (2006), Richard and Zhang (2007) and Bauwens and Galli (2009)). In this paper, we depart from this practice by introducing information from the data into the base-line importance density. More precisely, we define

$$f_i(z_i|z_{i-1}, x_i, \theta, \Delta) = \frac{f_i(z_i, x_i|z_{i-1}, \theta, \Delta)}{\int f_i(z_i, x_i|z_{i-1}, \theta, \Delta) dz_i},$$

i.e. the conditional transition densities given x_i, and set $m_i(z_i|z_{i-1}, x_i, \mathbf{0}_i) = f_i(z_i|z_{i-1}, x_i, \theta, \Delta)$.[3] Since $f_i(z_i, x_i|z_{i-1}, \theta, \Delta)$ is a bi-variate Gaussian density, $f_i(z_i|x_i, z_{i-1}, \theta, \Delta)$ is also Gaussian, with mean and standard deviation given by:

$$\begin{aligned}
\mu_{\mathbf{0}_i}(z_{i-1}, x_i) &= z_{i-1} + \Delta M(z_{i-1}) + \sigma\rho(x_i - \Delta(a + b\exp(z_{i-1})))\exp\left(z_{i-1}\left(\gamma - \frac{3}{2}\right)\right), \\
\Sigma_{\mathbf{0}_i}(z_{i-1}) &= \sigma\sqrt{\Delta(1-\rho^2)}\exp(z_{i-1}(\gamma - 1)),
\end{aligned}$$

respectively.

2.4.2 The parametrically extended importance density

The base-line importance density is in it self a valid, but not very efficient importance density. Consequently, we parametrically extend it to introduce more flexibility. Following Richard and Zhang (2007), each factor of the base-line importance density is (conditionally on z_{i-1}) perturbed within the univariate Gaussian family of distributions. Staying within the Gaussian family is numerically advantageous because sampling from m then becomes fast and conceptually simple. More precisely, the extension is done by multiplying $m_i(z_i|z_{i-1}, x_i, \mathbf{0}_i)$ by $\exp(\mathbf{a}_{i,1}z_i + \mathbf{a}_{i,2}z_i^2)$ and compensating with the appropriate normalization factor. More precisely, we write m_i as

$$m_i(z_i|z_{i-1}, x_i, \mathbf{a}_i) = \frac{B_i(z_i|z_{i-1}, x_i)\psi_i(z_i, \mathbf{a}_i)}{\chi_i(z_{i-1}, x_i, \mathbf{a}_i)},$$

where

$$\begin{aligned}
\log B_i(z_i|z_{i-1}, x_i) &= -\frac{(z_i - \mu_{\mathbf{0}_i}(z_{i-1}, x_i))^2}{2\Sigma_{\mathbf{0}_i}(z_{i-1})^2}, \\
\log \psi_i(z_i, \mathbf{a}_i) &= \mathbf{a}_{i,1}z_i + \mathbf{a}_{i,2}z_i^2, \\
\chi_i(z_{i-1}, x_i, \mathbf{a}_i) &= \int B_i(z_i|z_{i-1}, x_i)\psi_i(z_i, \mathbf{a}_i)dz_i.
\end{aligned}$$

[3] $\mathbf{0}_i$ should be read as the ith row of $\mathbf{a}$ with the elements all equal to 0

An explicit expression for χ_i is given in appendix A. The mean and the standard deviation of $m_i(z_i|z_{i-1}, x_i, \mathbf{a}_i)$ which are used when sampling from $m(\mathbf{z}|\mathbf{a}, \mathbf{x}, z_0)$ have the form:

$$\mu_{\mathbf{a}_i}(z_{i-1}, x_i) = \frac{\mu_{\mathbf{0}_i}(z_{i-1}, x_i) + \mathbf{a}_{i,1}\Sigma_{\mathbf{0}_i}(z_{i-1})^2}{1 - 2\mathbf{a}_{i,2}\Sigma_{\mathbf{0}_i}(z_{i-1})^2}, \tag{7}$$

$$\Sigma_{\mathbf{a}_i}(z_{i-1}) = \frac{\Sigma_{\mathbf{0}_i}(z_{i-1})}{\sqrt{1 - 2\mathbf{a}_{i,2}\Sigma_{\mathbf{0}_i}(z_{i-1})^2}}. \tag{8}$$

For each m_i to have finite variance, it is clear from (8) that $\mathbf{a}_{i,2}$ must have the restriction $\mathbf{a}_{i,2} < 1/(2\Sigma^2_{\mathbf{0}_i})$.

2.4.3 The EIS regressions

The final piece of notation introduced is the fraction

$$\xi_i(z_{i-1}, x_i) = \frac{f_i(z_i, x_i|z_{i-1}, \theta, \Delta)}{B_i(z_i|z_{i-1}, x_i)}.$$

As B_i is the shape of the conditional density $f_i(z_i|z_{i-1}, x_i, \theta, \Delta)$, $\xi(z_{i-1}, x_i)$ is constant as a function z_i. The expression for ξ_i is given in appendix A.

