8-2011

# Simulated Maximum Likelihood Estimation for Latent Diffusion Models

Tore Selland Kleppe

Jun Yu
*Singapore Management University*, yujun@smu.edu.sg

Hans J. Skaug

SMU
SINGAPORE MANAGEMENT
UNIVERSITY

# Simulated Maximum Likelihood Estimation for Latent Diffusion Models

## Tore Selland Kleppe, Jun Yu, Hans J. Skaug

August 2011

Paper No. 10-2011

# Simulated Maximum Likelihood Estimation for Latent Diffusion Models*

Tore Selland Kleppe                    Jun Yu

*University of Bergen*        *Singapore Management University*

Hans J. Skaug

*University of Bergen*

July 23, 2011

**Abstract**

In this paper a method is developed and implemented to provide the simulated maximum likelihood estimation of latent diffusions based on discrete data. The method is applicable to diffusions that either have latent elements in the state vector or are only observed at discrete time with a noise. Latent diffusions are very important in practical applications in financial economics. The proposed approach synthesizes the closed form method of Aït-Sahalia (2008) and the efficient importance sampler of Richard and Zhang (2007). It does not require any infill observations to be introduced and hence is computationally tractable. The Monte Carlo study shows that the method works well in finite sample. The empirical applications illustrate usefulness of the method and find no evidence of infinite variance in the importance sampler.

*JEL classification:* C11, C15, G12
*Keywords*: Closed-form approximation; Diffusion Model; Efficient importance sampler

# 1    Introduction

Diffusion models have proven to be very useful in economics and finance. For example, it provides a convenient mathematical framework for the development of financial economics and option pricing theory (Black and Scholes, 1973; Heston, 1993; Duffie and Kan, 1996) and for a separate treatment of stock variables and flow variables in macroeconomics (Bergstrom, 1984). Not surprisingly, estimation of diffusion models has received a great deal of attention in econometrics. One main difficulty in estimating diffusion models is that, although the model is formulated in continuous time, the observed

data are always collected in discrete time. This misalignment renders the calculation of likelihood and exact moments difficult. In particular, the difficulty in calculating the likelihood function is due to the lack of an analytical expression for the transition probability density (TPD). A simple solution to this misalignment is to discretize the continuous time model using the Euler-Maruyama (EM) method. However, the EM approximation introduces discretization bias, with the magnitude of the bias depending on the length of the sampling interval which in general is fixed.

Many methods have been proposed to reduce the discretization bias. Motivated by the fact that the discretization bias is smaller if the sampling intervals are shorter, the so-called infill method introduces additional latent variables holding the value of the diffusion at time points between the observations. To obtain the required TPD at the original frequency, these latent observations are integrated out from the product of the TPDs for the increased frequency. As the integrations are high dimensional, importance sampling is often used to evaluate the integrals numerically. This is the basic idea behind the infill maximum likelihood method of Pedersen (1995) and Durham and Gallant (2002). The more latent variables are introduced, the finer the partition becomes and consequently the discretization bias is reduced. However, as the dimension of the integration goes up the computational cost of achieving a specified numerical accuracy increases.

Aït-Sahalia (1999, 2002b) proposed a powerful alternative to address the problem of the discretization bias. The procedure is based on a series of closed-form expressions that can approximate arbitrarily well the true TPD at the original frequency as the number of the terms in the series expansion increases. It has been shown that in all practical situations this closed-form approach is able to approximate the TPD very accurately even with only a few terms included in the series expansion; see Aït-Sahalia (1999, 2002b) and Aït-Sahalia (2008). In addition, the approach is computationally efficient because it does not require infill observations or any Monte Carlo simulations. Aït-Sahalia (2008) generalized the technique to irreducible diffusions and multivariate diffusions.

While the closed-form method can essentially remove the discretization bias completely and is computationally inexpensive, a key assumption for its implementation is that the state variables, that are assumed to follow a diffusion, are observable. When some or all of the state variables are latent, the closed-form approach is not directly applicable. Examples of useful latent diffusion models include the entire class of continuous time stochastic volatility models with the volatility being the latent state; see Hull and White (1987), Heston (1993), Andersen and Lund (1997) and Duffie et al. (2000). A second example is the continuous time stochastic mean model of Balduzzi et al. (1998), in

which the mean is the latent state. Thirdly, the presence of market microstructure noise prevents the state variables from being directly observed. Hence, extension of the closed-form maximum likelihood estimation (MLE) method to cover latent diffusion models is important.

