10-2010

Estimating the GARCH Diffusion: Simulated Maximum Likelihood in Continuous Time

Tore Selland KLEPPE
University of Bergen

Jun YU
Singapore Management University, yujun@smu.edu.sg

Hans J. SKAUG
University of Bergen

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Estimating the GARCH Diffusion: Simulated Maximum Likelihood in Continuous Time

Tore Selland Kleppe, Jun Yu & Hans J. Skaug

October 2010

Paper No. 13-2010

Estimating the GARCH Diffusion: Simulated Maximum Likelihood in Continuous Time*

Tore Selland Kleppe
University of Bergen

Jun Yu
Singapore Management University

Hans J. Skaug
University of Bergen

January 5, 2010

Abstract

A new algorithm is developed to provide a simulated maximum likelihood estimation of the GARCH diffusion model of Nelson (1990) based on return data only. The method combines two accurate approximation procedures, namely, the polynomial expansion of Aït-Sahalia (2008) to approximate the transition probability density of return and volatility, and the Efficient Importance Sampler (EIS) of Richard and Zhang (2007) to integrate out the volatility. The first and second order terms in the polynomial expansion are used to generate a base-line importance density for an EIS algorithm. The higher order terms are included when evaluating the importance weights. Monte Carlo experiments show that the new method works well and the discretization error is well controlled by the polynomial expansion. In the empirical application, we fit the GARCH diffusion to equity data, perform diagnostics on the model fit, and test the finiteness of the importance weights.

JEL classification: C11, C15, G12
Keywords: Efficient importance sampling; GARCH diffusion model; Simulated Maximum likelihood; Stochastic volatility

1 Introduction

Inference for stochastic volatility (SV) models using simulated maximum likelihood (SML) estimation has attracted extensive attention in recent years. Important works include Danielsson 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.

Richard (1993), Danielsson (1994), Shephard and Pitt (1997), Sandmann and Koopman (1998), Liesenfeld and Richard (2003, 2006), Durham (2006, 2007), and Richard and Zhang (2007). Yu (2010) reviewed various SML algorithms proposed in the literature. With few exceptions,[1] these studies deal with the discrete-time log-normal SV model of Taylor (1982) and its extensions. On the other hand, in the theoretical finance literature, much attention has been paid to continuous-time SV models (see e.g. Hull and White (1987); Stein and Stein (1991); Heston (1993); Lewis (2000); Aït-Sahalia and Kimmel (2007)) in the form of diffusion models. One reason for favoring continuous-time models is that they allow for a rich and convenient option pricing theory. For instance, the no-arbitrage-condition is characterized by a Martingale-measure, and a large class of options may be priced by solving the partial differential equations corresponding to conditional expectations of given functionals under this measure.

The most well known continuous time SV model is Heston's model (Heston, 1993). Much of its popularity is due to the fact that a nearly-closed-form expression for European option prices is available for this model. As a result, a number of parameter estimation procedures have been proposed for this specification based on return data only, including generalized method of moments (GMM) (Chacko and Viceira, 2003), Bayesian Markov Chain Monte Carlo (MCMC) (Eraker et al., 2003; Jones, 2003), efficient method of moments (Chernov and Ghysels, 2000), SML (Durham, 2006), and methods based on the empirical characteristic function (ECF) (e.g. Singleton (2001)). However, several studies have found strong empirical evidence against Heston's specification; see, e.g. Andersen et al. (2002), Jones (2003), and Aït-Sahalia and Kimmel (2007).[2]

The main contribution of this paper is to develop a SML procedure to estimate the GARCH diffusion model of Nelson (1990) based on return data only. There are several important reasons why we choose to estimate the GARCH diffusion. First, although the GARCH diffusion model is not the most widely used continuous time SV model in the option pricing literature, its discrete time ARCH model, GARCH(1,1), has been one of the most popular specifications in the discrete time literature

[1] For example, the Euler-Maruyama-based discretized Heston's model in Durham (2006) and the inverted gamma model of Richard and Zhang (2007) belong to these exceptions.

[2] While we focus on the estimation of the physical measure, there are studies in the literature on estimating the Heston model using options data only to learn about the risk neutral measure (Bakshi et al., 1997) and using both options and return data to learn about the physical and the risk neutral measures Aït-Sahalia and Kimmel (2007).

and received the most empirical applications within the ARCH family. As a consequence, not surprisingly, some recent studies have found that the GARCH diffusion model is able to capture the dynamics of stock prices better than Heston's model; see, for example, Jones (2003) where Bayesian MCMC was used for the empirical analysis. Moreover, the GARCH diffusion model offers highly accurate approximations to real option prices (Barone-Adesi et al., 2005).

Second, contrary to a common belief, the maximum likelihood (ML) estimates obtained under the GARCH model are not asymptotically equivalent to those obtained under the GARCH diffusion. As shown in Nelson (1990), GARCH(1,1) converges weakly to the GARCH diffusion process. This property implies that GARCH diffusion will share a similar empirical success to the discrete time GARCH(1,1) model. Although it is attempting to suggest the idea of estimating the GARCH diffusion using a discrete time GARCH(1,1) model, as done in several studies (e.g. Lewis (2000), App. 1.1 and Javaheri (2005)), unfortunately, this suggestion is not theoretically correct because the two models are not asymptotically equivalent in terms of Le Cam's deficiency distance (Wang, 2002).

Third, to the best of our knowledge, there is no significant development of the estimation technique for the GARCH diffusion based on return data only. This is a challenging task for at least three reasons. First, unlike the Heston model, the GARCH diffusion does not have a closed form expression for the characteristic function, making the ECF based approach infeasible.[3] This feature is obviously shared by model specifications outside of the affine family (Duffie et al., 2000). Second, unlike the discrete time SV models, the GARCH diffusion does not have a closed form expression for the joint transition probability density (TPD) of the return and the volatility. The lack of analytical expression for the joint TPD is generally true for continuous time models. Third, even if the joint TPD is available, ML is still not straightforward because the volatility is an unobservable state variable and has to be integrated out from the joint density. It is well known that such an integration is of high dimension and numerical techniques are required (see e.g. Danielsson (1994); Shephard and Pitt (1997); Sandmann and Koopman (1998)). To develop the ML procedure for the GARCH diffusion, obviously the last two difficulties cannot be circumvented.

[3] However, as shown in Meddahi (2002), the moments of GARCH diffusion is available and hence a GMM procedure may be developed.

The algorithm developed in the present paper combines two accurate approximation procedures, namely, the polynomial expansions of Aït-Sahalia (2008) to approximate the transition density of the return and the volatility, and the Efficient Importance Sampler (EIS) of Richard and Zhang (2007) to integrate out the volatility. The first and second order terms in the polynomial expansion are used to generate a base-line importance density for an EIS algorithm.

There are two state-of-the-art techniques to approximate the TPD of the return and the volatility of a continuous time model – in-fill simulations and series expansions. The in-fill simulation approach was proposed and refined in Pedersen (1995), Elerian et al. (2001) and Durham and Gallant (2002) in the case of discretely observed processes. For the series expansions, seminal contributions include Aït-Sahalia (2002b, 2008) and Aït-Sahalia and Yu (2006). The expansions proposed by Aït-Sahalia (2002b, 2008) are based on Hermite polynomials and Taylor-like polynomials respectively while Aït-Sahalia and Yu (2006) propose the saddlepoint approximation. We shall follow the series expansion approach and apply a bi-variate polynomial expansion for irreducible diffusions (from now on polynomial expansion) as described in Aït-Sahalia (2008) to approximate to arbitrary precision the transition TPD. These expansions are of closed form, are highly accurate and allow for fast repeated evaluation. Aït-Sahalia (2002a) compared the improved in-fill simulation method of Durham and Gallant (2002) to the Hermite expansions for the TPD and found that the Hermite expansions can obtain more accurate evaluations in much shorter CPU time. It will be made clear soon that the computational cost at this stage is essential because the evaluation of the TPD is used in combination with the EIS methods.

Our work is related to the earlier work by Aït-Sahalia and Kimmel (2007) where the polynomial expansions were used to provide ML estimation of continuous time SV models and Jones (2003) where a Bayesian MCMC method was used to estimate in-filled Euler-Maruyama (EM) discretized continuous time SV models. However, in Aït-Sahalia and Kimmel (2007) and Jones (2003) the estimation was performed based on the assumption that both the return and the volatility are observed. In their empirical applications, the volatility was assumed to be the same as the implied volatility. Our approach does not require the volatility be observed and we integrate out the latent volatility using the EIS algorithm. To account for the non-Gaussianity in the polynomial expansion,

some further adjustments must be made to the existing EIS algorithm.

Our work is also related to Eraker (2001) and Durham and Gallant (2002) where certain continuous time SV models were estimated without using the EM discretization. Their techniques are based on in-fill simulation approach while our technique is based on the series expansion. Moreover, Eraker (2001) uses a Bayesian MCMC method whereas ours is a ML approach.

Although we only estimate the GARCH diffusion model in the paper, we need to point out that our method is not limited to any particular continuous time SV model. Neither the polynomial expansion nor the EIS require a linear function form in the drift function or the diffusion function.

The rest of the paper is organized as follows. Section 2 reviews the GARCH-diffusion model and introduces the SML algorithm for approximating the log likelihood function. In Section 3, we perform a Monte Carlo study to check the statistical performance of the proposed SML. In Section 4, SML is applied to a real data set, and some diagnostic tests are performed. Finally, Section 5 provides some discussion.

