Simulated maximum likelihood for general stochastic volatility models: a change of variable approach

Kleppe, Tore Selland and Skaug, Hans J.

10 July 2008

Simulated maximum likelihood for general stochastic volatility models: a change of variable approach

Tore Selland Kleppe*and Hans Julius Skaug†

July 10, 2008

Abstract

Maximum likelihood has proved to be a valuable tool for fitting the log-normal stochastic volatility model to financial returns time series. Using a sequential change of variable framework, we are able to cast more general stochastic volatility models into a form appropriate for importance samplers based on the Laplace approximation. We apply the methodology to two example models, showing that efficient importance samplers can be constructed even for highly non-Gaussian latent processes such as square-root diffusions.

Keywords: Change of Variable, Heston Model, Laplace Importance Sampler, Simulated Maximum Likelihood, Stochastic Volatility

*Department of Mathematics, University of Bergen, Johannes Brunsgate 12, 5008 Bergen, Norway, Email: tore.kleppe@math.uib.no (Corresponding author)

†Department of Mathematics, University of Bergen, Johannes Brunsgate 12, 5008 Bergen, Norway, Email: skaug@math.uib.no

1 Introduction

During the last two decades, a vast literature on fitting stochastic volatility (SV) models to price return data has emerged. Parameter estimation in such models is made difficult by the presence of a latent volatility process. The recent approaches follow essentially two lines of attack for integrating out the volatility: simulated maximum likelihood (SML) (e.g. Danielsson and Richard (1993); Danielsson (1994); Shepard and Pitt (1997); Sandmann and Koopman (1998); Liesenfeld and Richard (2003, 2006); Durham (2006, 2007); Richard and Zhang (2007)) and Markov chain Monte Carlo (MCMC) (e.g. Omori et al. (2007) and refrences therein). In this paper we seek to extend the class of SV models that can be efficiently fit using SML, and hence to provide access to the maximum likelihood toolbox.

Earlier SML approaches are mainly focused around extensions of the discrete-time log-normal SV model (Taylor, 1982)

$$X_t = \sqrt{V_t}\,\eta_t,\ t = 1, \ldots, n \tag{1}$$

$$V_t = \sigma_X^2 \exp(U_t) \tag{2}$$

$$U_t = \phi U_{t-1} + \sigma Z_t,\ t = 2, ..., n \tag{3}$$

$$U_1 = \frac{\sigma}{\sqrt{1-\phi^2}} Z_1 \tag{4}$$

where (ϕ, σ, σ_X) are parameters and $(\eta_t, Z_t),\ t = 1, \ldots, n$ are i.i.d. standard Gaussian variates. Here, X_t is the process of price log-returns and V_t is the volatility process. The Gaussian AR(1) process U_t is typically highly autocorrelated, with a small transition variance σ^2. The success of Laplace approximation (applied to $\mathbf{U} = [U_1, \ldots, U_n]'$), and the associated SML, may be traced back to the fact that the conditional probability density function (PDF) of $\mathbf{U}$ given $\mathbf{X}$ deviates only to a little extent from a multivariate Gaussian PDF.

A large number of other SV models, in particular in continuous-time, have been proposed in literature (e.g. Nelson (1990); Heston (1993); Barndorff-Nielsen and Shephard (2001); Durbin and Koopman (2001); Jones (2003); Nielsen and Shephard (2003)) and are typically specified as a bi-variate stochastic process $\{(X_t, V_t)\}$ where $V_t > 0$ is a Markov process and t indexes either continuous- or discrete-time. Inspired by the high accuracy of Laplace-based SML for (1 - 4) in U_t, we introduce a Gaussian white noise process Z_t and re-specify the original model (X_t, V_t) as (X_t, Z_t) so that the marginal distribution (and hence the likelihood) of X_t sampled at discrete times $\mathbf{X}$ is equal in both models. Application of Laplace importance samplers in the corresponding $\mathbf{Z}$ yield a rapidly converging approximation of the likelihood function, and for a fixed number of importance samples, this approximate likelihood function can be maximized to obtain parameter estimates and approximate test statistics.

A prerequisite for this approach is that the PDF of $(\mathbf{X}, \mathbf{V})$ at a collection of time points $1, \ldots, n$ may be written on the form

$$p_{\mathbf{X},\mathbf{V}}(\mathbf{x}, \mathbf{v}) = p_{X_1,V_1}(x_1, v_1) \prod_{i=2}^{n} p_{X_i,V_i|V_{i-1}}(x_i, v_i|v_{i-1}) \tag{5}$$

and that the transition probability density (TPD) $p_{X_i,V_i|V_{i-1}}(x_i, v_i|v_{i-1})$ can be evaluated efficiently. For continuous-time models, the observations need not to be at equidistant times as long as the TPDs can be calculated accordingly.