Using this notation, we have

$$\begin{aligned}\frac{p(\mathbf{z}, \mathbf{x}|\theta, z_0)}{m(\mathbf{z}|\mathbf{a}, \mathbf{x}, z_0)} &= \prod_{i=1}^{n} \frac{f_i(z_i, x_i|z_{i-1}, \theta)}{m_i(z_i|z_{i-1}, x_i, \mathbf{a}_i)} = \prod_{i=1}^{n} \frac{\xi_i(z_{i-1}, x_i)\chi_i(z_{i-1}, x_i, \mathbf{a}_i)}{\psi_i(z_i, \mathbf{a}_i)} \\ &= \xi_1(z_0, x_1)\chi_1(z_0, x_1, \mathbf{a}_1) \left[\prod_{i=1}^{n-1} \frac{\xi_{i+1}(z_i, x_{i+1})\chi_{i+1}(z_i, x_{i+1}, \mathbf{a}_{i+1})}{\psi_i(z_i, \mathbf{a}_i)}\right] \frac{1}{\psi_n(z_n, \mathbf{a}_n)}. \end{aligned} \tag{9}$$

This last representation enables us to work out how the parameter $\mathbf{a}$ should be chosen to minimise MC variance using EIS type regressions. Firstly, we set $\mathbf{a}_n = [0, 0]$ so that the last fraction is equal to 1 for all z_n. In fact, setting $\mathbf{a}_n$ to zero effectively integrates out z_n analytically, and thus for $n = 1$ the procedure is exact. Secondly, under the assumption that z_0 is non-stochastic, $\xi_1\chi_1$ is also constant for fixed values of z_0 and does not add to the variance of the importance sampling procedure.

Finally, we notice that the log-variation (as a function of $\mathbf{z}$) for each of the factors in the bracketed product of (9) depends only on a single z_i. This gives rise to a recursive set of EIS ordinary least squares regression on the form

$$\begin{aligned}\log \xi_{i+1}(\tilde{z}_i^{(j)}, x_{i+1}) + \log \chi_{i+1}(\tilde{z}_i^{(j)}, x_{i+1}, \mathbf{a}_{i+1}) &= c_i + \log \psi_i(\tilde{z}_i^{(j)}, \mathbf{a}_i) + \eta_i^{(j)} \\ &= c_i + \mathbf{a}_{i,1}\tilde{z}_i^{(j)} + \mathbf{a}_{i,2}(\tilde{z}_i^{(j)})^2 + \eta_i^{(j)},\ i = 1, \ldots, n-1,\ j = 1, \ldots, S, \end{aligned} \tag{10}$$

where $\eta_i^{(j)}$ are the regression residuals. The constant term c_i is estimated jointly with $\mathbf{a}_i$. In particular, we notice that the regressions are linear in $\mathbf{a}_i$, suggesting that computationally efficient linear least squares algorithms may be applied. Notice also that the EIS regressions needs to be calculated backwards in time, as the ith regression depends on $\mathbf{a}_{i+1}$.

The MC variance of (5), represented by $\eta_i^{(j)}$, stems from the fact that the left hand side of (10) is non-quadratic (in z_i) and thus deviates from the quadratic model represented by $\log \psi_i$. Still, since

$\tilde{z}_i^{(j)}$, $j = 1, \ldots, S$, are typically strongly located by the information provided from the base-line density, the quadratic approximation works reasonably well.

A fortunate by-product of the EIS regressions is that the log-weights in the likelihood estimate (5) are directly expressible in terms of the regression residuals. More precisely, (9) provides us with the following expression:

$$\log \frac{p(\tilde{\mathbf{z}}^{(j)}, \mathbf{x}|\theta, z_0)}{m(\tilde{\mathbf{z}}^{(j)}|\mathbf{a}, \mathbf{x}, z_0)} = \log(\xi_1 \chi_1) + \sum_{i=1}^{n-1} \left[c_i + \eta_i^{(j)} \right], \; j = 1, \ldots, S, \tag{11}$$

provided that we have set $\mathbf{a}_n$ to zero. Thus, the estimate of the likelihood function can be calculated with very small effort once the relevant quantities in the regression models are calculated.

2.4.4 Iterative EIS and implementation

Since the $\tilde{\mathbf{z}}^{(j)}$s depend on $\mathbf{a}$, and $\mathbf{a}_{i+1}$ needs to be known to calculate $\mathbf{a}_i$, we may treat the regressions (10) as a fixed point condition for $\hat{\mathbf{a}}$, towards which we generate a convergent sequence of $\mathbf{a}^{(k)}$ for integers k. This is done using the following steps:

1. Set $\mathbf{a}^{(0)} = \mathbf{0}$, $k = 0$ and let $\mathbf{w} \in \mathbb{R}^{n \times S}$ denote a matrix filled with independent standard normal variates.

2. Simulate the paths $\tilde{\mathbf{z}}^{(j)} = \tilde{\mathbf{z}}^{(j)}(\mathbf{a}^{(k)})$, $j = 1, \ldots, S$ forward in time (i.e. for $i = 1 \to n-1$) using
$$\tilde{z}_i^{(j)} = \mu_{\mathbf{a}_i^{(k)}}(\tilde{z}_{i-1}^{(j)}, x_i) + \Sigma_{\mathbf{a}_i^{(k)}}(\tilde{z}_{i-1}^{(j)}) \mathbf{w}_{i,j} \text{ for } j = 1, \ldots, S, \; i = 1, \ldots, n-1,$$
where for simplicity we define $\tilde{z}_0^{(j)} = z_0$.