2 Model and Methodology

Let S_t denote the log-price of some asset, and V_t the volatility of this asset. Then the GARCH diffusion is given as the solution to the Itô stochastic differential equation (SDE)

$$d\begin{bmatrix} S_t \\ V_t \end{bmatrix} = \begin{bmatrix} a \\ \alpha + \beta V_t \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma V_t \end{bmatrix} \begin{bmatrix} dB_{t,1} \\ dB_{t,2} \end{bmatrix},$$

where $B_{t,1}$ and $B_{t,2}$ denotes a pair of independent canonical Brownian motions. Here, $\theta = [\alpha, \beta, \sigma, \rho, a]$ are the parameters to be determined. Provided that $\beta < 0$, the volatility process V_t is mean reverting with long run mean equal to $-\alpha/\beta$. The stationary distribution is the inverse Gamma with shape parameter $\tilde{\alpha} = 1 - 2\beta/\sigma^2$ and scale parameter $\tilde{\beta} = 2\alpha/\sigma^2$ (see e.g. Nelson (1990) and Barone-Adesi et al. (2005)).

For convenience, we follow Aït-Sahalia (2002b) or Durham and Gallant (2002) and apply the variance stabilizing transformation to the volatility. More precisely, we define $Z_t = \log(V_t)$ and

apply Ito's lemma to find the joint dynamics of S_t and Z_t to be

$$d\begin{bmatrix} S_t \\ Z_t \end{bmatrix} = \begin{bmatrix} a \\ (\beta - \frac{1}{2}\sigma^2) + \alpha\exp(-Z_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)}\exp\left(\frac{1}{2}Z_t\right) & \rho\exp\left(\frac{1}{2}Z_t\right) \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} dB_{t,1} \\ dB_{t,2} \end{bmatrix}. \tag{1}$$

Clearly, the resulting latent process Z_t is a non-linear Ornstein-Uhlenbec process.

In the rest of this paper, we assume that observations of the log-price process are only available at discrete times, and that the volatility is unobserved. Namely, only discrete observations on return are available. More precisely, we assume that we are given $n+1$ regularly spaced observations $\mathbf{s} = [s_0, s_1, s_2, \ldots, s_n] = [S_0, S_\Delta, S_{2\Delta}, \ldots, S_{n\Delta}]$, but the equal spacing assumption can easily be relaxed. Correspondingly, we use the notation $\mathbf{z} = [z_0, z_1, \ldots, z_n]$ for (unobserved) discretely sampled log-volatilities at times corresponding to those of $\mathbf{s}$. In the following, we suppress all dependencies on the parameter vector θ and the time step Δ to keep the notation simple.

2.1 TPDs, joint densities and their approximations

Let $p(s_i, z_i|s_{i-1}, z_{i-1})$ denote the TPD of solution process of the SDE (1). Due to the Markov property (Øksendal, 2003) of the solution process of (1), the joint density of $\mathbf{s}$ and $\mathbf{z}$ is given as

$$p(\mathbf{s}, \mathbf{z}) = p(s_0, z_0) \prod_{i=1}^{n} p_i(s_i, z_i|s_{i-1}, z_{i-1}). \tag{2}$$

Given the special structure of the model (1), the log-price enters the TPD only through the log-return denoted by $x_i := s_i - s_{i-1}$. We shall use the short hand notation $p_i(s_i, z_i|s_{i-1}, z_{i-1}) = p_i(x_i, z_i|z_{i-1})$ and let $\mathbf{x}$ denote the n-vector of log-returns $[x_1, \ldots, x_n] = [s_1 - s_0, \ldots, s_n - s_{n-1}]$. This, in turn, implies that we need only to specify an initial density for the stationary Z_t at $t = 0$, and the joint density of $(\mathbf{x}, \mathbf{z})$ takes the form

$$p(\mathbf{x}, \mathbf{z}) = p_0(z_0) \prod_{i=1}^{n} p_i(x_i, z_i|z_{i-1}). \tag{3}$$

Unfortunately, the TPD of the GARCH-diffusion is not known in closed form, and in general we need to approximate the joint density $p(\mathbf{x}, \mathbf{z})$. This is done by simply substituting each $p_i,\ i = 0, \ldots, n$ with generic approximations $\bar{p}_i$.

Two classes of approximate TPDs are employed for $\bar{p}_i$, namely EM-TPDs ($\bar{p}_i^{(E)}$) and the polynomial expansions for non-reducible diffusions of order K given in Aït-Sahalia (2008), denoted by $\bar{p}_i^{(K)}$. Though the EM-TPDs are conceptually simple and have some good properties when constructing the importance sampler (Kleppe et al., 2009), their fixed accuracy for fixed Δ may lead to an unacceptable bias in the resulting approximate integrated ML procedure (relative to the use of exact TPDs). Though more cumbersome to derive and having somewhat higher computational cost, the polynomial expansions of Aït-Sahalia are attractive in that they have closed form but still adjustable accuracy for varying order K. This enables us to study the error resulting from the EM-TPDs by considering a sequence of Aït-Sahalia expansions of increasing order.

The EM-TPDs $\bar{p}_i(x_i, z_i|z_{i-1})$ are bivariate Gaussian densities characterized by the mean vector and covariance matrix

$$\begin{bmatrix} \Delta a \\ z_{i-1} + \Delta((\beta - \frac{1}{2}\sigma^2) + \alpha \exp(-z_{i-1})) \end{bmatrix} \text{ and } \Delta \begin{bmatrix} \exp(z_{i-1}) & \sigma\rho \exp\left(\frac{1}{2} z_{i-1}\right) \\ \sigma\rho \exp\left(\frac{1}{2} z_{i-1}\right) & \sigma^2 \end{bmatrix} \tag{4}$$

respectively. In the special case of the GARCH diffusion model with state variables (S_t, Z_t), the Aït-Sahalia expansions of order K have the form

$$\log \bar{p}_i^{(K)}(x_i, z_i|z_{i-1}) = -\log(2\pi\Delta) - \frac{1}{2}\left(z_i + \log(\sigma^2(1-\rho^2))\right) + \frac{C_{-1}(x_i, z_i, z_{i-1})}{\Delta} + \sum_{k=0}^{K} C_k(x_i, z_i, z_{i-1}) \frac{\Delta^k}{k!}. \tag{5}$$

Clearly, the expansion has the interpretation as a functional power series in Δ (plus some additional terms). The form of the coefficients C_k are found by solving both the Forward- and Backward Kolmogorov partial differential equations to the appropriate orders in Δ using the algorithms outlined in Aït-Sahalia (2008). The actual expressions for C_k are in general complicated, and we obtained

these using Maple. Their exact specification is available upon request in computer form from the first author. It is worth noticing that the polynomial expansions are not proper densities as they do not integrate to exactly one. However, in our experience the expansions are very accurate for the GARCH diffusion model so that re-normalization is unnecessary.

The initial density p_0 does have a closed form, namely the density of the logarithm of inverse Gamma variate, but we take $\bar{p}_0$ to be the Gaussian Laplace approximation to p_0, i.e. the Gaussian density with the same mode and same second derivative as p_0 at the common mode. The mean and standard deviation characterizing this Gaussian approximation are given as

$$\mu_{\mathbf{0}_0} = -\log\left(\frac{\sigma^2 - 2\beta}{2\alpha}\right), \tag{6}$$

$$\Sigma_{\mathbf{0}_0} = \frac{\sigma^2}{\sigma^2 - 2\beta}. \tag{7}$$

This simplification is mainly done for convenience when constructing the importance sampler, and the errors committed are asymptotically small when n increases.

2.2 Simulated maximum likelihood for the GARCH diffusion model

As explained in the introduction, the second obstacle for the likelihood analysis in the GARCH-diffusion is that the volatility is unobserved, and needs to be integrated out of the joint likelihood (3). In this work, we adapt the EIS procedure outlined in Kleppe et al. (2009) to work with the approximate TPDs described above. The EIS is chosen as it does not rely on a global near-Gaussian assumption on $p(\mathbf{z}|\mathbf{x})$ which is required by the Laplace importance sampler (Shephard and Pitt, 1997). We first review the EIS procedure originally proposed by Liesenfeld and Richard (2003) and further explained in Richard and Zhang (2007).

2.2.1 An efficient importance sampling procedure

The idea behind importance sampling for calculating the marginalization integral is choose an auxiliary importance density $m(\mathbf{z})$ so that

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{z}) d\mathbf{z} = \int \frac{p(\mathbf{x}, \mathbf{z})}{m(\mathbf{z})} m(\mathbf{z}) d\mathbf{z}. \tag{8}$$

For $M \to \infty$, (8) can be approximated by the Monte Carlo estimate

$$p(\mathbf{x}) \approx \tilde{l}(\theta|\mathbf{x}) = \frac{1}{M} \sum_{j=1}^{M} \frac{p(\mathbf{x}, \tilde{\mathbf{z}}^{(j)})}{m(\tilde{\mathbf{z}}^{(j)})}, \tag{9}$$

where $\tilde{\mathbf{z}}^{(j)} \sim m(\mathbf{z})$ for $j = 1, \ldots, M$. The importance weights $p(\mathbf{x}, \tilde{\mathbf{z}}^{(j)})/m(\tilde{\mathbf{z}}^{(j)})$ are denoted by $w(\tilde{\mathbf{z}}^{(j)})$. We shall refer to the approximate ML estimator,

$$\hat{\theta} = \arg\max_{\theta} \log \tilde{l}(\theta|\mathbf{x}), \tag{10}$$

as the SML estimator.

The EIS algorithm of Richard and Zhang (2007) provides a method for choosing an optimal importance density $m(\mathbf{z}) = m(\mathbf{z}|\hat{\mathbf{a}})$ within a prescribed class of auxiliary importance densities indexed by a $n + 1 \times 2$ dimensional parameter $\mathbf{a}$. The optimality is in the sense that the variance of $\tilde{l}(\theta|\mathbf{x})$, for a fixed number of importance draws M, is minimized within the class of admissible values of $\mathbf{a}$. We refer to Richard and Zhang (2007) for a more detailed description of the EIS and its optimality properties, as a full recapture of their work is beyond the scope of this paper.