In this paper, we extend the closed-form MLE method of Aït-Sahalia (2008) to estimate models involving latent diffusions. The approach synthesizes the closed form method and the efficient importance sampler (EIS) of Richard and Zhang (2007) to provide the so-called simulated maximum likelihood (SML) estimator.

To deal with the challenge of latent variables in the context of continuous time stochastic volatility, Jones (2003) and Aït-Sahalia and Kimmel (2007) proposed to estimate the model using data from both the underlying spot and the options markets. Option prices were used to extract volatility, making the integration of volatility out of the joint TPDs unnecessary. It is well known that option prices are derived from the risk-neutral measure. Consequently, a benefit of using data from both the spot market and the options market jointly is that one can learn about the physical and the risk-neutral measures. However, this benefit comes at expense. To connect the physical and the risk-neutral measures, the functional form of the market price of risk has to be specified. If one's interest is to learn about the physical measure only, the implied volatility is less useful. Moreover, in some cases, such as for models with stochastic mean, it is not clear how to extract latent variables from derivative prices. In contrast, our SML approach does not rely on option pricing data and hence it allows us to estimate the model in the physical measure without worrying about the potential mis-specification of the market price of risk.

The paper is organized as follows. Section 2 discusses two classes of models and introduces the estimation method. Section 3 explains how to implement the method in two distinct examples, the GARCH diffusion model and a CEV model observed with a noise. In addition, in Section 3 we also examine the accuracy of the method using simulated data. In Section 4, we apply this estimation method to real data. Section 5 concludes.

## 2  Methodology

### 2.1  Model specifications

Let the time-homogeneous diffusion be denoted by

$$dX_\tau = a(X_\tau; \theta)d\tau + b(X_\tau; \theta)dB_\tau, \tag{1}$$

where $X_\tau$ and $a(X_\tau)$ are $q$-vectors, and $b(X_\tau)$ is a $q \times q$ matrix, with $B_\tau$ being an $q$-dimensional uncorrelated Brownian motion. $\theta$ is the vector of parameters to be estimated. We assume that (1) admits a unique solution. Let $x_t = X_{t\Delta}$ $(t = 1, \ldots, T)$ be the value of $X_\tau$ which is sampled at frequency $1/\Delta$ and $\mathbf{x} = (x_1, \ldots, x_T)$ be the collection of such values. We consider two different cases for the latent structure of $\mathbf{x}$, and the implications for likelihood inference about $\theta$:

**Case 1** $x_t$ is partly observed. Denote the observed part of $x_t$ by $y_t$, and the latent part by $z_t$, so that $x_t = [y_t \; z_t]'$. The likelihood is given as

$$L(\theta|\mathbf{y}) = p(\mathbf{y}; \theta) = \int p(\mathbf{y}, \mathbf{z}; \theta)d\mathbf{z}, \tag{2}$$

where $\mathbf{y} = (y_1, \cdots, y_T)$ and $\mathbf{z} = (z_1, \cdots, z_T)$, and $p$ represents a generic probability density.

**Case 2** The entire $x_t$ is unobserved, but an observation vector $y_t$, associated with $x_t$, is available. Denote by $p(y_t|x_t; \theta)$ the conditional density of $y_t$ given $x_t$. Jointly, $\mathbf{x}$ and $\mathbf{y}$ constitute a discrete time state space model, with likelihood function given by the integral

$$L(\theta|\mathbf{y}) = p(\mathbf{y}; \theta) = \int \left[ \prod_{t=1}^{T} p(y_t|x_t; \theta) \right] p(\mathbf{x}; \theta)d\mathbf{x}. \tag{3}$$

In both cases, the calculation, and later the maximization of $l(\theta|\mathbf{y}) := \log L(\theta|\mathbf{y})$ pose two substantial problems. Firstly, $p(\mathbf{x}; \theta)$ is available in closed form only in very special cases. Secondly, both (2) and (3) involve integration over high-dimensional spaces, with neither of the integrals having a closed form expression. To overcome these obstacles, we use the closed form expansions of Aït-Sahalia (2008) and the EIS algorithm of Richard and Zhang (2007), respectively, as outlined in Section 2.3. From now on, we make the dependence on the parameter vector $\theta$ implicit in the notation.