3. Calculate $\mathbf{a}_i^{(k+1)}$ backwards in time (i.e. for $i = n-1 \to 1$) by estimating the regression models (10) based on $\mathbf{a}_{i+1}^{(k+1)}$ in χ_{i+1} and the paths $\tilde{\mathbf{z}}^{(j)}(\mathbf{a}^{(k)})$.

4. Calculate the logarithm of the likelihood estimate (9) using the quantities calculated for the regressions in step 3, and stop the iteration if this estimate has converged to the desired precision.

5. $k \leftarrow k+1$ and return to step 2.

Following Richard and Zhang (2007), we apply the same set of canonical standard normal variates $\mathbf{w}$ to generate the paths in step 2 for each iteration. Moreover, this same set of canonical variates is used for each evaluation of the simulated log-likelihood function when doing the likelihood maximisation. This usage of common random numbers ensures the smoothness of the simulated log-likelihood function and allows us to apply a BFGS quasi-Newton optimiser (Nocedal and Wright (1999)) based on finite difference gradients. Another measure to keep the simulated log-likelihood function smooth is to terminate the EIS iteration when the change (from iteration k to $k+1$) in log-likelihood value is a small number. We have used a change of log-likelihood value of $< 1.0e-9$ as our stopping criterion.

The choice to apply a gradient-based optimisation algorithm stems from the fact that the model has up to eight parameters, and a simplex-type optimisation algorithm will generally require too many function evaluations to converge for such problems. The computational cost of the extra

EIS iterations needed to obtain this high precision are thus typically won back when using a faster optimisation algorithm. The typical number of EIS iterations required is between 20 and 40 to obtain precision of the order of $1.0e-9$. However, once such an evaluation is complete, computing the log-likelihood values needed for finite difference gradients can be much faster since we may start the EIS iteration using the previous $\hat{\mathbf{a}}$ and apply it for a slightly perturbed parameter vector. Typically, this approach requires about 5 to 10 iterations to converge.

One final detail to improve numerical stability is to add a simple line search, similar to those applied in line-searching optimisation algorithms, to the EIS iteration. This is done by regarding the difference in iterates $\mathbf{d}^{(k+1)} = \mathbf{a}^{(k+1)} - \mathbf{a}^{(k)}$ as a "search direction", which we may when necessary take shorter steps along. More precisely, when completing step 3 above, we set $\mathbf{a}^{(k+1)} = \mathbf{a}^{(k)} + \omega \mathbf{d}^{(k+1)}$, $\omega \in (0,1)$, if the "raw" iterate in step 3 leads to an infinite variance in the importance density or some other pathology.

Typical computing times for our FORTRAN90 implementation[4] range from 30 to 1000 seconds for locating a maximum likelihood estimator for data sets with around 2000 observations using a standard PC. The LAPACK (Anderson et al., 1999) routine `dgels` was used for the linear regressions, and all the random numbers where generated using Marsaglia's KISS generator (see e.g. Leong et al. (2005)).

3 Monte Carlo Experiments

To assess the statistical properties of the EIS-MC procedure outlined in Section 2, we have conducted some Monte Carlo experiments. The main objectives of these experiments are to quantify the error arising from the EM discreteization and from using EIS to integrate out the latent process.

The main sources (with no particular ordering) of statistical bias for the EIS-MC procedure are:

- Discreteization of the continuous time model using a EM-discreteization. An indirect way of diagnosing this bias is to look for unacceptable errors using EM-based maximum likelihood when the latent process is observed. In this manner, importance sampling is avoided, and hence can be eliminated as a source of error.
- Small sample biases from using the integrated likelihood function. Diagnostics for this is provided by comparing the EM-based MLEs when the log-volatility is observed and un-observed.
- MC errors from using the MC estimate (5) instead of using exact integration. These errors will be discussed in the next section by using many different random number seeds in the program.

All of the computations are done using a yearly time scale and with daily observations, corresponding to $\Delta = 1/252$. We use $S = 32$ paths in the importance sampler throughout both the MC simulations and the application to real data. All the MC experiments are done with both $\mathbf{z}$ observed and unobserved, i.e. by maximising (4) and (5), respectively, with respective to θ. Under observed $\mathbf{z}$, the simulated z_0 is applied, whereas under unobserved $\mathbf{z}$, we estimate z_0 along with the other parameters.

The "true" parameter values used to generate the synthetic data sets are the empirical parameter estimates obtained from the Standard & Poor 500 data set which we consider in Section 4. We

[4] The source code is availible on request from the first author.

parameter	true value	bias	std	MSE
Observed volatility, $n = 2022$				
α	0.2109	0.0081	0.0320	0.0011
β	-7.7721	-0.4366	1.4286	2.2272
σ	0.3774	-0.0042	0.0059	0.0001
ρ	-0.3162	0.0044	0.0190	0.0004
a	0.0591	-0.0072	0.0854	0.0073
b	1.6435	0.4320	4.0063	16.2035
Unobserved volatility, $n = 2022$				
α	0.2109	-0.0040	0.0601	0.0036
β	-7.7721	-0.1068	2.4411	5.9577
σ	0.3774	-0.0342	0.0493	0.0036
ρ	-0.3162	0.0194	0.1209	0.0150
a	0.0591	0.0344	0.1277	0.0175
b	1.6435	-1.0805	5.5070	31.4291

Table 1: Heston model MC study results. The bias is reported as $E[\hat{\theta} - \theta_{\text{true}}]$. "std" denotes the statistical standard errors and MSE denotes the mean square error around the "true parameter".

first consider the Heston model and the GARCH diffusion model, and then consider the full CEV model under two different simulation designs.