The EIS algorithm sampler employed here is derived as follows. Firstly, we introduce the base-line importance density $m(\mathbf{z}|\mathbf{0})$ where $\mathbf{a} = \mathbf{0}$ is taken element wise. The base-line importance density is derived from the approximate joint density of $(\mathbf{x}, \mathbf{z})$ and plays an important role as the full auxiliary importance density will be expanded around it. Secondly, we add the flexibility indexed by the parameter $\mathbf{a}$ by parametrically extending the base-line importance density at each dimension within the Gaussian class of densities. Finally, we derive the linear least-squares regressions that

are used to locate the optimal parameter $\mathbf{a}$ in an iterative manner.

2.2.2 The Base-line importance density

In most earlier applications of EIS (see e.g. Liesenfeld and Richard (2003, 2006); Richard and Zhang (2007); Bauwens and Galli (2009)), the base-line importance density $m(\mathbf{z}|\mathbf{0})$ is taken to be the natural sampler, i.e. the marginal density of the latent process $p(\mathbf{z})$. In this work we follow Kleppe et al. (2009) and use the product of Gaussian approximations (exact for EM-TPDs) to the conditional transition density, given the observed return x_i. Clearly, taking $m(\mathbf{z}|\mathbf{0}) = p(\mathbf{z}|\mathbf{x})$ would result in an importance sampler with zero variance. However, conditioning on the whole $\mathbf{x}$ is as hard as the initial problem, and we shall therefore only condition on individual elements x_i, as all the resulting formulae are of closed form. Conditioning on larger portions of $\mathbf{x}$ forward in time would typically result in even smaller variance, but this is not practical as it would involve numerical integration. We refer to Kleppe and Skaug (2009) for more details on constructing importance samplers around products of conditional-on-data transition densities.

The Gaussian approximations to the conditional (on data) TPDs are derived by factoring the approximate z_i-variation of the $\bar{p}_i$ into a constant part, a "Gaussian part" and some residual variation. More precisely, we write

$$\bar{p}(x_i, z_i|z_{i-1}) = A_i(z_{i-1}, x_i)B_i(z_i, z_{i-1}, x_i)R_i(z_i, z_{i-1}, x_i) \tag{11}$$

where $A_i(z_{i-1}, x_i)$ does not depend on z_i,

$$B_i(z_i, z_{i-1}, x_i) = \exp\left(-\frac{(z_i - \mu_{\mathbf{0}_i}(z_{i-1}, x_i))^2}{2\Sigma^2_{\mathbf{0}_i}(z_{i-1}, x_i)}\right) \tag{12}$$

is a Gaussian kernel in z_i and $R(z_i, z_{i-1}, x_i)$ is a slowly varying function that we shall refer to as the residual variation.[4] We will choose $\mu_{\mathbf{0}_i}(z_{i-1}, x_i)$ and $\Sigma_{\mathbf{0}_i}(z_{i-1}, x_i)$ so that as much as possible of the z_i-variation is accounted for in B_i. Consequently, this makes R_i close to constant. Hence A_iB_i has the interpretation of being an un-normalized Gaussian approximation to the conditional

[4] Here the subscript $\mathbf{0}_i$ should be read as the ith row of $\mathbf{a}$ with the elements set to zero.

on x_i transition density of z_i from z_{i-1}.

In particular for the bi-variate Gaussian $\bar{p}_i^{(E)}$, B_i represents the exact shape of the conditional density $\bar{p}_i^{(E)}(z_i|z_{i-1}, x_i)$ and $R_i = 1$. In accordance with Kleppe et al. (2009), the expressions for $\mu_{\mathbf{0}_i}$, $\Sigma_{\mathbf{0}_i}$ and A_i under the EM discretization are given as

$$\mu_{\mathbf{0}_i,(E)}(z_{i-1}, x_i) = z_{i-1} + \Delta((\beta - \frac{1}{2}\sigma^2) + \alpha \exp(-z_{i-1})) + \sigma\rho(x_i - \Delta a)\exp(-\frac{z_{i-1}}{2}), \tag{13}$$

$$\Sigma_{\mathbf{0}_i,(E)}(z_{i-1}, x_i) = \sigma\sqrt{\Delta(1-\rho^2)}, \tag{14}$$

$$A_{i,(E)}(z_{i-1}, x_i) = \frac{1}{\sqrt{2\pi\Sigma^2_{\mathbf{0}_i,(E)}}}, \tag{15}$$

and thus have we fully characterized the factorization (11) of the EM-TPDs.

For the expansions of Aït-Sahalia, C_k, $k = -1, \ldots, K$ are polynomials in x_i and $z_i - z_{i-1}$ of order $2(K-k)$ [5], but the polynomial coefficients generally depend on z_{i-1} in a non-polynomial way. Thus is $\bar{p}_i^{(K)}(z_i|z_{i-1}, x_i)$ not exactly Gaussian and we derive the factors (11) using the following Taylor series argument: Since $\log \bar{p}_i^{(K)}(x_i, z_i|z_{i-1})$ is a polynomial in $z_i \mapsto (z_i - z_{i-1})$, we may rearrange the terms according to their order in $(z_i - z_{i-1})$ rather than Δ. This may be done as the Taylor series in z_i around the old state z_{i-1}:

$$\log \bar{p}_i^{(K)}(x_i, z_i|z_{i-1}) = \sum_{l=0}^{L} D_{l,(K)}(x_i, z_{i-1})(z_i - z_{i-1})^l \tag{16}$$

where $L = 2(K+1)$ is the highest order of the polynomials used. Notice that the approximate log-TPD is a polynomial in z_i, so no additional error is committed by introducing the Taylor series representation. Another rearrangement by keeping all the terms up to the second order gives us

$$\log \bar{p}_i^{(K)}(x_i, z_i|z_{i-1}) = \underbrace{D'_{0,(K)}(x_i, z_{i-1})}_{\log A_{i,(K)}} \underbrace{- \frac{(z_i - \mu_{\mathbf{0}_i,(K)})^2}{2\Sigma^2_{\mathbf{0}_i,(K)}}}_{\log B_{i,(K)}} + \underbrace{\sum_{l=3}^{L} D_{l,(K)}(x_i, z_{i-1})(z_i - z_{i-1})^l}_{\log R_{i,(K)}} \tag{17}$$

[5] We follow Aït-Sahalia and Kimmel (2007)s $2(K-k)$ rather than Aït-Sahalia (2008)s $2(K+1-k)$ on the choice of polynomial order for computational convenience.

where

$$D'_{0,(K)} = D_{0,(K)} + \frac{D^2_{1,(K)}}{4D_{2,(K)}} \tag{18}$$

$$\mu_{\mathbf{0}_i,(K)} = z_{i-1} - \frac{D_{1,(K)}}{2D_{2,(K)}} \tag{19}$$

$$\Sigma_{\mathbf{0}_i,(K)} = \frac{1}{\sqrt{-2D_{2,(K)}}} \tag{20}$$

Notice that this Taylor series argument would have produced the same results as in (13 - 15) if it was applied to the EM-TPDs. However, for the non-Gaussian Aït-Sahalia expansions there is in general no guarantee that this argument would produce a valid Gaussian approximation. Still, due to the fact that we are working with the log-volatility with near-Gaussian conditional-on-data TPDs, the second order series $\log A_i + \log B_i$ provides a precise approximation, and in practice has non-valid Gaussian approximations not been a problem for this model.

For both the EM-TPDs and the polynomial expansions, we use the Gaussian approximation $\bar{p}_0$ to the stationary density. Analogously to the above introduced notation, we write $\bar{p}_0(z_0) = A_0 B_0(z_0)$ where

$$A_0 = \frac{1}{\sqrt{2\pi\Sigma^2_{\mathbf{0}_0}}}, \quad \log B_0(z_0) = -\frac{(z_0 - \mu_{\mathbf{0}_0})^2}{2\Sigma^2_{\mathbf{0}_0}} \tag{21}$$

With this generic notation in place, we may define the locally Gaussian base-line importance density as the product of Gaussian approximations to conditional-on-data TPDs

$$m(\mathbf{z}|\mathbf{0}) = m_0(z_0|\mathbf{0}_0) \prod_{i=1}^{n} m_i(z_i|z_{i-1}, x_i, \mathbf{0}_i) \tag{22}$$

where

$$m_0(z_0|\mathbf{0}_0) = \frac{B_0(z_0)}{\sqrt{2\pi\Sigma^2_{\mathbf{0}_0}}} \text{ and } m_i(z_i|z_{i-1}, x_i, \mathbf{0}_i) = \frac{B_i(z_i, z_{i-1}, x_i)}{\sqrt{2\pi\Sigma^2_{\mathbf{0}_i}}}, \text{ for } i = 1, \ldots, n. \tag{23}$$

Notice in particular that $m(\mathbf{z}|\mathbf{0})$ contain information from the data, and thus should its shape be closer to posterior density of the log-volatility given the data, i.e $p(\mathbf{z}|\mathbf{x})$, than the natural sampler.