The rest of the paper is laid out as follows. Section 2 is devoted to an outline of the proposed change of variable methodology, and we illustrate some important features through simple examples. Some issues of implementation are also discussed. In section 3 we illustrate the methodology through fitting two example models to the classic Dollar/Pound dataset (Harvey et al., 1994). The first model considered is the log-normal model (1 - 4), but we depart from earlier work by disregarding the underlying Gaussian process U_t. The second

model is a semi-discrete version of the Heston model (Heston, 1993), where the volatility process follows a continuous-time square-root diffusion. Finally, section 4 provides some discussion.

2 Methodology

In this section, we outline the proposed methodology to transform a general stochastic volatility model likelihood problem into a form suitable for Laplace importance sampler analysis. First, we review some basic facts concerning variable transforms and the Laplace approximation. Then we motivate and state the general sequential change of variable map that constitutes the core of this work. Finally we consider implementation issues.

2.1 Changes of variables and Laplace approximations

At the core of any textbook in multivariate calculus is the change of variable formula for integrals

$$\int_{\mathbb{R}^n} f(\mathbf{v})d\mathbf{v} = \int_{\mathbb{R}^n} f(\psi(\mathbf{z}))|\nabla\psi(\mathbf{z})|d\mathbf{z}, \tag{6}$$

where $\psi : \mathbb{R}^n \to \mathbb{R}^n$ and one-to-one, and $|\nabla\psi(\mathbf{z})|$ is the Jacobian determinant of ψ (Sydsæter et al., 1999). When n is large, the Laplace approximation (Barndorff-Nielsen and Cox, 1989) and related importance samplers (Kuk, 1999; Skaug, 2002) are the workhorses for approximating integrals on this form. The Laplace approximation is given as

$$\int_{\mathbb{R}^n} f(\mathbf{v})d\mathbf{v} \simeq f(\hat{\mathbf{v}})\frac{(2\pi)^{n/2}}{\sqrt{|\nabla^2 g(\hat{\mathbf{v}})|}},\ \hat{\mathbf{v}} = \arg\min_{\mathbf{v}}\ g(\mathbf{v}),\ g(\mathbf{v}) = -\log f(\mathbf{v}) \tag{7}$$

where $|\nabla^2 g(\hat{\mathbf{v}})|$ is the determinant of the Hessian of $g(\mathbf{v})$ at the minimizer $\hat{\mathbf{v}}$. It is easily verified that the Laplace approximation is nothing more than approximating the integrand with an un-normalized $N(\hat{\mathbf{v}}, \left[\nabla^2 g(\hat{\mathbf{v}})\right]^{-1})$ density and calculating the exact integral over the approximate integrand. We refer to this Gaussian PDF as the Laplace approximating density (LAD). In itself, the Laplace approximation is fairly accurate over a large class of integrals, but the fixed accuracy is a drawback. To work around this, we use Laplace importance samplers (LIS)

$$I_{LIS} = \frac{1}{m} \sum_{i=1}^{m} \frac{f(\mathbf{v}_i)}{p_{LAD}(\mathbf{v}_i)}$$

with $\mathbf{v}_i, i = 1, \ldots, m$ iid sampled from the LAD. Provided that $var(I_{LIS}) < \infty$, I_{LIS} converges strongly, as $m \to \infty$, to the exact value of the integral. We shall use the terms Laplace approximation and Laplace importance samplers more or less interchangeably.

Combining (6) with the LIS sets the stage for the rest of the discussion here. Since the change of variable map ψ is to our disposition, we hope that by approximating the right hand side of (6) with a LIS, the Monte Carlo variance can be reduced significantly compared to applying a LIS to the left hand side.

Some properties of the Laplace approximation deserve mentioning here. The Laplace approximation is exact when the integrand is an un-normalized Gaussian density. As noted in Butler (2007), the Laplace approximation is invariant under affine changes of variable in the original integral. Both of these properties extends trivially to the LIS and points out some directions as how to choose ψ. An affine ψ will not improve anything, and the right hand side integrand in (6) should be close to an un-normalized Gaussian density.

If the Laplace approximation is applied to a marginalization integral in a latent variable model, say $(\mathbf{X}, \mathbf{V})$ with $\mathbf{V}$ latent, then the mode $\hat{\mathbf{v}}$ is the empirical Bayes MAP estimate of $\mathbf{V}$ provided that the parameters are at their maximum

likelihood estimate (Carlin and Louis, 1996). In the context of SV models, this may be applied to smooth and filter the volatility process.