The synthetic data sets are simulated using the EM scheme with a time step of $\Delta/2048$, so that the simulated data sampled at the Δ-time-grid can be regarded as sampled from the continuous time model. The first 3000 data points for each simulation are discarded so that the simulated distribution of z_0 is close to the marginal distribution of z_0 dictated by the model. For all the experiments, we simulate and estimate 500 synthetic data sets.

3.1 Heston model

The results for the simulations under the Heston model are given in Table 1. We use sample size $n = 2022$, equal to the number of observations in real data set considered in Section 4.[5]

It is well known that MLEs of the mean reversion parameter tend to be biased toward faster mean reversion in finite samples in the context of diffusion processes with a linear drift (see for example, Phillips and Yu (2005) and Phillips and Yu (2009b)). For the CEV model specified in (1), a faster mean reversion rate corresponds to higher negative values of β. This is also seen under the Heston model both for observed and unobserved log-volatility. Interestingly, we find the effect is stronger for observed log-volatility. Still, the bias under the Heston model is smaller than the corresponding statistical standard errors, both for the observed and the unobserved log-volatility. Thus, it seems that all three sources of bias discussed above are controlled for this model and under the amount of data.

The loss of precision when using the integrated likelihood procedure ranges by a factor from 2 to 10 in increased statistical standard errors. In particular, the estimation precision of the volatility-of-volatility parameter σ and the leverage parameter ρ are increased by a factor close to 10.

[5] Under this simulation regime, 8% of the estimation replica for the unobserved log-volatility failed to converge and were subsequently ignored.

parameter	true value	bias	std	MSE
Observed volatility, $n = 2022$				
α	0.2411	0.0015	0.0323	0.0010
β	-9.3220	-0.3056	2.0270	4.1937
σ	2.8202	-0.0715	0.0430	0.0070
ρ	-0.2920	0.0020	0.0195	0.0004
a	0.1019	-0.0039	0.0881	0.0078
b	0.1139	0.3217	4.4037	19.4570
Unobserved volatility, $n = 2022$				
α	0.2411	0.0117	0.0756	0.0058
β	-9.3220	-0.8100	3.6413	13.8878
σ	2.8202	-0.0760	0.4254	0.1864
ρ	-0.2920	0.0371	0.1156	0.0147
a	0.1019	0.0407	0.1320	0.0190
b	0.1139	-1.4421	6.1166	39.4159

Table 2: GARCH diffusion model MC study results. See the caption of Table 1 for details.

3.2 The GARCH diffusion model

Simulation results for the GARCH diffusion model are summarized in Table 2. As before we use $n = 2022$.[6] For both observed and unobserved log-volatility, there is a negative bias in the mean reversion parameter β as one would expect. Once again we see that no biases are larger in magnitude than the corresponding standard errors when the log-volatility is unobserved. A downward bias in σ is seen for the observed log-volatility to be larger in magnitude than the corresponding standard error, but the bias is still small comparing with the "true" parameter. Again the loss of precision from using integrated MLE is the largest for the parameters σ and ρ, where again about a 10-fold increase in standard error is seen.

3.3 The CEV model

For the CEV model, we have performed two simulation studies which are summarized in Tables 3 and 4. We denote these parameter settings as P1 and P2, respectively, corresponding to columns 3 and 4 in Table 5.[7]

Under P1 as "true parameters" are the estimates obtained using the full set of S&P 500 returns, which include the October 1987 crash. The experiment is done using both $n = 2022$ and $n = 5000$ data points. From Table 3 we see that the EIS-MLE procedure produces downward biased estimates of γ when the log-volatility is unobserved. When the log-volatility is observed, this effect is negligible. The bias in γ also leads to substantial bias in the other parameters governing the volatility, as we expect the MLE estimates of them must have a strong correlation structure in distribution, even asymptotically. Increasing the sample size from 2022 to 5000 decreases the biases slightly, but it seems that very long time series will be needed to identify γ with a decent precision for when the true parameter is in this range.

[6] For this model, 2.8% of the simulation replications under unobserved log-volatility failed to converge.

[7] For P1, 4.2% of the replications failed to converge, while for P2, 6.2 % of the replications failed to converge.

parameter	true value	bias	std	MSE
Observed volatility, $n = 2022$				
α	0.0434	0.0102	0.0232	0.0006
β	-0.4281	-0.5903	1.5432	2.7252
σ	13.6298	-0.3713	1.3716	2.0153
ρ	-0.3317	0.0013	0.0188	0.0004
γ	1.5551	-0.0070	0.0266	0.0008
a	0.0820	-0.0095	0.0923	0.0086
b	0.8716	0.5788	4.4829	20.3912
Unobserved volatility, $n = 2022$				
α	0.0434	0.0634	0.0530	0.0068
β	-0.4281	-3.4599	2.6160	18.7998
σ	13.6298	-8.6680	3.7726	89.3375
ρ	-0.3317	0.0227	0.1355	0.0188
γ	1.5551	-0.3539	0.2202	0.1736
a	0.0820	0.0030	0.1191	0.0142
b	0.8716	0.1687	5.4970	30.1819
Observed volatility, $n = 5000$				
α	0.0434	0.0036	0.0138	0.0002
β	-0.4281	-0.2046	0.9445	0.9322
σ	13.6298	-0.4053	0.8616	0.9051
ρ	-0.3317	0.0014	0.0128	0.0002
γ	1.5551	-0.0066	0.0169	0.0003
a	0.0820	-0.0053	0.0557	0.0031
b	0.8716	0.2240	2.5162	6.3687
Unobserved volatility, $n = 5000$				
α	0.0434	0.0480	0.0320	0.0033
β	-0.4281	-2.6796	1.5569	9.5995
σ	13.6298	-8.5886	2.5369	80.1870
ρ	-0.3317	0.0309	0.0787	0.0071
γ	1.5551	-0.3082	0.1444	0.1158
a	0.0820	0.0154	0.0720	0.0054
b	0.8716	-0.5007	3.1073	9.8867