2.2.3 The parametrically extended importance density

Let $\{\mathbf{a}_i\}_{i=0}^n$ denote the rows of $\mathbf{a}$, and let $a_{i,1}$ and $a_{i,2}$ denote the elements of $\mathbf{a}_i$. Following the earlier literature on EIS, we extend each element m_i of the (22) within the family of Gaussian densities. In practice, this is done by multiplying the elements with $\psi_i(z_i|\mathbf{a}_i) = \exp(a_{i,1}z_i + a_{i,2}z_i^2)$ and compensating with the appropriate normalization factor. Thus we get the representation

$$m_0(z_i|\mathbf{a}_0) = \frac{B_0(z_i)\exp(a_{0,1}z_i + a_{0,2}z_i^2)}{\chi_0(\mathbf{a}_0)}, \tag{24}$$

$$m_i(z_i|z_{i-1}, x_i, \mathbf{a}_i) = \frac{B_i(z_i|z_{i-1}, x_i)\exp(a_{i,1}z_i + a_{i,2}z_i^2)}{\chi_i(z_{i-1}, x_i, \mathbf{a}_i)}, \tag{25}$$

where

$$\chi_0(\mathbf{a_0}) = \int_{\mathbb{R}} B_0(z_0)\psi_0(z_0|\mathbf{a}_0)dz_0, \tag{26}$$

$$\chi_i(z_{i-1}, x_i, \mathbf{a_i}) = \int_{\mathbb{R}} B_i(z_i|z_{i-1}, x_i)\psi_i(z_i|\mathbf{a}_i)dz_i. \tag{27}$$

The explicit expression for $\log\chi_i$ is given in Appendix A. Simple calculations yields that the m_is are the Gaussian with means and standard deviations given as,

$$\mu_{\mathbf{a}_0} = \frac{\mu_{\mathbf{0}_0} + a_{0,1}\Sigma^2_{\mathbf{0}_0}}{1 - 2a_{0,2}\Sigma^2_{\mathbf{0}_0}}, \qquad \mu_{\mathbf{a}_i}(z_{i-1}, x_i) = \frac{\mu_{\mathbf{0}_i}(z_{i-1}, x_i) + a_{i,1}\Sigma^2_{\mathbf{0}_i}(z_{i-1}, x_i)}{1 - 2a_{i,2}\Sigma^2_{\mathbf{0}_i}(z_{i-1}, x_i)},\ i = 1, \ldots, n, \tag{28}$$

$$\Sigma_{\mathbf{a}_0} = \frac{\Sigma_{\mathbf{0}_0}}{\sqrt{1 - 2a_{0,2}\Sigma^2_{\mathbf{0}_0}}}, \qquad \Sigma_{\mathbf{a}_i}(z_{i-1}, x_i) = \frac{\Sigma_{\mathbf{0}_i}(z_{i-1}, x_i)}{\sqrt{1 - 2a_{i,2}\Sigma^2_{\mathbf{0}_i}(z_{i-1}, x_i)}},\ i = 1, \ldots, n. \tag{29}$$

Inspection of these expressions suggest that $m(\mathbf{z}|\mathbf{a})$ also has a Markov structure (conditionally on $\mathbf{x}$ and $\mathbf{a}$) with Gaussian transition densities, and that sampling from the importance density using (28) and (29) is therefore fast and conceptually simple. For the conditional standard errors to be finite, it is required that $a_{i,2} < 1/(2\Sigma^2_{\mathbf{0}_i})$. In practice, the $\hat{a}_{i,2}$s are typically negative. Heuristically, this is reasonable as $p(\mathbf{z}|\mathbf{x})$ carries more information regarding $\mathbf{z}$ than $m(\mathbf{z}|\mathbf{0})$ does. As $m(\mathbf{z}|\hat{\mathbf{a}})$ may be viewed upon as an (un-normalized) approximation to $p(\mathbf{z}|\mathbf{x})$, it is reasonable that the optimal EIS parameter $\hat{\mathbf{a}}$ will shrink the transition standard deviations by attaining negative $a_{i,2}$ values.

2.2.4 Collecting factors and the EIS regressions

Recall that our aim using the EIS is to minimize the variance of the importance sampler weights $w(\mathbf{z})$ under the importance law on $\mathbf{z}$. The product of Gaussian kernels $\prod_{i=0}^{n} B_i$ cancels in the weights, and thus do not contribute to the weight variance. The remaining factors may be written as (see Appendix A for details)

$$w(\mathbf{z}) = \frac{p(\mathbf{x},\mathbf{z})}{m(\mathbf{z}|\mathbf{a})} = \chi_0(\mathbf{a}_0)A_0 \left[\frac{\chi_1(z_0,x_1,\mathbf{a}_1)A_1(z_0,x_1)}{\psi_0(z_0|\mathbf{a}_0)}\right]\left[\frac{R_n(z_n,z_{n-1},x_n)}{\psi_n(z_n|\mathbf{a}_n)}\right] \times \prod_{i=1}^{n-1}\left[\frac{\chi_{i+1}(z_i,x_{i+1},\mathbf{a}_{i+1})A_{i+1}(z_i,x_{i+1})R_i(z_i,z_{i-1},x_i)}{\psi_i(z_i,\mathbf{a}_i)}\right]. \quad (30)$$

Notice that we have collected the factors so that variation within each bracket is mainly for each $z_i,\ i=0,\ldots,n$. Still, the R_is prevents us from obtaining a perfect factorization in the general case.

Let $\tilde{\mathbf{z}}^{(j)} = [\tilde{z}_0^{(j)},\ldots,\tilde{z}_n^{(j)}],\ j=1,\ldots,M$ be draws from $m(\mathbf{z}|\mathbf{a})$. To minimize the variation of $w(\mathbf{z})$ where $m(\mathbf{z}|\mathbf{a})$ has significant mass, we follow the EIS strategy and introduce the regression models corresponding to the log of each of the bracketed factors in (30) (using that $\log\psi_i = a_{i,1}z_i + a_{i,2}z_i^2$):

$$\log\chi_1(\tilde{z}_0^{(j)},x_1,\mathbf{a}_1) + \log A_1(\tilde{z}_0^{(j)},x_1) = c_0 + a_{0,1}\tilde{z}_0^{(j)} + a_{0,2}(\tilde{z}_0^{(j)})^2 + \varepsilon_0^{(j)} \quad (31)$$

$$\log\chi_{i+1}(\tilde{z}_i^{(j)},x_{i+1},\mathbf{a}_{i+1}) + \log A_{i+1}(\tilde{z}_i^{(j)},x_{i+1}) + \log R_i(\tilde{z}_i^{(j)},\tilde{z}_{i-1}^{(j)},x_i) = c_i + a_{i,1}\tilde{z}_i^{(j)} + a_{i,2}(\tilde{z}_i^{(j)})^2 + \varepsilon_i^{(j)} \quad (32)$$

$$\log R_n(\tilde{z}_n^{(j)},\tilde{z}_{n-1}^{(j)},x_n) = c_n + a_{n,1}\tilde{z}_n^{(j)} + a_{n,2}(\tilde{z}_n^{(j)})^2 + \varepsilon_n^{(j)} \quad (33)$$

where (32) applies for $i=1,\ldots,n-1$. Here $\varepsilon_i^{(j)}$ are residuals and c_i are constant terms that may be included without contributing the variance of w. Notice that the right hand sides in (31 - 33) are linear in $(c_i,a_{i,1},a_{i,2})$, and may thus be estimated using computationally cheap linear least squares routines. The constant factor $\chi_0(\mathbf{a}_0)A_0$ is kept out of regressions, as it does not result in added variance.

As mentioned above, for the EM-TPDs, $\log R_i = 0$ for $i = 1, \ldots, n$, and thus we set $\mathbf{a}_n = (0, 0)$ when using these TPD approximations. The Monte Carlo (MC) variance stems from that the left hand sides of the regressions are non-linear functions in z_i, and thus does the quadratic right hand side models not capture completely the variation. In addition, the $\log R_i$ cannot be split completely into terms that vary only with z_i and z_{i-1}. Still, since the draws from the importance density are highly located, the quadratic models performs very well as will be apparent when we discuss the numerical accuracy of the procedure.

2.2.5 Iterative EIS and implementation

Upon inspection of the EIS regressions above, and the fact that $\tilde{\mathbf{z}}^{(j)}$ themselves depend on $\mathbf{a}$, it is clear that (31 - 33) must be regarded as a fix-point condition satisfied by the optimal EIS parameter $\hat{\mathbf{a}}$. We generate a convergent sequence $\{\mathbf{a}^{(k)}\}_k$ towards $\hat{\mathbf{a}}$ in the following manner:

1. Set $\mathbf{a}^{(0)} = \mathbf{0},\ k = 0$ and let $W_{i,j},\ i = 0, \ldots, n,\ j = 1 \ldots, M$ be $(n+1)M$ independent standard Gaussian variates.

2. Sample $\tilde{\mathbf{z}}^{(j)},\ j = 1, \ldots, M$ from $m(\mathbf{z}|\mathbf{a}^{(k)})$ forward in time using

$$\tilde{z}_i^{(j)} = \mu_{\mathbf{a}_i^{(k)}}(\tilde{z}_{i-1}^{(j)}, x_i) + \Sigma_{\mathbf{a}_i^{(k)}}(\tilde{z}_{i-1}^{(j)}, x_i) w_{i,j} \tag{34}$$

   for $i = 0, \ldots, n,\ j = 1, \ldots, M$.

3. Calculate $\mathbf{a}_i^{(k+1)}$ using the regression models (31 - 33) backwards in time (i.e. $i = n \to 0$) based on $\tilde{\mathbf{z}}^{(j)}$ and $\mathbf{a}_{i+1}^{(k+1)}$ in the $\log \chi_{i+1}$ terms (with obvious alterations for the first regression).