## 2.2 Transition density function approximations

Due to the Markovian property of the solution process $X_\tau$, the joint probability density function (PDF) of $\mathbf{x}$, conditional on $x_0$, may be written as

$$p(\mathbf{x}|x_0) = \prod_{t=1}^{T} p_t(x_t|x_{t-1}), \tag{4}$$

where $p_t = p_t(x_t|x_{t-1})$ is the TPD associated with (1). As the TPD of (1) has a closed form expression only for a few special cases, approximations are inevitable in general, and we shall denote by $\bar{p}_t = \bar{p}_t(x_t|x_{t-1})$, a generic TPD approximation. The approximate joint PDF of $\mathbf{x}$ is simply obtained by substituting $\bar{p}_t$ for $p_t$ in (4).

One simple way to approximate the TPD is to use the EM approximation, defined by

$$\bar{p}_t^{(E)}(x_t|x_{t-1}) = \mathcal{N}\left(x_t; x_{t-1} + \Delta a(x_{t-1}), \Delta b(x_{t-1})b(x_{t-1})'\right)$$

where $\mathcal{N}(x; m, \Sigma)$ is the PDF of $N(m, \Sigma)$, evaluated at $x$. However, for fixed $\Delta$ the EM approximation may lead to an unacceptable discretization bias.

It is known from the differential equation literature that the discretization bias decreases as the sampling interval decreases. One way of making the sampling interval arbitrarily small is to further partition the original interval into the sufficiently fine subintervals so that the discretization bias becomes negligible at the increased frequency. Consequently, one inevitably introduces latent variables between $x_{t-1}$ and $x_t$. To calculate $p_t$, we express it as the product of the TPDs evaluated via the EM method at the increased frequency. However, these latent observations must be integrated out. When the partition becomes finer, the discretization bias is closer to 0 but the dimension of the required integrations becomes higher. Pedersen (1995) proposed an importance sampling technique, based on the multivariate standard normal, to evaluate the integral numerically. Durham and Gallant (2002) suggests ways to improve computational efficiency of this simulated infill ML method of Pedersen.

In this paper we bypass infill simulations by employing the closed-form expansion approximations to TPDs for irreducible diffusions of Aït-Sahalia (2008). Though more cumbersome to derive, the Aït-Sahalia expansions are attractive in that they have closed form with adjustable accuracy. This enables us to study the errors resulting from applying the EM-TPDs by considering a sequence of Aït-Sahalia expansions.

The Aït-Sahalia expansions of order $K$ have the form

$$\log \bar{p}_t^{(K)}(x_t|x_{t-1}) \equiv -\frac{q}{2}\log(2\pi\Delta) - D_v(x_t) + \frac{C_{-1}^{j_{-1}}(x_t|x_{t-1})}{\Delta} + \sum_{k=0}^{K} C_k^{j_k}(x_t|x_{t-1})\frac{\Delta^k}{k!},$$

where

$$D_v(x) = \frac{1}{2}\log(\mathrm{Det}(b(x)b(x)')).$$

Clearly, the expansion has the interpretation as a functional power series in $\Delta$ (plus some additional terms). Increasing precision in the sense described in Aït-Sahalia (2008) is obtained by increasing $K$. The coefficients $C_k^{j_k}$ are polynomials of the form

$$C_k^{j_k}(r|s) = \sum_{|i| \le j_k} c_i^{(k)}(r_1 - s_1)^{i_1}(r_2 - s_2)^{i_2}\cdots(r_q - s_q)^{i_q},$$

where $i = (i_1, \ldots, i_m)$ is a multi-index with trace $|i|$ at most $j_k = 2(K - k)$.[1] The form of the coefficients $c_i^{(k)}$ are found by solving both the Forward- and Backward Kolmogorov partial differential equations to the appropriate orders in $\Delta$ using the algorithms outlined in Aït-Sahalia (2008). The actual expressions for $C_k^{j_k}$ for each particular model are in general complicated, and we obtained these using the symbolic manipulation software Maple. Their exact specification is available upon request in computer form from the authors.

It is worth noticing that the Aït-Sahalia expansions are not proper densities as they do not exactly integrate to one. However, in our experience the expansions are very accurate for the models considered here, so that a re-normalization is unnecessary.

## 2.3 Efficient importance sampling

The second obstacle faced is the calculation of the marginalization integrals as in (2) and (3), for which no closed form expression can be found. Again we need approximation. What we propose is to use a Monte Carlo integration method, namely, the EIS method of Richard and Zhang (2007). The EIS is chosen as it does not rely on a global near-Gaussian kernel assumption of the integrand, which is required by the Laplace approximation (Shephard and Pitt (1997); Durbin and Koopman (1997)).