2.2 A toy example

Consider the following one-dimensional example. Let $X \sim N(0, V)$ and $V \sim lognormal(0, \sigma)$. This may be thought of as a special case of the SV model (1 - 4) with only one observation. To calculate the PDF of X marginally, i.e. the likelihood, we need to integrate out the latent V:

$$\begin{aligned} p_X(x) &= \int p_{X|V}(x|v)p_V(v)dv \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi v}} \exp(-\frac{x^2}{2v}) \frac{1}{v\sigma\sqrt{2\pi}} \exp(-\frac{(\log(v))^2}{2\sigma^2}) \mathbf{1}_{\{v>0\}}(v)dv, \end{aligned} \quad (8)$$

where $\mathbf{1}$ denotes the indicator function. Now, consider the change of variable map $v = \psi(z) = F^{-1}(\Phi(z)) = \exp(\sigma z)$, where Φ denotes the standard Gaussian cumulative distribution function (CDF) and F is the CDF associated with p_V. Simple manipulations yield that the integral (8) may be rewritten as

$$p_X(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi \exp(\sigma z)}} \exp(-\frac{x^2}{2\exp(\sigma z)}) \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2})dz. \quad (9)$$

Simple probabilistic arguments suggest that the variance of LISs applied to both integrals converges to 0 as $\sigma \to 0$. However, the order of magnitude in the mean square error (MSE) of LISs applied to the two differ significantly as depicted in Figure 1. There are several partial explanations to this. First, since the integral (8) is taken over the positive half line, there is always positive probability that samples from the LAD hit outside the support, slowing down convergence of the importance sampler. By the simple change of variable, this is fixed for in (9). The most prominent feature of this comparison is still the fact that the rate at

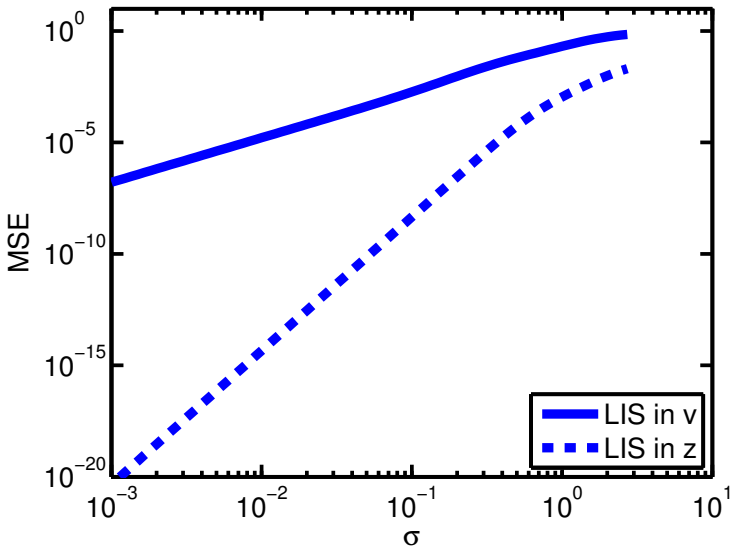

Figure 1: Monte Carlo estimates of the mean square error (MSE) in the calculation of p_X in section 2.2 using LISs with $m = 32$ importance samples for different values of the parameter σ. The MSEs are estimated over 1000 runs of the importance samplers for each value of σ. "LIS in v" refers to the LIS applied to (8) whereas "LIS in z" refers to a LIS applied to (9).

which (9) approaches an un-normalized Gaussian PDF is much higher, yielding a faster converging importance sampler.

All previous authors who have applied SML to the SV model (1 - 4) have applied an approximation to an integral corresponding to (9), and the above example may thus seem somewhat artificial. However, more general SV models are naturally cast in terms of the non-Gaussian variable V, leading to an integral similar to (8). Maps on the form $\psi(z) = F^{-1}(\Phi(z))$ constitute the backbone in our methodology for transforming integrals from the form (8) to (9).

Another change of variable in (9), say $u = \sigma z$, would have brought us closer to the practice of integrating in the $\mathbf{U}$ variables in the log-normal SV model. This illustrates that it irrelevant whether the "small number" representing the small transition variance in the in the volatility process sits in ψ or in the un-

derlying Gaussian as the Laplace methodology is invariant under linear variable changes. Throughout this paper, we will for simplicity, leave it in ψ and let the "driving" vector $\mathbf{Z}$ be standard Gaussian white noise.