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

We follow Richard and Zhang (2007) and use the same set of standard normal variates for each iteration and for each evaluation of the simulated likelihood to ensure a smooth surface for the log-likelihood function. A total of 8 iterations are performed for each function evaluation. If the simulated likelihood based on the polynomial expansions is computed, we do the 4 first iterations using EM-TPDs, as this is computationally faster and more stable. A small amount of parameter

shrinkage on $a_{i,2}$ (0.001 added to the corresponding diagonal element of the normal equations matrix) is introduced to make the computations more stable, but this small bias does not affect the numerical accuracy to any significant extent.

The algorithm is implemented in `FORTRAN90`. Following Skaug (2002) and Bastani and Guerrieri (2008), we use a algorithmic differentiation (AD) tool to generate code for the exact gradient of the simulated likelihood function. Specifically, we used `Tapenade` (Hascoët and Pascual, 2004) in multidirectional forward mode to complete a gradient in one forward sweep.

Finally, a line searching BFGS-quasi-Newton optimizer (Nocedal and Wright, 1999) is applied to maximize the simulated likelihood function using function values and the AD-generated gradients. None of the estimation replica presented in the next two sections failed to converge.

3 A Monte Carlo Study

To study the statistical properties of the proposed methods on daily data (where $\Delta = 1/252$), we conduct a Monte Carlo experiment. We shall consistently use the acronyms EUL for SML based on EM-TPDs and AS1, AS2 and AS3 for SML based on the polynomial expansions of order $K = 1$, 2 and 3. For the SML, we consistently use $M = 32$ draws both for the MC study and the application to real data.

The setup for the study is as follows. We use two sample sizes – $n = 2022$ (matching the sample size of the real data discussed shortly) and $n = 5000$, corresponding to roughly 8 and 20 years of data respectively. For each of the sample sizes, we simulate 1000 data sets using the EM scheme with time step $\Delta/256$. The resulting data on the Δ-grid should thus have very similar statistical properties to data from a discretely observed GARCH diffusion. Estimators for the relevant parameters are then obtained both with observed volatility and with unobserved volatility.

This simulation study setup is designed to attempt to heuristically disentangle the three main sources of statistical bias involved in this problem.

- The error committed when using the approximate TPDs comparing with applying the exact TPDs. As we employ a sequence of polynomial expansions in addition to the EM discretiza-

method	α	β	σ	ρ	a
True parameters	0.2231	-8.4650	2.7059	-0.3047	0.0955
$n = 2022$ observed log-volatility					
EUL	0.0009	-0.2547	-0.0657	0.0003	-0.0010
	(0.0300)	(1.8389)	(0.0423)	(0.0198)	(0.0482)
AS1	0.0056	-0.3294	0.0014	-0.0009	-0.0028
	(0.0312)	(1.8816)	(0.0435)	(0.0199)	(0.0479)
AS2	0.0075	-0.4404	0.0008	-0.0013	-0.0009
	(0.0314)	(1.8988)	(0.0435)	(0.0199)	(0.0482)
AS3	0.0082	-0.4756	0.0009	-0.0013	-0.0010
	(0.0316)	(1.9095)	(0.0435)	(0.0199)	(0.0482)
$n = 2022$ unobserved log-volatility					
EUL	0.0140	-0.8182	-0.0464	0.0288	0.0073
	(0.0740)	(3.3553)	(0.4282)	(0.0997)	0.0513
AS1	0.0034	-0.2168	-0.0608	0.0003	-0.0038
	(0.0658)	(2.9505)	(0.4065)	(0.1014)	(0.0520)
AS2	0.0149	-0.6748	0.0134	-0.0057	-0.0027
	(0.0716)	(3.1933)	(0.4324)	(0.1014)	(0.0520)
AS3	0.0212	-0.9348	0.0476	-0.0048	-0.0029
	(0.0793)	(3.5065)	(0.4621)	(0.1013)	(0.0520)
$n = 5000$ observed log-volatility					
EUL	-0.0046	0.0525	-0.0650	0.0010	-0.0008
	(0.0185)	(1.1443)	(0.0267)	(0.0127)	(0.0305)
AS1	-0.0001	-0.0127	0.0004	-0.0003	-0.0025
	(0.0192)	(1.1671)	(0.0276)	(0.0128)	(0.0303)
AS2	0.0017	-0.1209	-0.0001	-0.0006	-0.0007
	(0.0193)	(1.1775)	(0.0276)	(0.0128)	(0.0305)
AS3	0.0024	-0.1538	-0.0000	-0.0006	-0.0008
	(0.0194)	(1.1839)	(0.0276)	(0.0128)	(0.0305)
$n = 5000$ unobserved log-volatility					
EUL	-0.0003	-0.1809	-0.0862	0.0290	0.0082
	(0.0399)	(1.8212)	(0.2513)	(0.0616)	(0.0324)
AS1	-0.0082	0.2913	-0.0902	0.0003	-0.0026
	(0.0367)	(1.6593)	(0.2461)	(0.0625)	(0.0330)
AS2	0.0017	-0.1019	-0.0235	-0.0050	-0.0016
	(0.0399)	(1.7933)	(0.2601)	(0.0625)	0.0330
AS3	0.0054	-0.2525	-0.0016	-0.0043	-0.0018
	(0.0422)	(1.8907)	(0.2702)	(0.0625)	(0.0330)

Table 1: Results from the Monte Carlo experiment. All results are taken over 1000 synthetic data sets simulated under the parameters given in the "True parameters" row. Estimated bias (no parenthesis) is calculated as estimated parameters minus the true parameter. Statistical standard errors are given in parenthesizes.

tion, this source of error may be assessed quite rigorously by comparing the different TPD approximations. In particular, a convergence of the estimates obtained for the higher order polynomial expansions suggests that we have sufficient precision in the TPDs.

- The finite sample bias of using the integrated likelihood function. It is well known that ML tends to produce a finite sample bias for the mean reversion parameter for observed diffusion processes. In particular, Phillips and Yu (2009) show that ML estimates tend to be biased towards a faster mean reversion. This claim may be checked in our ML estimates of β when the volatility is observed or unobserved.

- The errors committed by using MC methods, in place of exact integration, to compute the integrated likelihood function. Comparison of estimators based on observed and unobserved volatility gives us some clues as to whether faith should be put into the importance sampler. This source of error will also be addressed in Section 4.1.2, where we test the finiteness of the variance of the importance sampling weights, and thus assess the convergence properties of the proposed importance sampling algorithm.

The parameter estimates obtained under AS2 for the real data discussed in section 4 are used as the "true parameters" throughout the complete experiment. These parameters, along with results from the MC study are summarized in Table 1. The mean computing times for locating the SML estimates ranges from 14 seconds (EUL, $n = 2022$) to 185 seconds (AS3 $n = 5000$) on a typical modern desktop computer.

Table 1 reports the bias and the standard error of each estimate obtained from 1000 replications. From first and third panels where volatility is observed, we see that there is some differences in the estimates obtained using the different TPD approximations. The most striking difference is the underestimation of the σ parameter under the EM-TPDs, whereas these biases are much smaller for the polynomial expansions. This result seems to be consistent with what has been found in Aït-Sahalia (1999). The expected bias towards faster mean reversion is also seen as an underestimation of the β parameter for all the TPDs. The bias gets smaller when the sample size is increased, as one would expect. The estimates obtained using AS2 and AS3 are consistently more similar than the

others, suggesting these approximations represent sufficiently precise approximations to the true TPDs for our needs.

Comparing the estimators obtained with and without observed volatility (i.e. panel 1 with panel 2 and panel 3 with panel 4), we see that the loss of statistical precision is most significant for the σ parameter where a ten-fold increase in the standard error is seen. The parameters governing the linear drift of the volatility, α and β, are subject to about a doubling of the statistical standard errors when the log-volatility is integrated out.

4 Empirical Application to Equity Data

In the empirical application, we employ the Standard & Poor 500 data previously used in Jacquier et al. (1994) and Yu (2005)[6] and later applied in Kleppe et al. (2009). The time-series of log-returns covers the period January 1980 to December 1987 and consists of a total of $n = 2022$ observations. In particular, the data covers the October 1987 crash.

Parameter estimates for the data using the four different estimation procedures are presented in Table 2. For the SML methods, the estimates are calculated as the mean across 100 estimates with different random number seeds in the importance sampler. In addition to parameter estimates and statistical standard errors taken from Table 1, we present standard errors due to the application of Monte Carlo methods for calculating the marginalization integrals. Consistently, the Monte Carlo standard errors are small comparing with statistical standard errors. As additional references for the Monte Carlo standard errors, we may mention that Liesenfeld and Richard (2006) obtains a standard error of 0.11 (0.0120 % of a log-likelihood value of 918) under the log-normal SV model using 30 draws using the EIS and that Durham (2006) obtains a standard error of 2.49 (0.0135 % of a log-likelihood value of 18473) under an EM discretized Heston's model using 1024 draws in a Laplace importance sampler. The corresponding value under the AS2 procedure presented here is 0.0751 (0.0011 % of 6540.9) and must be said to be quite impressing considering the non-linear and non-Gaussian nature of the model and the relatively modest number of draws in the importance

[6] We multiply the data of Yu (2005) with 0.01

method	α	β	σ	ρ	a	log-likelihood
EUL	0.2417	-9.3401	2.8072	-0.2914	0.1042	6541.1
	(0.0740)	(3.3553)	(0.4282)	(0.0997)	(0.0513)	
	[0.0023]	[0.0986]	[0.0127]	[0.0006]	[0.0001]	[0.0849]
AS1	0.2096	-7.9072	2.6283	-0.2989	0.0942	6540.5
	(0.0658)	(2.9505)	(0.4065)	(0.1014)	(0.0520)	
	[0.0012]	[0.0495]	[0.0073]	[0.0006]	[0.0001]	[0.0685]
AS2	0.2231	-8.4650	2.7059	-0.3047	0.0955	6540.9
	(0.0716)	(3.1933)	(0.4324)	(0.1014)	(0.0520)	
	[0.0028]	[0.1180]	[0.0142]	[0.0008]	[0.0001]	[0.0751]
AS3	0.2280	-8.6695	2.7313	-0.3045	0.0954	6541.0
	(0.0793)	(3.5065)	(0.4621)	(0.1013)	(0.0520)	
	[0.0033]	[0.1396]	[0.0167]	[0.0008]	[0.0001]	[0.0782]

Table 2: Parameter estimates and log-likelihood values for the S&P500 data using the four different estimation procedures. The parameter estimates are taken as the mean over 100 replications using different random number seeds in the importance sampler. Statistical standard errors taken from Table 1 with $n = 2022$ and unobserved log-volatility, and are presented in parenthesizes. The estimates of the standard errors due to the EIS MC variation are included in square parenthesizes.

sampler. Much of this improved accuracy (comparing with earlier EIS application such as Liesenfeld and Richard (2003)) comes from the fact that we use the product of conditional-on-data TPDs as our base-line importance density rather than the marginal density of the latent process.