Here we shall explain a version of the EIS with restricted generality, relying on Gaussian local samplers, as this will suffice for our needs. The choice of locally Gaussian samplers relies on the observation that the TPDs of diffusions can often be approximated quite accurately by Gaussian densities for "short" time steps $\Delta$. For the examples considered here, we show later that the sparsely parameterized locally Gaussian samplers suffice to get well-performing importance sampling procedures. For

---

[1] 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.

ease of exposition, we shall restrict ourselves to univariate latent states, but multivariate latent states are possible using multivariate locally Gaussian EIS samplers (Liesenfeld and Richard, 2003). For a more general exposition of the EIS we refer to Richard and Zhang (2007).

Let the integrand in the integral $I$ that we wish to approximate have a simple factorization

$$I = \int \varphi(\lambda_{(T)})d\lambda_{(T)} = \int \prod_{t=0}^{T} \varphi_t(\lambda_t|\lambda_{t-1})d\lambda_{(T)}, \ \lambda_{(T)} = (\lambda_0, \ldots, \lambda_T),$$

with $\lambda_t \in \mathbb{R}$, which is the case for the problems considered here. For notational ease, $\varphi_0(\lambda_0|\lambda_{-1}) = \varphi(\lambda_0)$. Our aim is to find an importance density $m(\lambda_{(T)})$ so that the importance sampling estimate, represented by the right hand side of

$$I = \int \frac{\varphi(\lambda_{(T)})}{m(\lambda_{(T)})}m(\lambda_{(T)})d\lambda_{(T)} \approx \hat{I} = \frac{1}{M}\sum_{j=1}^{M} \frac{\varphi(\lambda_{(T)}^{(j)})}{m(\lambda_{(T)}^{(j)})}, \ \lambda_{(T)}^{(j)} \sim \text{i.i.d. } m(\lambda_{(T)}), \tag{5}$$

has as small variance as possible, under the restriction that sampling from $m$ is computationally easy. The smaller the variance of $\varphi(\lambda_{(T)})/m(\lambda_{(T)})$, the smaller value for $M$, the number of random draws, is needed. The EIS restricts the importance density $m(\lambda_{(T)})$ to have a Markovian structure

$$m(\lambda_{(T)}) = m_0(\lambda_0) \prod_{t=1}^{T} m_t(\lambda_t|\lambda_{t-1}), \tag{6}$$

with each factor $m_t$ specified as

$$m_t(\lambda_t|\lambda_{t-1}; \mathbf{a}_t) = \frac{k_t(\lambda_t|\lambda_{t-1})\psi_t(\lambda_t; \mathbf{a}_t)}{\chi_t(\lambda_{t-1}; \mathbf{a}_t)}, \ \mathbf{a}_t = (a_{t,1}, a_{t,2})', \tag{7}$$

$$\psi_t(\lambda_t; \mathbf{a}_t) = \exp(a_{t,1}\lambda_t + a_{t,2}\lambda_t^2), \tag{8}$$

$$\chi_t(\lambda_{t-1}; \mathbf{a}_t) = \int k_t(\lambda_t|\lambda_{t-1})\psi_t(\lambda_t; \mathbf{a}_t)d\lambda_t, \tag{9}$$

$$\log k_t(\lambda_t|\lambda_{t-1}) = \sum_{q=0}^{2} \frac{F^{(q)}(\lambda_{t-1})}{q!}(\lambda_t - \lambda^*(\lambda_{t-1}))^q, \tag{10}$$

with $m_0(\lambda_0|\lambda_{-1}) \equiv m(\lambda_0)$, and correspondingly for the other expressions. The definitions of $F^{(q)}$ and $\lambda^*$ will be given below. Notice that $m_t$ is a Gaussian density with mean and variance given as

$$\mu_t = -\frac{F^{(1)} - F^{(2)}\lambda^* + a_{t,1}}{F^{(2)} + 2a_{t,2}}, \ \Sigma_t = -\frac{1}{F^{(2)} + 2a_{t,2}}, \tag{11}$$

7

which makes sampling from (6) conceptually simple and computationally fast.