2.3 A sequential change of variable framework

The map introduced in the previous example generalizes naturally to the sequential change of variable framework

$$\mathbf{v} = \psi(\mathbf{z}) = \begin{bmatrix} \psi_1(\mathbf{z}) \\ \psi_2(\mathbf{z}) \\ \vdots \\ \psi_{n-1}(\mathbf{z}) \\ \psi_n(\mathbf{z}) \end{bmatrix} = \begin{bmatrix} F_1^{-1}(\Phi(z_1)) \\ F_2^{-1}(\Phi(z_2), v_1) \\ \vdots \\ F_{n-1}^{-1}(\Phi(z_{n-1}), v_{n-2}) \\ F_n^{-1}(\Phi(z_n), v_{n-1}) \end{bmatrix}. \tag{10}$$

For now, we let the F_is be absolutely continuous CDFs in their first argument. Specific choices are discussed shortly. It is easily verified, by an induction argument forward in time, that (10) is a one-to-one mapping.

The strict one-step-backward dependence is introduced to mimic the Markov structure of $(\mathbf{X}, \mathbf{V})$ and has the pleasant feature of a triangular Jacobian. This last fact leads to a particularly simple formula for the log-Jacobian determinant:

$$\log|\nabla\psi(\mathbf{z})| = -\frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\frac{1}{2}z_i^2 - \log p_1(v_1(\mathbf{z})) - \sum_{i=2}^{n}\log p_i(v_i(\mathbf{z}), v_{i-1}(\mathbf{z})),$$

where p_i is the PDF (in its first argument) defined by $p_i(v_i, v_{i-1}) = \partial/\partial v_i\, F(v_i, v_{i-1})$. Combining the log-determinant with the joint PDF of the SV model (5) yields

the general modified negative log-integrand in $\mathbf{z}$.

$$g^*(\mathbf{z}) = \frac{n}{2}\log(2\pi)+\frac{1}{2}\sum_{i=1}^{n} z_i^2-[\log p_{X_1|V_1}(x_1|v_1(\mathbf{z}))+\log p_{V_1}(v_1(\mathbf{z}))-\log p_1(v_1(\mathbf{z}))]$$
$$-\sum_{i=2}^{n}\left[\log p_{X_i|V_i}(x_i|v_i(\mathbf{z})) + \log p_{V_i|V_{i-1}}(v_i(\mathbf{z})|v_{i-1}(\mathbf{z})) - \log p_i(v_i(\mathbf{z}), v_{i-1}(\mathbf{z}))\right]. \tag{11}$$

2.4 Choices of F_i

In this work, we consider two specific choices of the F_is that are rather apparent given (11). These are chosen under the constraints that they are easy to evaluate and invert, and that they apply to a broad class of models.

If we take $F_i,\ i = 2,\ldots,n$ to be the CDFs of $V_i|V_{i-1}$, and let F_1 be the limiting CDF of V_t, we get a sequential change of variable map which we will denote $\psi^{(1)}$. Under this formulation, the negative log-integrand (11) takes the form:

$$g^*_{(1)}(\mathbf{z}) = \frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{i=1}^{n} z_i^2 - \sum_{i=1}^{n}\log p_{X_i|V_i}(x_i|v_i(\mathbf{z}))$$
$$= \frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{i=1}^{n} z_i^2 - \sum_{i=1}^{n} M_i^{(1)}(v_i(\mathbf{z})). \tag{12}$$

Typically, if the volatility process has small variance relative to 1, $\mathbf{v}(\mathbf{z})$ will vary slowly and the standard Gaussian part will dominate the total variation of $g^*_{(1)}$. Under $\psi^{(1)}$, the distributions of $\mathbf{V}$ and $\psi^{(1)}(\mathbf{Z})$ are equal, and we may interpret the methodology as simply adding a new level in the hierarchical representation of the SV model.

We denote by $\psi^{(2)}$ the sequential change of variable map when $F_i,\ i = 2,\ldots,n$ are taken to be the CDFs of $V_i|V_{i-1}, X_i$ and F_1 to be the CDF of $V_1|X_1$. Since $p_i,\ i > 2$ in this case may be written as $p_{X_i,V_i|V_{i-1}}/p_{X_i|V_{i-1}}$, the

negative log-integrand has the form

$$g^*_{(2)}(\mathbf{z}) = \frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{i=1}^{n} z_i^2 - \sum_{i=2}^{n} \log p_{X_i|V_{i-1}}(x_i|v_{i-1}(\mathbf{z})) - p_{X_1}(x_1)$$
$$= \frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{i=1}^{n} z_i^2 - \sum_{i=2}^{n} M_i^{(2)}(v_{i-1}(\mathbf{z})) - M_1^{(2)}.$$

It is reasonable to assume, and we shall see this in the examples, that $\psi^{(2)}$ should lead to faster convergence of the LIS since $M^{(2)}$ depends more weakly on $\mathbf{z}$ than does $M^{(1)}$. The drawback is that the $\psi^{(2)}$ formulation typically leads to non-standard CDFs that need to be evaluated using numerical integration techniques.