From the Table, we see that the parameter estimates for the four SML methods are fairly consistent. The SML estimates for ρ are very much in accordance with the leverage parameter found in Yu (2005) (posterior mean = -0.3179) using the log-normal model and Bayesian estimation. In particular, we see that the parameter estimates for AS2 and AS3 are more similar than comparing say AS1 and AS2. This suggests that $K = 2$ is a sufficient order in the polynomial expansions for most practical applications under these ranges of parameters and time-steps.

4.1 Diagnostics

In addition to estimating parameters, we have also considered two forms of diagnostics for the above presented parameter estimation procedures. Firstly, we perform a battery of test to the residuals to assess the model fit. Secondly, we follow Koopman et al. (2009) and perform some tests on a finite variance of importance weights assumption in SML procedures.

4.1.1 Model fit: tests on residuals

Residuals for stochastic volatility models are not as standard as in the discrete time GARCH case, but some work has been done in Kim et al. (1998), Liesenfeld and Richard (2003) and Durham (2006). Here we propose to use a very simple estimator

$$y_i = \frac{x_i - \Delta\hat{a}}{\sqrt{\Delta \exp(\hat{z}_i)}} \tag{35}$$

for the standardized Brownian increments involved in the price process as the basis for the residual analysis. Here $\hat{a}$ is the SML estimate of a and we take $\hat{\mathbf{z}}$ to be the empirical Bayes smoothing estimator (Carlin and Louis, 1996), i.e. $\hat{\mathbf{z}} = E_{\theta=\hat{\theta}}[\mathbf{z}|\mathbf{x}]$, of the log-volatility. Under a severely miss-specified model, one would expect that $\mathbf{y}$ should deviate from being a vector of i.i.d. standard Gaussian variates.

The empirical Bayes estimator is particularly attractive from a computational perspective, as only minor adjustments to the EIS-SML code is needed to obtain an Independent Metropolis-Hastings (IMH) MCMC algorithm for computing the posterior mean $\hat{\mathbf{z}}$ (or any other desired moment). See, for example Robert and Casella (2004) p. 276 for a general treatment of IMH and Liesenfeld and Richard (2006) and Liesenfeld and Richard (2008) for IMH in the context of proposal distributions located using EIS. For each of the four SML procedures, we first locate an EIS importance density and then draw i.i.d. proposals $\mathbf{z}_p$ from the importance density. The acceptance probability, given the current state $\mathbf{z}_c$, has the form

$$p_{\text{accept}} = \min\left(\frac{w(\mathbf{z}_p)}{w(\mathbf{z}_c)}, 1\right) \tag{36}$$

The posterior means used here are based on chains of length 10000 and the acceptance rate for all four SML methods are between 0.7 and 0.8.

Normal QQ-plots for the residuals (35) are given in Figure 1 for the four SML procedures described above. None of the QQ-plots suggest any severe model-miss-specification. In addition, we have performed the standard battery of tests for temporal independence and normality of the

method	L-B(10)	L-B(15)	L-B(20)	J-B	K-S	M($2n^{1/3}$)	M($4n^{1/3}$)
EUL	0.0322	0.1262	0.3116	0.0409	0.0547	-6.4814	-9.2376
AS1	0.0287	0.1162	0.2906	0.1424	0.0698	-6.8556	-9.8941
AS2	0.0287	0.1158	0.2891	0.1467	0.0719	-6.4089	-8.8807
AS3	0.0286	0.1155	0.2886	0.1469	0.0719	-6.3747	-8.7028

Table 3: Various test-statistics. L-B(lags) contains the p-values for the Ljung-Box temporal dependence test for lags 10,15 and 20. J-B and K-S contains the p-values for the Jarque-Bera and Kolmogorov-Smirnov normality tests respectively. M(k) denotes the Monahan test with truncation k. The test statistics are asymptotically standard normal under the null-hypothesis that the importance weights have borderline infinite variance, and large negative values suggest a rejection towards finite variance.

residuals. The relevant test-statistics are summarized in Table 3. The Ljung-Box tests show that there may be some unexplained dependence for small numbers of lags. The EUL residuals yield borderline Jarque-Bera and Kolmogorv-Smirnov test-statistics for normality, whereas a suspicion of miss-specification cannot be supported by these normality tests when the polynomial expansions are used.

4.1.2 Tests for the importance weight variance

Recall that to have $\sqrt{M}$ convergence and asymptotic normality of the integral estimate (9), a finite variance of the importance weights is required. Recently, Koopman et al. (2009) proposed several tests for finite variance based on extreme value theory, and we shall apply some of their suggested methods here. Throughout this section, we consider the scaled importance weights $w' = \exp(\log w - 6540)$ as the values of w are too large for the floating point numerics used. This re-scaling does not affect the results presented, as the test statistics are invariant under re-scaling.

The tests are based on $N = 1000M = 32000$ importance weights obtained by evaluation of the EIS procedures 1000 times on the real data at the parameter estimated obtained for AS2. The 100 largest scaled weights, along with a histogram of the scaled weights are presented in Figures 2 and 3 for each of the four SML procedures. These preliminary diagnostics do not suggest infinite variance problems under any of the SML procedures.

More formal tests can be based on the peak over threshold methodology for i.i.d. observations.

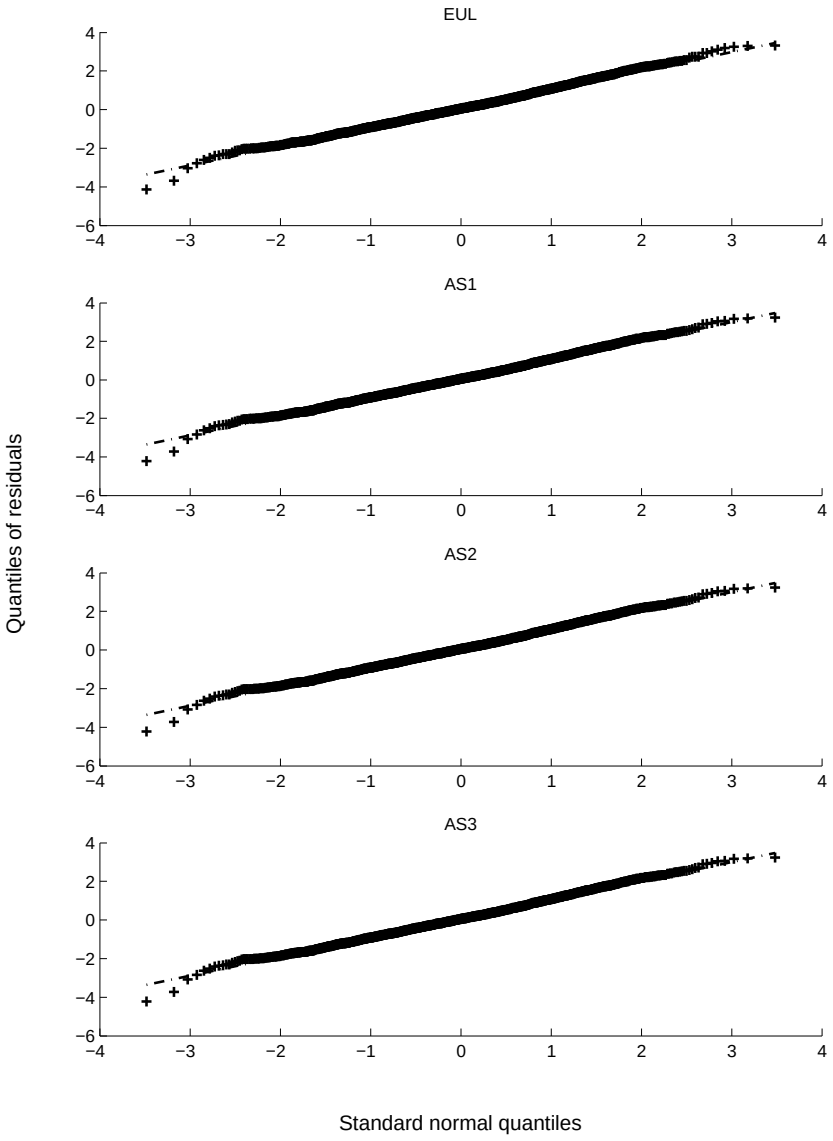

Figure 1: QQ-plot of the residuals (35).