Flexibility in $m(\lambda_{(T)}) = m(\lambda_{(T)}; \mathbf{a})$ is introduced by the parameters $\mathbf{a} = (\mathbf{a}_0, \cdots, \mathbf{a}_T)$, and the aim of the EIS is to choose $\mathbf{a}_t$ to minimize the variance of $\hat{I}$. Plugging (7) into (5), we obtain

$$I = \chi_0(\mathbf{a}_0) \int \left[ \prod_{t=0}^{T} \frac{\varphi_t(\lambda_t|\lambda_{t-1})\chi_{t+1}(\lambda_t; \mathbf{a}_{t+1})}{k_t(\lambda_t|\lambda_{t-1})\psi_t(\lambda_t; \mathbf{a}_t)} \right] m(\lambda_{(T)}; \mathbf{a})d\lambda_{(T)}, \tag{12}$$

where $\chi_{T+1} \equiv 1$. The EIS proceeds by introducing draws $\left\{ \lambda_{(T)}^{(j)} \right\}_{j=1}^{M} \sim m(\lambda_{(T)}; \mathbf{a})$, and minimizes the Monte Carlo variance of the logarithm of each factor in the product of (12) as

$$\hat{\mathbf{a}}_t, \hat{c}_t = \arg\min_{\mathbf{a}_t, c_t} \sum_{j=1}^{M} \left[ \log\left( \frac{\varphi_t(\lambda_t^{(j)}|\lambda_{t-1}^{(j)})\chi_{t+1}(\lambda_t^{(j)}; \mathbf{a}_{t+1})}{k_t(\lambda_t^{(j)}|\lambda_{t-1}^{(j)})} \right) - c_t - a_{t,1}\lambda_t^{(j)} - a_{t,2}(\lambda_t^{(j)})^2 \right]^2, \tag{13}$$

for $t = T, T-1, \ldots, 0$. Clearly, due to the parameterization of $m_t$, the above minimization problem is a linear least squares problem, which admits the application of computationally attractive linear regression routines to calculate the solutions.[2]

## 2.4 EIS samplers for diffusion models

With the generic EIS algorithm in place, we are ready to specify the functional form of $\varphi_t$ and $k_t$ for the two models at hand.

### 2.4.1 Case 1

In this case, the latent state is denoted by $\lambda_t = z_t$, and we write $\bar{p}_t(x_t|x_{t-1}) = \bar{p}_t(z_t, y_t|z_{t-1}, y_{t-1})$. The components in the integrand may be written as

$$\varphi_t(z_t|z_{t-1}) = \begin{cases} \bar{p}_0(z_0) & \text{for } t = 0, \\ \bar{p}_t(z_t, y_t|z_{t-1}, y_{t-1}) & \text{for } t = 1, \ldots, T. \end{cases}$$

The initial sampler kernel $k_t$ is obtained using a second order Taylor approximation around an expansion point $z^*(z_{t-1})$ as

---

[2] As the draws $\left\{ \lambda_{(T)}^{(j)} \right\}_{j=1}^{M}$ themselves depend on $\mathbf{a}$, the regression problems should be regarded as a fixed point condition, towards we generate a convergent sequence (Richard and Zhang, 2007).

$$\log \bar{p}_t(z_t, y_t | z_{t-1}, y_{t-1}) = \underbrace{\sum_{q=0}^{2} \frac{F_t^{(q)}(z_{t-1})}{q!} (z_t - z^*(z_{t-1}))^q}_{\log k_t(z_t | z_{t-1})} + Q_t(z_t | z_{t-1}),$$

$$F_t^{(q)}(z_{t-1}) \equiv \left[ \frac{\partial^q}{\partial z_t^q} \log \bar{p}_t(z_t, y_t | z_{t-1}, y_{t-1}) \right]_{z_t = z^*(z_{t-1})}.$$

For the EM-TPD, this reduces to $k_t = \bar{p}_t$. For the Aït-Sahalia expansions, finding $k_t$ is mainly a matter of rearranging the polynomial terms in ascending order in $(z_t - z^*(z_{t-1}))$ rather than in $\Delta$. However, the non-linear function $D_v(x_t)$ needs to be Taylor expanded. Still, no additional error is committed since all the residual variation is retained in $Q_t$. The expansion point $z^*(z_{t-1})$ should be chosen so the $k_t$ closely approximates $\bar{p}(z_t | y_t, z_{t-1}, y_{t-1})$. In the GARCH diffusion model considered below, we set $z^*(z_{t-1}) = E_{\bar{p}^{(E)}}[z_t | y_t, z_{t-1}, y_{t-1}]$ which has a simple closed form expression. Under these specifications, the EIS regresses

$$\log Q_t(z_t^{(j)} | z_{t-1}^{(j)}) \chi_{t+1}(z_t^{(j)}; \mathbf{a}_{t+1}) \quad \text{on} \quad \text{constant} + z_t^{(j)} + (z_t^{(j)})^2, \ t = T, \dots, 1. \tag{14}$$