By construction, choosing one of these maps resolves the issues concerning the joint support of the latent process. This is of particular importance in models where the support of $V_i|V_{i-1}$ depends on V_{i-1}, such as in the exponential AR(1) process proposed by Nielsen and Shephard (2003) or the gamma AR(1) processes proposed by Durbin and Koopman (2001).

A few more points deserve mentioning here. In both formulations, it is not essential to have an explicit limiting distribution for the volatility process. This can be worked around by starting the latent process in some fixed point at some time prior to where we have data, and then let it be integrated out. Missing data can be handled equally well in this manner. Moreover, the explicit splitting of the PDFs into transition and observation parts seen in (11) is not strictly necessary as the cancellation effects can be made implicitly in the computations.

2.5 A simple explicit example

Consider a modified version of the SV model proposed in Nielsen and Shephard (2003),

$$\begin{aligned} X_t =& \sqrt{V_t + \gamma}\, \eta_t,\ t = 1, \ldots, n \\ V_t =& \phi V_{t-1} + \lambda \varepsilon_t,\ t = 2, \ldots, n \\ V_1 =& \frac{\lambda}{1-\phi} \varepsilon_1, \end{aligned}$$

where $\eta_t, t = 1, \ldots, n$ are iid standard Gaussian, $\varepsilon_t,\ t = 1, \ldots, n$ are i.i.d. $exp(1)$ and $\lambda, \gamma > 0$, $\phi \in [0, 1)$ are parameters. For simplicity, we have chosen the start condition so that it matches the mean of the limiting distribution. Under $\psi^{(1)}$, this model admits an explicit sequential change of variable map given by the recursion

$$\begin{aligned} v_1 =& -\frac{\lambda}{1-\phi} \log(1 - \Phi(z_1)) \\ v_i =& \phi v_{i-1} - \lambda \log(1 - \Phi(z_i)),\ i = 2, \ldots, n \end{aligned}$$

and it is easy to verify that the gradient and Hessian of $g^*_{(1)}$ are on the form

$$\begin{aligned} \left[\nabla g^*_{(1)}(\mathbf{z})\right]_i =& z_i + O(\lambda) \\ \left[\nabla^2 g^*_{(1)}(\mathbf{z})\right]_{ij} =& \begin{cases} 1 + O(\lambda) & \text{if } i = j \\ O(\lambda^2) & \text{otherwise} \end{cases} \end{aligned}$$

for fixed values of ϕ, γ and small values of λ. Moreover all the higher derivatives are at least $O(\lambda)$ as one would expect. This suggests that for small λ, the transition standard deviation, a high degree of accuracy could be expected when using Laplace methods. Moreover, the transition to a degenerate volatility

process, i.e. X_t is i.i.d. $N(0, \gamma)$, is smooth both theoretically and numerically.

2.6 Derivative calculations and implementation issues

For the proposed methodology to be applicable, it is essential that the first and second order derivative arrays of the negative modified log-integrand $g^*(\mathbf{z})$ can be evaluated efficiently. The sparsity pattern in the Hessian matrix seen in the classical treatment of the log-normal SV model (1 - 4) is in general lost with the introduction of the change of variable map. Still, we can find efficient methods for calculating derivatives using a combination of forward and backward mode algorithmic differentiation (AD) techniques (Griewank, 2000).

In the current work, the derivative calculations are done in two phases based on the local preaccumulation principle in Griewank (2000). First, we calculate g^* forward in time, and concurrently, calculate and store the first and second order partial derivatives of M_i and ψ_i. Since the dependence structure is completely given by (10), backward accumulation of the gradient and Hessian can be done secondly in only $\mathcal{O}(n)$ and $\mathcal{O}(n^2)$ operations respectively, with only a small multiple of n extra memory required. Note that the computational cost of the determinant evaluation in (7) is $\mathcal{O}(n^3)$, and hence the cost of the derivative calculation will be neglectible when n gets large. We use the AD tool Tapenade (Hascoët and Pascual, 2004) to generate multi-directional forward mode code for the derivatives of ψ_i and M_i, whereas the backward accumulation schemes have been hand coded. This balance between backward and forward AD yields a fast and memory-efficient code, which is easily adapted to new models.

All non-trivial one-dimensional integrals are evaluated using the integrals of Chebyshev interpolating polynomials (Press et al., 1992) to obtain fast repeated evaluation of the approximate CDFs and smooth code. The optimization problem in (7) is solved using a two-step procedure following Skaug and Fournier

(2006). First a BFGS quasi-Newton solver (Nocedal and Wright, 1999), then a full Newton solver (Nocedal and Wright, 1999). The BFGS solver provide stability over a large part of the parameter space, whereas the Newton solver lets us solve the inner problem to machine precision to keep the simulated likelihood function smooth. Common random numbers in the importance samplers is another remedy applied for keeping the likelihood approximation smooth. For each of the i.i.d. standard Gaussian vectors used to simulate from the LAD, we use four antithetics to balance for location and scale (Durbin and Koopman, 2001).