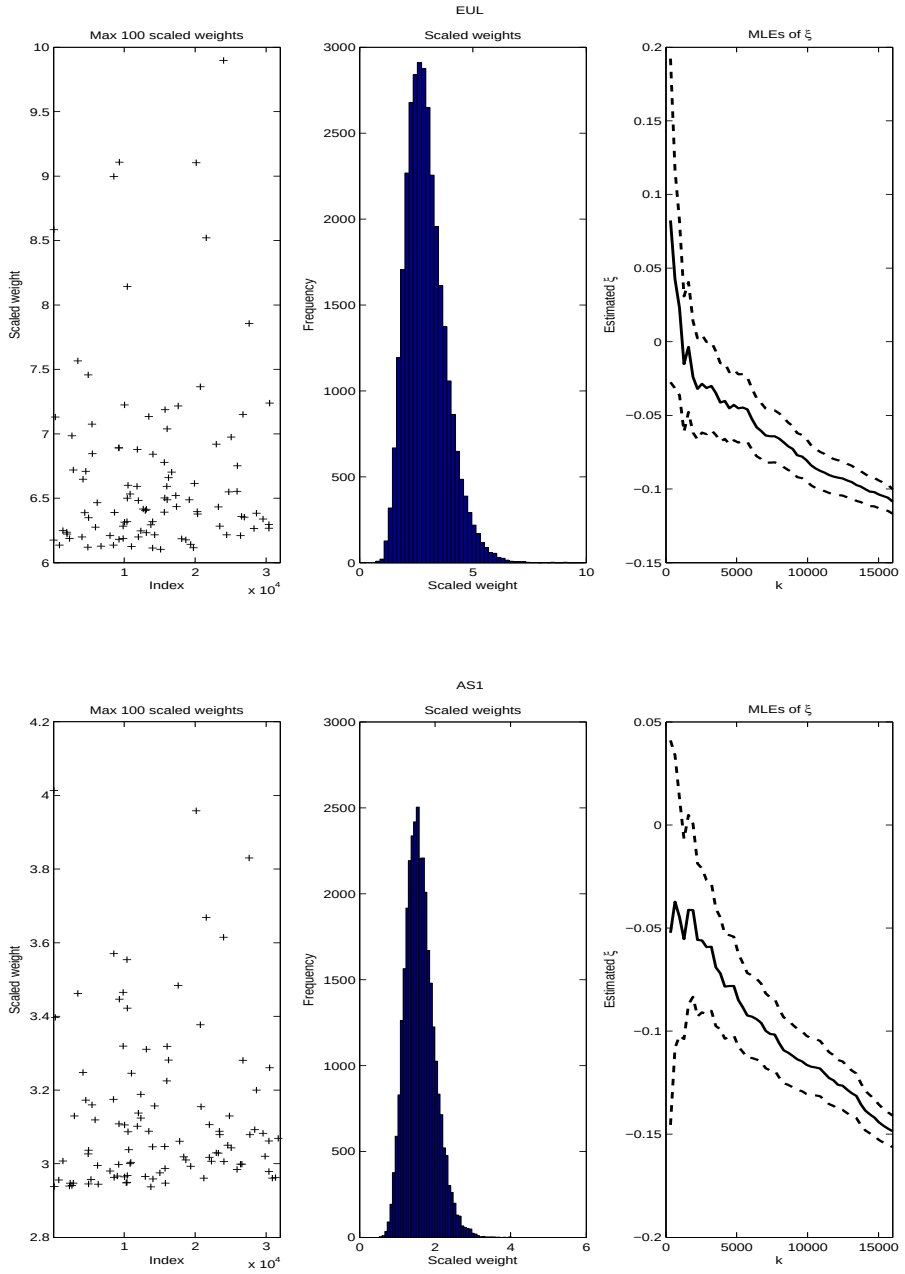

Figure 2: Finite variance diagnostics for EUL and AS1. The left hand side panels present the 100 largest scaled weights. The middle panels are histograms of all the scaled weights. The right hand side panel plots the maximum likelihood estimates of ξ (solid) along with 95% confidence bands (dashed) for different values of the truncation parameter k.

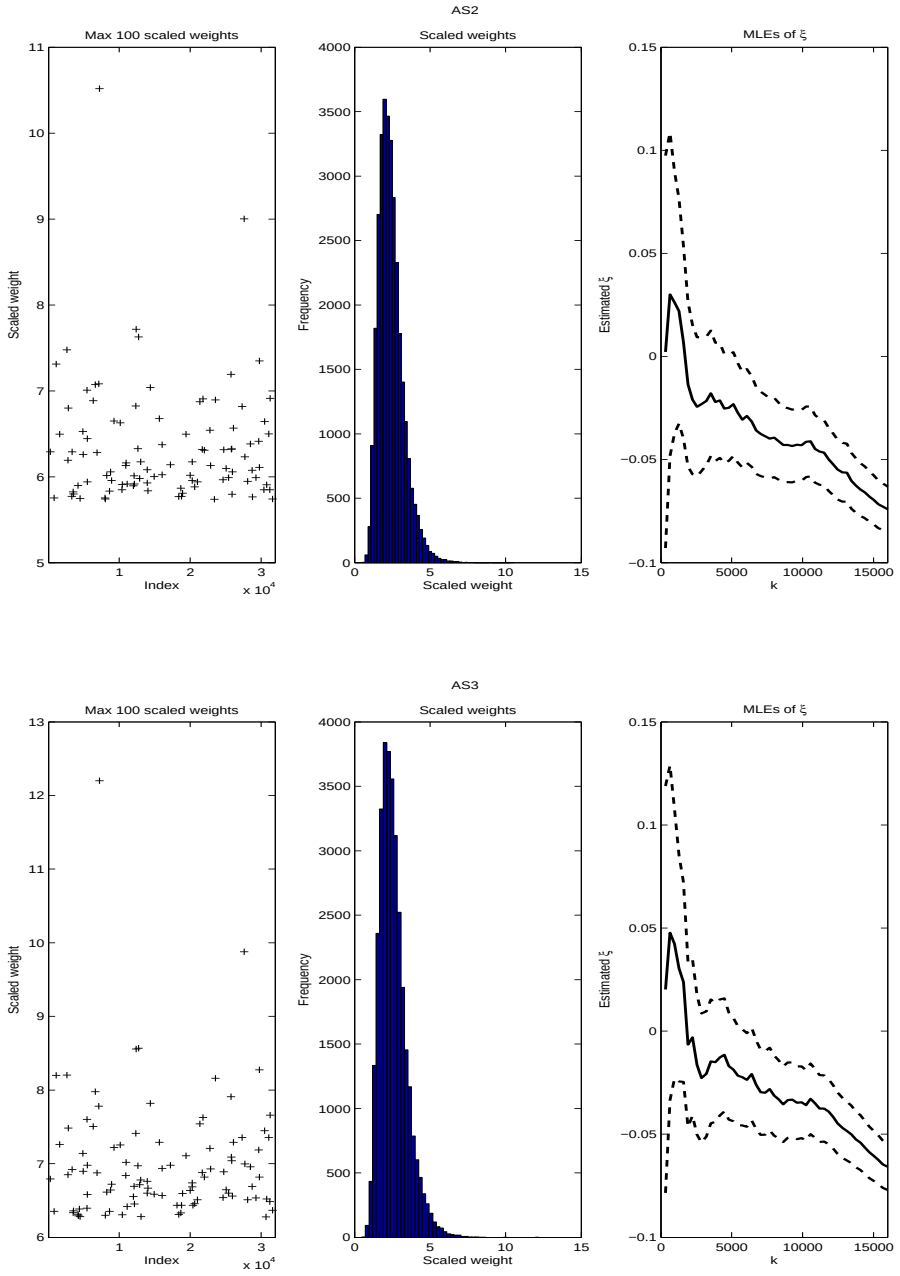

Figure 3: Finite variance diagnostics for AS2 and AS3. The left hand side panels present the 100 largest scaled weights. The middle panels are histograms of all the scaled weights. The right hand side panel plots the maximum likelihood estimates of ξ (solid) along with 95% confidence bands (dashed) for different values of the truncation parameter k.

A caveat here is that the importance weights are not exactly independent when they stem from the same EIS evaluation. Still, since the tests are invariant to a reordering of the data, we disregard this fact and proceed as if the data were i.i.d. Let $\{w'_{(j)}\}$ denote the scaled weights sorted in descending order. We define the "over threshold" weights (OTW) as $u_i = w'_{(i)} - w'_{(N-k)},\ i = 1, \ldots, k$ where k is a tuning parameter. Our aim is to measure the tail thickness of the OTWs as only the tails determine the finiteness of variance. The central tool for inference is the generalized Pareto distribution with density

$$f(u; \xi, b) = \frac{1}{b}\left(1 - \xi\frac{u}{b}\right)^{-\frac{1}{\xi}-1} \tag{37}$$

for which we fit to $\{u_i\}_{i=1}^N$ using two different methods. The parameter ξ determines the tail thickness, and in particular does $\xi < 1/2$ correspond to a finite variance. For $\xi < 0$, the Generalized Pareto distribution has finite support, and thus trivially finite variance. The parameter b is a scale parameter, whose actual value is of little interest for our application.

ML estimates of ξ are plotted in the rightmost plots of Figures 2 and 3 along with 95% confidence bands for values of k ranging from $[0.01N]$ to $[0.5N]$ where $[\cdot]$ denotes the integer part.[7] From the Figures, we see that the MLEs of ξ stay consistently below $1/2$ for any reasonable truncation parameter k.