The handling of the initial latent state $z_0$ is model specific. One possibility is to treat $z_0$ as fixed and known. In the GARCH diffusion example considered below, $\bar{p}_0$ is taken to be a Gaussian approximation to $p(Z_\tau)$. This leads to $k_0 = \bar{p}_0$ and an initial regression of

$$\log \chi_1(z_0^{(j)}) \quad \text{on} \quad \text{constant} + z_0^{(j)} + (z_0^{(j)})^2. \tag{15}$$

### 2.4.2 Case 2

In this case, the whole diffusion state is unobserved and $\lambda_t = x_t$. The factorization of the integrand based on approximate TPDs may be written as

$$\varphi_t(x_t | x_{t-1}) = \begin{cases} \bar{p}_0(x_0) & \text{for } t = 0, \\ \bar{p}_t(x_t | x_{t-1}) g(y_t | x_t) & \text{for } t = 1, \dots, T. \end{cases}$$

For $t = 1, \dots, T$, the initial sampler kernel $k_t$ is again obtained as a second order Taylor-approximation of $\log \bar{p}_t$ around an expansion point $x^*(x_{t-1})$, i.e.

$$\log \bar{p}_t(x_t|x_{t-1}) \quad = \quad \underbrace{\sum_{q=0}^{2} \frac{F_t^{(q)}(x_{t-1})}{q!}(x_t - x^*(x_{t-1}))^q}_{\log k_t(x_t|x_{t-1})} + Q_t(x_t|x_{t-1}),$$

$$F_t^{(q)}(x_{t-1}) \quad \equiv \quad \left[\frac{d^q}{dx_t^q}\log \bar{p}_t(x_t|x_{t-1})\right]_{x_t = x^*(x_{t-1})}.$$

In the CEV diffusion model considered below, the expansion point is simply set to $x^* = x_{t-1}$. The regression problem (13) reduces to regressing

$$\log Q_t(x_t^{(j)}|x_{t-1}^{(j)})g(y_t|x_t^{(j)})\chi_{t+1}(x_t^{(j)}; \mathbf{a}_{t+1}) \quad \text{on} \quad \text{constant} + x_t^{(j)} + (x_t^{(j)})^2, \ t = T, \dots, 1.$$

and $x_0$ is treated as known and fixed.

# 3 Specific models and simulation Study

To examine the performance of the proposed procedure, we estimate two diffusion models using simulated data. The first model is the GARCH diffusion of Nelson (1990), a special case of Case 1. The second model is the CEV diffusion observed with an i.i.d. noise, a special case of Case 2.

For both models, the algorithms are implemented in `FORTRAN90`. Following Skaug (2002) and Bastani and Guerrieri (2008), we use algorithmic differentiation to generate code for the exact gradient of the simulated likelihood function. A line searching BFGS-quasi-Newton optimizer (Nocedal and Wright, 1999) is applied to maximize the simulated likelihood function using function values and exact gradients.

## 3.1 The GARCH diffusion model

Let $Y_t$ denote the log-price of some asset, and $V_t$ the volatility of this asset. Then the GARCH diffusion model is given by

$$d\begin{bmatrix} Y_t \\ V_t \end{bmatrix} = \begin{bmatrix} a \\ \alpha + \beta V_t \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma V_t \end{bmatrix} \begin{bmatrix} dB_{t,1} \\ dB_{t,2} \end{bmatrix}, \tag{16}$$

where $B_{t,1}$ and $B_{t,2}$ denote a pair of independent canonical Brownian motions. The parameters to be determined are $\theta = [\alpha, \beta, \sigma, \rho, a]$. Provided that $\beta < 0$, the volatility process $V_t$ is mean reverting to the long run mean, $-\alpha/\beta$. 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)). The parameter $\rho \in (-1, 1)$ represents the so called leverage effect (Yu, 2005). The model was first obtained by Nelson as a continuous time limit of the discrete time GARCH(1,1) model of Bollerslev (1986). Duan and Yeh (2011) recently showed that this model provides much better empirical fit to actual data than the square root stochastic volatility model of Heston (1993).