Table 3: CEV model (P1) MC study results. See the caption of Table 1 for details.

parameter	true value	bias	std	MSE
Observed volatility, $n = 1800$				
α	0.0754	0.0097	0.0216	0.0006
β	-3.4022	-0.5524	1.3048	2.0042
σ	1.7587	0.0538	0.2565	0.0686
ρ	-0.3912	0.0037	0.0191	0.0004
γ	1.0804	0.0072	0.0359	0.0013
a	-0.1811	-0.0204	0.1160	0.0138
b	13.8246	1.2152	6.2209	40.0973
Unobserved volatility, $n = 1800$				
α	0.0754	0.0231	0.0499	0.0030
β	-3.4022	-1.3109	2.7270	9.1392
σ	1.7587	-0.1653	1.6373	2.7025
ρ	-0.3912	0.0030	0.1618	0.0261
γ	1.0804	-0.1273	0.2449	0.0761
a	-0.1811	-0.0220	0.1755	0.0312
b	13.8246	1.5312	9.2881	88.4301
Observed volatility, $n = 5000$				
α	0.0754	0.0025	0.0122	0.0002
β	-3.4022	-0.1435	0.7612	0.5989
σ	1.7587	0.0370	0.1511	0.0242
ρ	-0.3912	0.0044	0.0120	0.0002
γ	1.0804	0.0067	0.0213	0.0005
a	-0.1811	-0.0078	0.0693	0.0049
b	13.8246	0.4200	3.6322	13.3426
Unobserved volatility, $n = 5000$				
α	0.0754	0.0044	0.0217	0.0005
β	-3.4022	-0.2826	1.1640	1.4320
σ	1.7587	-0.2897	0.8019	0.7257
ρ	-0.3912	0.0201	0.0876	0.0081
γ	1.0804	-0.0770	0.1500	0.0284
a	-0.1811	0.0101	0.0936	0.0089
b	13.8246	-0.3163	4.7425	22.5458

Table 4: CEV model (P2) MC study results. See the caption of Table 1 for details.

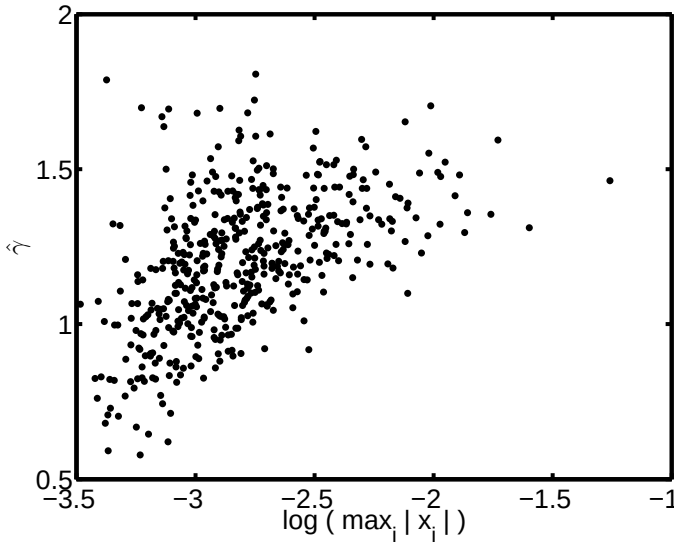

Figure 1: Estimated γ values explained by $\log(\max_i |x_i|)$ from the MC experiment with parameter setting 1 and $n = 2022$. The sample correlation equals 0.52.

To understand the impact of extreme observations (or "crashes") on this bias, we use the logarithm of the maximal absolute log-return as a simple proxy for "large" crash. In Figure 1, a scatter plot of the estimated γ against logarithm of the maximal absolute log-return is shown across all simulated paths under P1. A strong positive relationship is seen (correlation 0.52). This plot suggests that when the log-volatility is unobserved, extreme values in a sample are needed to identify high values of γ. As a reference, the maximal absolute log-return of the October 1987 crash is roughly $\exp(-1.64)$. In the simulated data sets, such extreme events occur in roughly 0.4% of the data sets.

For the P2 "true parameters" obtained using the data excluding the October 1987 crash, we see much smaller biases, and the downward bias in γ decrease substantially when the sample size is increased from 1800 to 5000. Still, all the biases are smaller in magnitude than the corresponding statistical standard error.

4 Application to real data

For our application to real data, we use Standard & Poor 500 log-return data previously used in Jacquier et al. (1994) and Yu (2005).[8] The data covers the period January 1980 to December 1987 and the sample has a total of $n = 2022$ log-return observations.