3 Application to real data

To illustrate the proposed methodology, we fit two models to the classic data set of log-returns of the Dollar/Pound exchange rate from 01/10/1981 to 28/06/1985 (Harvey et al., 1994). The data consist of a total of $n = 945$ log-return observations.

The first model is the log-normal SV model (1 - 4), but with the latent process taken to be V_t in (2), i.e. the process with log-normal marginals. This example is included for two reasons. First, it shows that the current methodology gives accurate results compared to highly model-specific methods in previous work. Secondly, it provides a reference for comparison between the two models. The second example considers a simplified semi-discretized version of the Heston model (Heston, 1993) where the volatility process follows a square-root diffusion. Simulated likelihood analysis under such a volatility process is, to our knowledge, new.

In both examples, we use $m = 128$ samples from the LAD, based on 32 i.i.d. standard Gaussian vectors.

3.1 The log-normal SV model

The log-normal SV model fitted to the classic Dollar/Pound data have been thoroughly analyzed by Harvey et al. (1994); Shepard and Pitt (1997); Durbin and Koopman (2001); Liesenfeld and Richard (2006) using simulated likelihood and MCMC methods. Common for the simulated likelihood approaches is that the latent process is taken to be the Gaussian AR(1) process (3), thus yielding a near-Gaussian integrand suitable for straight forward LIS. Here, we disregard the Gaussian process U_t, and apply the methodology directly to V_t, so that our $\psi^{(1)}$ map is equivalent to the approach taken by previous authors.

The setup is as follows. We first estimate the parameters for the Dollar/Pound data 100 times with different seeds in the importance sampler. This is done to assess the convergence of the importance samplers for the chosen value of m. For a single simulation (seed) in the $\psi^{(2)}$ formulation, the function minimizer failed to meet the convergence criterion, and this simulation replica was ignored. We take the mean of the parameter estimates (column 1 in Table 1) to be our "best estimates". Then we conduct a parametric bootstrap experiment using 100 simulated data sets each of length $n = 945$, with the mean parameter estimates as "true" parameters. Resulting estimates of bias, variance and correlation are summarized in Table 1.

The most prominent feature of this analysis is that the Monte Carlo standard errors for $\psi^{(2)}$ are reduced by a factor ≈ 4 relative to those of $\psi^{(1)}$ using the same number of importance samples. This supports the hypothesis that $\psi^{(2)}$ takes the modified negative log-integrand (11) closer to a quadratic form. The parameter estimates are very much in accordance with those presented in previous work, e.g. $(\phi, \sigma, \sigma_X) = (0.9731,\ 0.1726,\ 0.6338)$ in Durbin and Koopman (2001). The parametric bootstrap based standard errors and covariance matrices are also much in accordance with those reported in the previous work.

	Monte Carlo		Parametric bootstrap				
	MC Mean	MC S.E.	Bias	S.E.	Correlations		
					ϕ	σ	σ_X
			$\psi^{(1)}$				
ϕ	0.9742	0.0012	0.0095	0.0176	1	-0.5975	0.0403
σ	0.1709	0.0041	-0.0101	0.0344		1	-0.1279
σ_X	0.6317	0.0014	-0.0061	0.0709			1
l	-918.662	0.2554					
			$\psi^{(2)}$				
ϕ	0.9741	0.0003	0.0095	0.0177	1	-0.5963	0.0391
σ	0.1715	0.0007	-0.0105	0.0344		1	-0.1263
σ_X	0.6315	0.0003	-0.0061	0.0708			1
l	-918.648	0.0657					

Table 1: Parameter estimates for the log-normal SV model fit to the Dollar/Pound data. The first two columns show the mean and standard deviation calculated across the maximum likelihood estimations using different seeds in the importance sampler. l denotes the negative log-likelihood approximation. The remaining columns show the results of the parametric bootstrap experiment.