For convenience, we follow Aït-Sahalia (2002b) and Durham and Gallant (2002) and apply the variance stabilizing transformation to $V_t$, so that the transformed volatility has constant infinitesimal variance. There are two reasons for doing this. Firstly, it appears that $p(z_t|y_t, z_{t-1}, y_{t-1})$ is better approximated by a Gaussian importance distribution. Secondly, it is our experience that the Aït-Sahalia expansions converge faster when the domain of the diffusion is doubly unbounded. We define $Z_t = \log(V_t)$ and apply Ito's lemma to find the joint dynamics of $Y_t$ and $Z_t$

$$
d\begin{bmatrix} Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} a \\ (\beta - \frac{1}{2}\sigma^2) + \alpha \exp(-Z_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)} \exp\left(\frac{1}{2}Z_t\right) & \rho \exp\left(\frac{1}{2}Z_t\right) \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} dB_{t,1} \\ dB_{t,2} \end{bmatrix}.
$$
(17)

We assume that only $Y_t$ is observed at discrete times with time step $\Delta$. Moreover, the log-volatility $Z_t$ is assumed to be unobserved. 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 the standard deviation characterizing are given, respectively, as

$$
-\log\left(\frac{\sigma^2 - 2\beta}{2\alpha}\right) \quad \text{and} \quad \frac{\sigma^2}{\sigma^2 - 2\beta}.
$$
(18)

This simplification is mainly done for convenience when constructing the importance sampler, and the errors committed are asymptotically small when $T$ increases. We proceed with studying the statistical properties of the SML estimator for this model.

The setup for the study is as follows. We generate daily data (i.e. $\Delta = 1/252$) and use sample size $T = 2,022$ (matching the sample size in the real data discussed later) corresponding to roughly

8 years of data. We simulate 1,000 data sets using the EM scheme with time step $\Delta/256$. Since the time step is so small, the data can be regarded as coming from the continuous time model. We shall use the acronyms EUL for the method based on EM-TPDs and AS1, AS2 and AS3 for the method based on the closed-form expansions with $K = 1$, 2 and 3. For the SML, we consistently use $M = 16$ draws both for the MC study and the empirical study. (S)ML estimators are obtained when volatility is assumed to be observed and also assumed to be latent.

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

- The discretization error generated by the Euler method. As we employ a sequence of polynomial expansions in addition to the EM discretization, the discretization error may be assessed accurately by comparing the EM method and the closed-form method with higher order polynomial expansions.

- 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 completely observed diffusion processes. In particular, Phillips and Yu (2009) show that the ML estimate tends to be biased towards a faster mean reversion. This claim may be checked in our ML estimates of $\beta$ when the volatility is observed or unobserved.

- The errors generated from the Monte Carlo simulation at the EIS stage when volatility is assumed to be latent. This is possible because when volatility is observed, we do not need to integrate out latent variable and the closed-form method is directly applicable to $(Y_t, Z_t)$. Hence, the comparison of two sets of estimators, one based on observed volatility and one based on unobserved volatility, tells us whether faith can be put into the importance sampler. This source of errors will also be addressed in Section 4.1, 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 4.1 are used as the "true parameters" throughout the complete experiment. The mean computing times for locating the SML estimates ranges from 43 seconds (EUL) to 103 seconds (AS3) on a Dell PowerEdge R200 computer with an Intel Xeon X3330 2.66GHz Quad core processor. The routines for evaluating $\mu_t$, $\Sigma_t$ and $Q_t$ are distributed on the four cores of the computer. A total of 12 EIS iterations were used, with