We also fit the following continuous time log-normal (LN) model using the proposed method:

$$d\begin{bmatrix} \bar{s}_t \\ \bar{z}_t \end{bmatrix} = \begin{bmatrix} a + b\exp(\bar{z}_t) \\ \alpha + \beta \bar{z}_t \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)}\exp(\bar{z}_t/2) & \rho\exp(\bar{z}_t/2) \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} dB_{t,1} \\ dB_{t,2} \end{bmatrix}. \tag{12}$$

[8] The log-return data are multiplied by 0.01.

parameter	Heston	GARCH	CEV	LN	CEV1800	LN1800
α	0.2109	0.2411	0.0434	-29.5044	0.0754	-17.4672
	[0.0035]	[0.0020]	[0.0028]	[0.2304]	[0.0027]	[0.1234]
	(0.0601)	(0.0756)	(0.0530)	(8.7924)	(0.0499)	(8.5628)
β	-7.7721	-9.3220	-0.4281	-7.6953	-3.4022	-4.4413
	[0.1395]	[0.0860]	[0.1495]	[0.0593]	[0.1489]	[0.0313]
	(2.4411)	(3.6413)	(2.6160)	(2.2791)	(2.7270)	(2.2166)
σ	0.3774	2.8202	13.6298	2.4793	1.7587	1.2965
	[0.0036]	[0.0115]	[0.3321]	[0.0112]	[0.3088]	[0.0063]
	(0.0493)	(0.4254)	(3.7726)	(0.3313)	(1.6373)	(0.2664)
ρ	-0.3162	-0.2920	-0.3317	-0.3146	-0.3912	-0.3934
	[0.0017]	[0.0006]	[0.0023]	[0.0007]	[0.0010]	[0.0010]
	(0.1209)	(0.1156)	(0.1355)	(0.0946)	(0.1618)	(0.1525)
γ	0.5	1.0	1.5551	-	1.0804	-
	-	-	[0.0075]	-	[0.0432]	-
	-	-	(0.2202)	-	(0.2449)	-
a	0.0591	0.1019	0.0820	0.0683	-0.1811	-0.2021
	[0.0013]	[0.0005]	[0.0011]	[0.0005]	[0.0044]	[0.0008]
	(0.1277)	(0.1320)	(0.1191)	(0.1102)	(0.1755)	(0.1732)
b	1.6435	0.1139	0.8716	1.4183	13.8246	14.8801
	[0.0459]	[0.0209]	[0.0419]	[0.0196]	[0.2251]	[0.0409]
	(5.5070)	(6.1166)	(5.4970)	(4.9598)	(9.2881)	(8.7362)
log-likelihood	6514.90	6541.42	6552.09	6538.23	5916.20	5916.44
	[0.2457]	[0.0823]	[0.3494]	[0.1502]	[0.0457]	[0.0646]

Table 5: Parameter estimation results for the Standard and Poor 500 data. Monte Carlo standard errors are given squared brackets and statistical standard errors, taken from the MC experiments, are given in parentheses. For the Heston model and the CEV model, 1 of the 100 estimation replications failed to converge. Columns 5 and 6 report parameter estimates obtained when including only the first 1800 observations in the sample.

There are two purposes for fitting the LN model. First, it is used to illustrate the flexibility of our estimation methods. Second, it would be empirically interesting to compare the performance of the LN model with that of the CEV models. The discreteization and the recipe for adapting the EIS algorithm to this model are given in Appendix B.

Parameter estimates along with statistical- and MC standard errors for the Heston model, the GARCH diffusion, the full CEV model and the LN model are given in columns 1-4 in Table 5 respectively.

In addition, we have included parameter estimates for the CEV model and the LN model when only the first 1800 observations (and thus excluding the Oct-87 crash) are used in column 5 and 6. The statistical standard errors are taken from the MC experiments reported in Tables 1-4 for the CEV model instances. For the LN model, the statistical standard errors are MC estimates based on 100 synthetic data sets.

As both the Heston model and the GARCH diffusion model are special cases of the CEV model, it is sensible to compare the maximum likelihood values reported in the last row of the Table. The likelihood ratio test suggests that there is strong empirical evidence against the Heston model. This empirical result reinforces what have been found when both the spot prices and option prices are joint used to estimate the CEV-SV model (Jones (2003); Aït-Sahalia and Kimmel (2007)). Moreover, for the GARCH diffusion model, when the complete data set is used, the likelihood ratio test gives rejection for any practical p-value comparing with the complete CEV model. For the shorter data set, we see that the estimate for γ is less than one half standard error from that of the GARCH diffusion model.

The estimates for the leverage effect parameter ρ are very much in accordance with the estimates of Yu (2005) (posterior mean $= -0.3179$) under the log-normal stochastic volatility model. In all cases we obtained a positive estimate of b, suggesting a positive risk-return relation, but the parameter estimates are statistically insignificant.

The parameter estimates of the CEV model with and without the October 1987 crash differ significantly. This observation suggests a poor identification of γ and that the influence of the crash to the estimate of γ when the log-volatility is unobserved is substantial. The finding is consistent with that in Jones (2003), even though he uses data from 1986-2000 and 1988-2000 along with implied volatility data. For the data set including the October 87 crash, Jones (2003) obtains a posterior mean of 1.33 for the CEV parameter γ. The corresponding estimated value for data excluding the October 87 crash is 1.17. Our simulated maximum likelihood estimates for γ are 1.56 and 1.08 respectively. Jones (2003) argues that to accommodate the large spike in volatility represented by the October 87 crash, higher values of γ and σ are needed. Still, since Jones (2003) used both log-return and implied volatility data, it is expected that his parameter estimates differ less then ours with and without the October 1987 crash in the sample.