3.2 A semi-discrete version of the Heston model

The Heston model (Heston, 1993) in a fairly general form may be written as the stochastic differential equation (Aït-Sahalia and Kimmel, 2007)

$$d\begin{bmatrix} s_t \\ V_t \end{bmatrix} = \begin{bmatrix} a + bV_t \\ \beta(\alpha - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma\sqrt{V_t} \end{bmatrix} d\begin{bmatrix} B_t^1 \\ B_t^2 \end{bmatrix}, \quad (13)$$

where s_t is the log-price and (B_t^1, B_t^2) are independent standard Brownian motions. To be consistent with log-normal SV model we assume that observations are on a daily time scale, and simplify the model (to keep the exposition clear without too many technicalities) using Euler discreteization with time-step 1 on the price process and set $a = b = \rho = 0$. Then

$$s_{t+1} = s_t + \sqrt{V_t}\eta_t \implies X_t = s_{t+1} - s_t = \sqrt{V_t}\eta_t$$

where η_t are i.i.d. $N(0,1)$ shocks, X_t is the process of log-returns and V_t as in (13). These simplifications may be interpreted as observing the continuous-time volatility process with multiplicative noise at a collection of discrete times $t = 1, \ldots, n$.

The volatility process is strictly positive when $2\alpha\beta > \sigma^2$ and for $\beta > 0$, the process is stationary and has the property of mean reversion towards the long run mean α. The TPD can be shown to be a scaled non-central χ^2 (Cox et al., 1985) and the marginal distribution is a gamma. See Cox et al. (1985) for a complete characterization of these properties.

In all of the computations, we use a re-normalized saddlepoint approximation (Butler, 2007) to the non-central χ^2 density to obtain fast, stable and smooth evaluation over the whole parameter space. The setup is as in the previous example with 100 maximum likelihood estimations of the Dollar/Pound data set using different seeds and 100 parametric bootstrap estimations. In the $\psi^{(2)}$ computations, two runs for both Monte Carlo and parametric bootstrap failed to converge, and were consequently ignored.

The results are given in Table 2, and again we see that an approximate four-fold decrease in Monte Carlo error when switching from $\psi^{(1)}$ to $\psi^{(2)}$.

3.3 Comparison and diagnostics

Since the price return process in both models is linked to the volatility process in the same manner, we may directly compare the estimated volatility processes. A brief numeric summary of the volatility processes and resulting properties of the price return data are given in Table 3. We see that the volatility process of the semi-discrete Heston model is estimated to have a smaller marginal standard deviation and higher temporal correlation than the log-normal counterpart. This is further reflected in the empirical Bayes smoothed volatilities presented

	Monte Carlo		Parametric bootstrap				
	MC Mean	MC S.E.	Bias	S.E.	Correlations		
					α	β	σ
			$\psi^{(1)}$				
α	0.5425	0.0225	-0.0328	0.1154	1	-0.2206	0.0615
β	0.0194	0.0046	-0.0060	0.0120		1	0.7525
σ	0.0959	0.0079	-0.0043	0.0209			1
l	-920.322	1.0006					
			$\psi^{(2)}$				
α	0.5376	0.0049	-0.0408	0.1197	1	-0.2667	-0.0119
β	0.0200	0.0011	-0.0064	0.0126		1	0.7674
σ	0.0991	0.0025	-0.0055	0.0230			1
l	-920.148	0.2362					

Table 2: Parameter estimates and parametric bootstrap summary for the semi-discrete Heston model. See the caption of Table 1 for details.

Model	$E[V_t]$	$Std[V_t]$	$Skewness[V_t]$	$Corr(V_t, V_{t+1})$	$Std[X_t]$	$Kurtosis[X_t]$
Log-normal	0.5317	0.4689	8.1916	0.9662	0.7292	2.3327
Semi-discrete Heston	0.5376	0.3633	9.0309	0.9802	0.7332	1.3701
Data					0.7111	4.8619

Table 3: Properties of the two volatility process and the resulting log-returns, for the Dollar/Pound data. Moments are evaluated under a stationary assumption on $\{V_t\}$ and for parameter estimates based on $\psi^{(2)}$.

in Figure 2. It is immediate that the AIC based on the estimated likelihood values suggest that the log-normal SV model is the better model for these data as the two models have the same number of parameters. Discrepancy between the estimated and empirical standard deviations and kurtosis of the observed data presented in Table 3 further supports this.

4 Discussion

In the current paper, we have proposed a general methodology for casting non-Gaussian volatility problems into a form suitable for LIS-approximation of the likelihood function. The results we have presented suggest that models with a highly non-Gaussian volatility process can be handled in a similar way.