In addition to the ML estimation of ξ, we apply Hill's estimator (see Hill (1975) or Phillips et al. (1996)) for ξ in the Generalized Pareto distribution. This estimator is given as

$$\xi_H = \frac{1}{k}\sum_{j=1}^{k} \log w'_{(N-j+1)} - \log w'_{(N-k)}, \tag{38}$$

and has a known asymptotically normal limit under some conditions on the relative growth of N and K. We follow Monahan (1993) and Koopman et al. (2009) and use $k = [2N^{1/3}]$ and $k = [4N^{1/3}]$ for this test. The last two columns of Table 3 gives us test-statistics that are asymptotically standard normal under null-hypothesis that the true $\xi = 1/2$, i.e. borderline infinite variance in the weights. Large (comparing with the standard normal distribution) negative test-statistics suggest rejection towards smaller values of ξ and finite variance. From the Table, we see strong evidence against the

[7] Obtained using the `gpfit`-function in MATLAB.

null-hypothesis and towards finite variance. All in all, the tests for finite variance of the importance weights conclusively points towards finite variance.

5 Conclusion

In this paper, we have introduced a methodology for computing SML estimates under the GARCH diffusion model where the discretization errors are controlled by applying a sequence of TPD approximations. The SML procedure performs numerically very well and there is no evidence of infinite variance issues in the importance sampler. For the progressively precise TPD-approximations, we see that there is a decreasing difference in the resulting SML estimates, suggesting that arbitrary accurate approximations to the exact continuous time likelihood based on discrete data can be produced in this manner. Of course, there is a tradeoff here, as the cost of reducing the discretization error corresponds to increasing cost of evaluation of the polynomial expansions of higher order. This trade-off arises because the latent variable has to be integrated out. As a reference, the AS2 expansion requires about 100 lines of machine generated `FORTRAN90` code to be evaluated, whereas the corresponding figure for AS3 is about 550.

There is a scope for the other applications of our method. The polynomial expansions are by no means restricted to the GARCH diffusion model, and there should obviously be scope for applying the current methodology within a broader class of models. However, the sampling rate Δ and the degree of deviation from the normality of the latent process are important parameters for whether this would be successful. It is well known that for Brownian motion driven stochastic differential equations, the TPD converges to a normal distribution as $\Delta \to 0$, and thus should the above proposed methodology produce precise results for sufficiently small Δ. However, this limit argument may not be of practical interest as data may be available only for larger Δ. If this is the case, one may wish to consider exchange the locally Gaussian importance density with a more problem specific non-Gaussian importance density. Alternatively, the saddlepoint approximation, that makes use of a non-Gaussian distribution as the leading term, may be useful.

Another possible direction for future research may be to employ the here described Independent

Metropolis-Hastings EIS algorithm to update $\mathbf{z}$ in a Gibbs sampler-based MCMC algorithm. This was done under the log-normal model in Liesenfeld and Richard (2006). However, due to the complicated form (in the parameters) of the coefficients in the polynomial expansions, a Metropolis-Hastings algorithm will generally also be needed to update the parameter vector.

Another direction for possible research is to allow for jumps either in the volatility or in the price process or both. Yu (2007) provides the corresponding TPD-expansions for jump-diffusions. Coping with jumps in the EIS framework can be done by introducing latent process consisting of jump counts (such as a binomial or Poisson). By alternating between iterating the EIS algorithm on the volatility (conditionally on the jump counts) and the count process, another EIS approximation to $p(\mathbf{z}|\mathbf{x})$ is obtained.

References

Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance 54*(4), 1361–1395.

Aït-Sahalia, Y. (2002a). Comment: Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics 20*(3), 317–321.

Aït-Sahalia, Y. (2002b). 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.

Aït-Sahalia, Y. and J. Yu (2006). Saddlepoint approximations for continuous-time Markov processes. *Journal of Econometrics 134*(2), 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.

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *The Journal of Finance 52*(5), 2003–2049.

Barone-Adesi, G., H. Rasmussen, and C. Ravanelli (2005). An option pricing formula for the GARCH diffusion model. *Computational Statistics and Data Analysis 49*(2), 287 – 310.

Bastani, H. and L. Guerrieri (2008). On the application of automatic differentiation to the likelihood function for dynamic general equilibrium models. In C. Bischof, H. M. Bucker, and P. Hovland (Eds.), *Advances in Automatic Differentiation*. Springer.

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

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

Chacko, G. and L. M. Viceira (2003). Spectral GMM estimation of continuous-time processes. *Journal of Econometrics 116*, 259–292.

Chernov, M. and E. Ghysels (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics 56*(3), 407 – 458.

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.

Duffie, D., J. Pan, and K. 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. (2007). SV mixture models with application to S&P 500 index returns. *Journal of Financial Economics 85*, 822–856.

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.

Elerian, O., S. Chib, and N. Shephard (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica 69*(4), 959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics 19*, 177–191.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *The Journal of Finance 58*(3), 1269–1300.

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

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.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics 3*, 1163–1174.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Financ 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.

Javaheri, A. (2005). *Inside volatility arbitrage: the secrets of skewness*. Wiley.

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

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies 65*(3), 361–93.

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

Kleppe, T. S., J. Yu, and H. J. Skaug (2009). Simulated maximum likelihood estimation of continuous time stochastic volatility models. Accepted for publication in Advances in Econometrics 2010 issue.

Koopman, S. J., N. Shephard, and D. Creal (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics 149*(1), 2–11.

Lewis, A. L. (2000). *Option Valuation under Stochastic Volatility*. Finance Press.

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.

Liesenfeld, R. and J.-F. Richard (2008). Improving MCMC, using efficient importance sampling. *Comput. Stat. Data Anal. 53*(2), 272–288.

Meddahi, N. (2002). Moments of continuous time stochastic volatility models. Unpublished Paper, University of Montreal.

Monahan, J. (1993). Testing the behaviour of importance sampling weights. In *Computer Science and Statistics: Proceedings of the 25th Annual Symposium on the Interface*, pp. 112–117.

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.

Øksendal, B. (2003). *Stochastic differential equations - An introduction with applications* (6 ed.). Springer.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics 22*(1), 55–71.

Phillips, P. and J. Yu (2009). 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., J. W. McFarland, and P. C. McMahon (1996). Robust tests of forward exchange market efficiency with empirical evidence from the 1920s. *Journal of Applied Econometrics 11*(1), 1–22.

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

Robert, C. P. and G. Casella (2004). *Monte Carlo statistical methods*. Springer.

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

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

Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics 102*(1), 111 – 141.

Skaug, H. J. (2002). Automatic differentiation to facilitate maximum likelihood estimation innon-linear random effects models. *Journal of Computational and Graphical Statistics 11*, 458–470.

Stein, E. M. and J. C. Stein (1991). Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies 4*, 727–752.

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.

Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *The Annals of Statistics 30*, 754–783.

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.

Yu, J. (2010). Simulation-based estimation methods for financial time series models. In J.-C. Duan, J. E. Gentle, and W. Härdle (Eds.), *Handbook of Computational Finance*. Springer. Forthcoming.

A Explicit expressions

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

$$\log \chi_i(z_{i-1}, x_i, \mathbf{a}_i) = \frac{1}{2}\log(\pi) - \frac{1}{2}\log\left(\frac{1}{2\Sigma_{\mathbf{0}_i}(z_{i-1})^2} - \mathbf{a}_{i,2}\right) - \frac{\mu_{\mathbf{0}_i}(z_{i-1}, x_i)^2}{2\Sigma_{\mathbf{0}_i}(z_{i-1})^2} - \frac{\left(\frac{\mu_{\mathbf{0}_i}(z_{i-1},x_i)}{\Sigma_{\mathbf{0}_i}(z_{i-1})^2} + \mathbf{a}_{i,1}\right)^2}{4\left(\mathbf{a}_{i,2} - \frac{1}{\Sigma_{\mathbf{0}_i}(z_{i-1})^2}\right)}. \tag{39}$$

where obvious alterations apply for $i = 0$. The details for (30) are given as

$$\begin{aligned}
\frac{p(\mathbf{x},\mathbf{z})}{m(\mathbf{z}|\mathbf{a})} &= \frac{p(z_0)\prod_{i=1}^n p_i(z_i|z_{i-1},x_i)}{m_0(z_0|\mathbf{a}_0)\prod_{i=1}^n m_i(z_i|z_{i-1},x_i,\mathbf{a}_i)} \\
&= \frac{A_0 B_0(z_0)\prod_{i=1}^n A_i(z_{i-1},x_i)B_i(z_i,z_{i-1},x_i)R_i(z_i,z_{i-1},x_i)}{\frac{B_0(z_0)\psi_0(z_0|\mathbf{a}_0)}{\chi_0(\mathbf{a}_0)}\prod_{i=1}^n \frac{B_i(z_i,z_{i-1},x_i)\psi_i(z_i|\mathbf{a}_i)}{\chi_i(z_{i-1},x_i,\mathbf{a}_i)}} \\
&= \frac{\chi_0(\mathbf{a}_0)A_0\prod_{i=1}^n \chi_i(z_{i-1},x_i,\mathbf{a}_i)A_i(z_{i-1},x_i)R_i(z_i,z_{i-1},x_i)}{\psi_0(z_0|\mathbf{a}_0)\prod_{i=1}^n \psi_i(z_i|\mathbf{a}_i)} \\
&= \chi_0(\mathbf{a}_0)A_0\left[\frac{\chi_1(z_0,x_1,\mathbf{a}_1)A_1(z_0,x_1)}{\psi_0(z_0|\mathbf{a}_0)}\right]\left[\frac{R_n(z_n,z_{n-1},x_n)}{\psi_n(z_n|\mathbf{a}_n)}\right]\times \\
&\qquad \prod_{i=1}^{n-1}\left[\frac{\chi_{i+1}(z_i,x_{i+1},\mathbf{a}_{i+1})A_{i+1}(z_i,x_{i+1})R_i(z_i,z_{i-1},x_i)}{\psi_i(z_i,\mathbf{a}_i)}\right]
\end{aligned} \tag{40}$$