The estimation results based on the LN SV model are reported in Column 4 for the full sample and in Column 6 for the subsample. The LN SV model is also been successfully estimated in both cases. For example, when the full sample is used, the estimates for the leverage effect parameter ρ are very much in accordance with the estimates of Yu (2005). A comparison of the log-likelihood values of the CEV model and the LN model reveals that for the full sample the 7-parameter CEV model outperforms the 6-parameter LN model. However, for the subsample, the LN model has a slightly higher likelihood value even with fewer parameters.

To estimate the errors induced by integrating out the log-volatility using the above described EIS-MC method (comparing with the exact unknown integral), we repeat the estimation process 100 times using different random number seeds. These MC standard errors for the parameters and

maximum log-likelihood values are included in square-parenthesizes in Table 5. It is seen that the MC errors are generally small comparing with the statistical standard errors. Judging from the MC standard errors in maximum log-likelihood estimates, the EIS-MC method performs best when γ is close to 1.

As references for the standard errors of the maximum log-likelihoods, we may mention that Liesenfeld and Richard (2006) obtains a MC standard error of 0.11 (log-likelihood: 918) under a three parameter log-normal SV model with 945 latent variables using 30 paths in the importance sampler. For a 5 parameter time-discretetized Heston model, Durham (2006) obtains a MC standard error of 2.49 (log-likelihood: 18473) using 1024 draws in a Laplace importance sampler. As the latent process under consideration here is both non-linear and heteroskedastic, the standard errors reported in Table 5 are satisfactory. Comparing with the findings of Kleppe and Skaug (2009), much of this may be written back to constructing the importance sampler around the product of conditional transition densities, rather than around the natural sampler as is commonly done in other applications of the EIS algorithm to non-linear state space models.

5 Concluding Remarks

This paper outlines how the EIS algorithm may be applied to integrate out a latent process in an EM-discretetized stochastic differential equation model. In terms of numerical precision, we find that the algorithm performs very well when considering the non-linear and heteroskedastic structure of the latent process. In terms of the application to the CEV model, we find that the integrated MLEs obtained performs well for moderate values of γ, but the identification is more difficult for higher values of γ.

One direction for further research is to use the improved (relative to the EM) approximate continuous time TPDs proposed in Aït-Sahalia (2008) and for jump-diffusions in Yu (2007). Using a simple Taylor expansion of these approximations (in z_i) one can obtain estimates of the conditional transition densities (i.e. conditional on x_i) that stays within the locally Gaussian importance samplers. The inclusion of jumps in the model will probably also improve the identifiably of the complete CEV model, as large returns may be regarded as jumps rather than be caused by large spikes in the volatility process. As a result, the volatility series will be smoother and hence it can be expected that the finite sample estimation bias of the mean reversion parameter will be more serious.

Moreover, it should be noticed that this procedure is by no means restricted to the CEV family of models. As shown in Appendix B, only the three model-dependent functions $\mu_{\mathbf{0}_i}(z_{i-1}, x_i)$, $\Sigma_{\mathbf{0}_i}(z_{i-1})$ and $\xi_i(z_{i-1}, x_i)$ needs to be re-specified to implement a different model. The EM scheme suggests that any stochastic differential equation has an approximate Gaussian TPD for sufficiently short time-steps Δ. Thus, the technique of using the conditional-on-data EM-TPD can be applied provided that data are given over a fine enough time-grid. In particular, due to the explicit nature of Gaussian conditional densities, multivariate extensions (towards both multiple observed and unobserved processes) should also be straightforward.

It is also worth noticing that the above outlined procedure is closely related to the Laplace accelerated sequential importance sampling (LASIS) procedure of Kleppe and Skaug (2009). In the setting of the EM-discretized CEV model, their procedure would be equivalent to applying a Laplace importance sampler in $\mathbf{w}$ (which are standard normal) in stead of $\mathbf{z}$. This procedure would then bypass the much problems of heteroscedasticity and non-linearity in much the same manner as outlined here, but we do not make further comparisons here.

References

Aït-Sahalia, Y. (2002). Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. *Econometrica 70*, 223–262.

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Annals of statistics 36*(2), 906–937.

Aït-Sahalia, Y. and R. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics 134*, 507–551.

Andersen, T. G., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance 57*(3), 1239–1284.

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

Anderson, E., Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen (1999). *LAPACK Users' Guide* (Third ed.). Society for Industrial and Applied Mathematics.

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *Journal of Finance 52*, 2003–2049.

Bauwens, L. and F. Galli (2009). Efficient importance sampling for ml estimation of scd models. *Computational Statistics and Data Analysis 53*, 1974–1992.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

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

Durham, G. B. (2006). Monte carlo methods for estimating, smoothing, and filtering one and two-factor stochastic volatility models. *Journal of Econometrics 133*, 273–305.

Durham, G. B. and A. R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). *Journal of Business and Economic statistics 20*(3), 297–338.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bonds and currency options. *Review of financial studies 6*, 327–343.