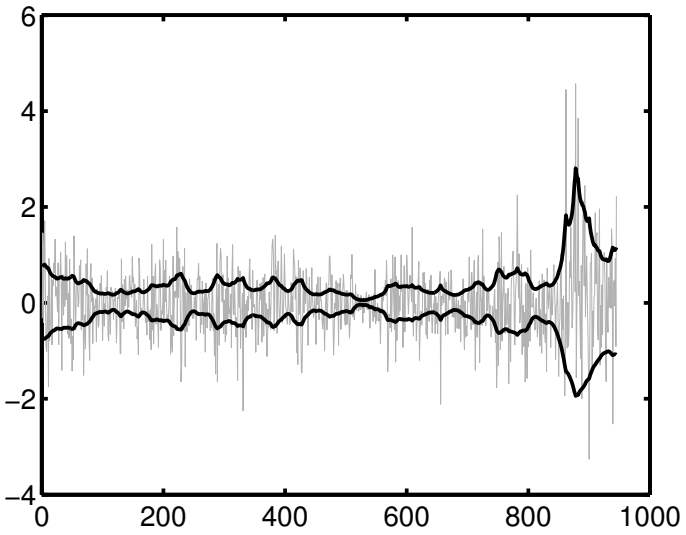

Figure 2: Empirical Bayes estimates of the volatility process based on the complete Dollar/Pound dataset under $\psi^{(2)}$. The upper (positive) estimate is from the log-normal SV model. The lower (negative) estimate is from the semi-discrete Heston model and has been negated to make the plot clearer. The gray curve is the actual data.

We have gone to some length to keep the proposed framework fairly general, so there is clearly scope for tuning for specific models. One such example, given that the specific volatility process limiting density has the same support as the transition density, would be to take the F_is to be the CDFs of $V_i|X_i$. This formulation has the fortunate property of preserving the tri-diagonal form of the Hessian, but the performance of the LIS is dependent on the mixing properties of the volatility process.

For a specific model, it is natural to ask which of the sequential change of variable maps is preferable from a computational perspective. Judging from the two example models considered here, one need about a 16-fold increase in the number of importance samples to obtain the same precision for $\psi^{(1)}$ as for that of $\psi^{(2)}$ (column 2 in Tables 1 and 2). For models that have an explicit $\psi^{(1)}$, this may still be preferable, as most of the computational burden in the current implementation lies in the evaluation and inversion of the F_is. If the sequential change of variable map cannot be made explicit, one is almost always certain to be better off with $\psi^{(2)}$.

The idea of changing the integration variables is not new (Mackay, 1998). Still, it is an area that has received very little attention in the latent variable maximum likelihood literature, considering its huge potential.

The examples presented in this work do by no means deplete the potential of these methods. There are no immediate reasons why models displaying asymmetries and other real-world properties of financial time series should not work, as long as the corresponding TPDs may be computed efficiently. In the context of continuous-time models, the TPD-expansions of Aït-Sahalia and Kimmel (2007) may prove valuable, but this is still a question that will require further research.

References

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

Barndorff-Nielsen, O. E. and D. R. Cox (1989). *Asymptotic techniques for use in statistics*. Chapman & Hall.

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

Butler, R. (2007). *Saddlepoint Approximations with Applications*. Cambridge University Press.

Carlin, B. P. and T. A. Louis (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall.

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.

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

Danielsson, J. and J. F. Richard (1993). Accelerated gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics 8*, 153–173.

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

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. (2007). Sv mixture models with application to s&p 500 index returns. *Journal of Financial Economics 85*, 822–856.

Griewank, A. (2000). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia.

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

Hascoët, L. and V. Pascual (2004). Tapenade 2.1 user's guide. Technical Report 0300, INRIA.

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

Jones, C. S. (2003). The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics 116*(1-2), 181–224.

Kuk, A. Y. C. (1999). Laplace importance sampling for generalized linear mixed models. *Journal of Statistical Computation and Simulation 63*, 143–158.

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

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

Mackay, D. (1998). Choice of basis for Laplace approximation. *Machine learning 33*, 77–86.

Nelson, D. B. (1990). Arch models as diffusion approximations. *Journal of Econometrics 45*(1-2), 7–38.

Nielsen, B. and N. Shephard (2003). Likelihood analysis of a first-order autoregressive model with exponential innovations. *Journal of Time Series Analysis 24*, 337–344.

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

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics 127*(2), 425–449.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical recipes in fortran* (second ed.). Cambridge university press.

Richard, J.-F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics 127*(2), 1385–1411.

Sandmann, G. and S. Koopman (1998). Maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics 63*, 289–306.

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

Skaug, H. and D. Fournier (2006). Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. *Computational Statistics & Data Analysis 56*, 699–709.

Skaug, H. J. (2002). Automatic differentiation to facilitate maximum likelihood

estimation innonlinear random effects models. *Journal of Computational and Graphical Statistics 11*, 458–470.

Sydsæter, K., A. Strøm, and P. Berck (1999). *Economists' Mathmatical Manual* (3 ed.). Springer.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes - a study of the daily sugar prices 1961-75. In O. D. Anderson (Ed.), *Time Series Analysis: Theory and Practice*, Number 1. North-Holland, Amsterdam.