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

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics 12*(4), 371–89.

Jones, C. S. (2003). The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of econometrics 116*, 181–224.

Kleppe, T. S. and H. Skaug (2009). Fitting general stochastic volatility models using laplace accelerated sequential importance sampling. Submitted for publication.

Kloeden, P. E. and E. Platen (1999). *Numerical solution of stochastic differential equations*. Springer-Verlag.

Leong, P. H. W., G. Zhang, D. U. Lee, W. Luk, and J. Villasenor (2005). A comment on the implementation of the ziggurat method. *Journal of Statistical Software 12*(7), 1–4.

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

Liesenfeld, R. and J.-F. Richard (2006). Classical and bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews 25*(2), 335–360.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics 45*, 7–38.

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

Phillips, P. and J. Yu (2005). Jackknifing bond option prices. *Review of Financial Studies 18*, 707–742.

Phillips, P. and J. Yu (2009a). Maximum likelihood and gaussian estimation of continuous time models in finance. In e. a. Torben Andersen (Ed.), *Handbook of Financial Time Series*, pp. 497–530. Springer.

Phillips, P. C. B. and J. Yu (2009b). Simulation-Based Estimation of Contingent-Claims Prices. *Rev. Financ. Stud. 22*(9), 3669–3705.

Rao, B. L. S. P. (1999). *Statistical inference for diffusion type processes*. Number 8 in Kendall's library of statistics. Arnold.

Richard, J.-F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics 141*(2), 1385 – 1411.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika 84*, 653–667.

Wiggins, J. (1987). Option values under stochastic volatility: Theory and empirical estimate. *Journal of Financial Economics 19*, 351–372.

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics 127*, 165–178.

Yu, J. (2007). Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the chinese yuan. *Journal of Econometrics 141*(2), 1245 – 1280.

A Appendix: Explicit expressions

The explicit expression for $\log \chi_i$ is given as

$$\log \chi_i(z_{i-1}, x_i, \mathbf{a}_i) = \frac{1}{2}\log(\pi) - \frac{1}{2}\log\left(\frac{1}{2\Sigma_{\mathbf{0}_i}(z_{i-1})^2} - \mathbf{a}_{i,2}\right) - \frac{\mu_{\mathbf{0}_i}(z_{i-1}, x_i)^2}{2\Sigma_{\mathbf{0}_i}(z_{i-1})^2} - \frac{\left(\frac{\mu_{\mathbf{0}_i}(z_{i-1}, x_i)}{\Sigma_{\mathbf{0}_i}(z_{i-1})^2} + \mathbf{a}_{i,1}\right)^2}{4\left(\mathbf{a}_{i,2} - \frac{1}{\Sigma_{\mathbf{0}_i}(z_{i-1})^2}\right)}.$$

Moreover, $\log \xi_i$ is given as

$$\log \xi_i(z_{i-1}, x_i) = -\log(2\pi\Delta) + \frac{z_{i-1}(1-2\gamma)}{2} - \frac{1}{2}\log\left(\sigma^2(1-\rho^2)\right) - \frac{\left(x_i - \Delta\left(a + b\exp(z_{i-1})\right)\right)^2}{2\Delta\exp\left(z_{i-1}\right)}.$$

B The log-normal model

In this paper, we use the following specification of the continuous time log-normal stochastic volatility model:

$$d\begin{bmatrix} \bar{s}_t \\ \bar{z}_t \end{bmatrix} = \begin{bmatrix} a + b\exp(\bar{z}_t) \\ \alpha + \beta\bar{z}_t \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)}\exp(\bar{z}_t/2) & \rho\exp(\bar{z}_t/2) \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} dB_{t,1} \\ dB_{t,2} \end{bmatrix}. \tag{13}$$

The Euler-Maruyama scheme yields the discrete time dynamics

$$\begin{bmatrix} x_{i+1} \\ z_{i+1} \end{bmatrix} = \begin{bmatrix} \Delta(a + b\exp(z_i)) \\ z_i + \Delta(\alpha + \beta z_i) \end{bmatrix} + \sqrt{\Delta}\begin{bmatrix} \sqrt{(1-\rho^2)}\exp(z_i/2) & \rho\exp(z_i/2) \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \end{bmatrix} \tag{14}$$

which with $a = b = 0$ is equivalent to the ASV1 specification in Yu (2005).

To adapt the above described EIS algorithm, the following functions needs to be altered:

$$\mu_{\mathbf{0}_i}(z_{i-1}, x_i) = z_{i-1} + \Delta(\alpha + \beta z_{i-1}) + \sigma\rho\exp(-z_{i-1}/2)(x_i - \Delta(a + b\exp(z_{i-1})), \tag{15}$$

$$\Sigma_{\mathbf{0}_i}(z_{i-1}) = \sigma\sqrt{\Delta(1-\rho^2)}, \tag{16}$$

$$\log \xi_i(z_{i-1}, x_i) = -\log(2\pi\Delta) - z_{i-1}/2 - \log(\sigma^2(1-\rho^2)) - \frac{\left(x_i - \Delta(a + b\exp(z_{i-1}))\right)^2}{2\Delta\exp(z_{i-1})}. \tag{17}$$

This highlights the fact that the above EIS algorithm is easily adapted to other models cast in the form of EM discretetized stochastic differential equations. For the computations summarized in Table 5, we use $M = 32$ draws in the importance sampler.