12

| method | $\alpha$ | $\beta$ | $\sigma$ | $\rho$ | $a$ |
|---|---|---|---|---|---|
| true parameters | 9.4753e-02 | -1.1754e+00 | 3.2607e+00 | -8.4668e-01 | -1.8307e-02 |
| observed log-volatility | | | | | |
| EUL | -1.1348e-02 | 1.3205e-01 | -5.3359e-02 | 3.4870e-03 | 2.1317e-02 |
| | (3.2606e-02) | (9.8615e-01) | (3.6189e-02) | (5.1728e-03) | (8.2130e-02) |
| AS1 | -3.5503e-03 | 1.2237e-01 | -2.2063e-03 | -4.3671e-05 | 4.6145e-03 |
| | (2.2912e-02) | (9.4922e-01) | (3.6171e-02) | (5.1024e-03) | (5.9712e-02) |
| AS2 | -1.6758e-03 | 4.1729e-02 | -2.2890e-03 | -1.4109e-04 | 3.8146e-03 |
| | (2.3461e-02) | (9.8974e-01) | (3.6064e-02) | (5.1145e-03) | (6.1019e-02) |
| AS3 | -1.4217e-03 | 3.5122e-02 | -2.7328e-03 | -6.9178e-05 | 3.3318e-03 |
| | (2.3029e-02) | (9.8181e-01) | (3.5930e-02) | (5.1032e-03) | (6.0604e-02) |
| unobserved log-volatility | | | | | |
| EUL | 1.7676e-03 | -5.5798e-01 | -2.0729e-01 | 2.8468e-02 | 1.4380e-02 |
| | (1.9159e-02) | (1.1675e+00) | (2.4401e-01) | (4.6443e-02) | (4.3687e-02) |
| AS1 | 4.4984e-04 | 5.5142e-02 | -4.5956e-02 | -4.3213e-03 | -6.4342e-03 |
| | (1.8587e-02) | (9.4814e-01) | (2.5800e-01) | (3.6043e-02) | (4.4929e-02) |
| AS2 | 4.7073e-03 | -1.0854e-01 | -1.2901e-02 | -6.7084e-03 | -6.5838e-03 |
| | (1.9204e-02) | (1.0473e+00) | (2.6269e-01) | (3.6248e-02) | (4.4012e-02) |
| AS3 | 4.3914e-03 | -1.1067e-01 | -1.4859e-02 | -5.8453e-03 | -5.9800e-03 |
| | (1.9189e-02) | (1.0458e+00) | (2.6225e-01) | (3.5774e-02) | (4.4464e-02) |

Table 1: Results from the Monte Carlo experiment for the GARCH diffusion. All results are based on 1,000 simulated data sets using the parameters given in the "True parameters" row. The bias (no parenthesis) is calculated as the average of the estimates minus the true parameter. Statistical standard errors are given in parenthesizes.

the 6 first iterations based on the EM approximation to ensure greater stability and faster convergence of the algorithm. In addition, we provide the following starting conditions

$$a_{t,1}^{(0)} = \frac{1}{2}\log(\max(y_{t+1}^2, 0.00001)/\Delta), \ t = 0, \ldots, T-1, \tag{19}$$

$$a_{t,2}^{(0)} = -\frac{1}{4}, \ t = 0, \ldots, T-1, \tag{20}$$

so that $\psi_t(z_t; \mathbf{a}_t^{(0)})$ is close to be proportional to $\bar{p}^{(E)}(y_{t+1}|z_t)$ for $t = 0, \ldots, T-1$. We also set $a_{T,1}^{(0)} = a_{T,2}^{(0)} = 0$ initially.

Table 1 reports the bias $(\widehat{\theta} - \theta)$ and the standard error of each estimate across the 1,000 simulation

replica. We see that there are differences in the estimates when volatility is assumed to be observed from those when volatility is assumed to be latent. A relatively more striking difference is the under-estimation of $\sigma$ under the EM-TPDs, whereas the bias is smaller for the polynomial expansions. This observation seems to be consistent with what has been found in Aït-Sahalia (1999). In addition, we see a larger bias in $\rho$ for the EUL-based routines that seem to be mitigated when the expansions are applied. The expected bias towards faster mean reversion is seen as an overestimation of $-\beta$ in the EUL, AS2 and AS3 procedures. Interestingly this does not occur when the volatility is observed. 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, we see that the loss of statistical precision is most significant for the $\sigma$ and $\rho$ parameters where a ten-fold increase in the standard error is seen. The parameters governing the linear drift of the volatility, $\alpha$ and $\beta$, have similar statistical standard errors when the log-volatility is integrated out.

## 3.2 CEV diffusion observed with noise

In this example, we shall consider the constant elasticity of volatility (CEV) short term interest rate model of Chan et al. (1992), but disturb it with an independent and identically distributed (i.i.d.) noise. The model for the (unobserved true) interest rate $R_\tau$ is specified as

$$dR_\tau = (\alpha + \beta R_\tau)d\tau + \sigma R_\tau^\gamma dB_\tau, \tag{21}$$

and we assume we have noisy observations

$$y_t = r_t + \sigma_y \varepsilon_t, \; r_t = R_{t\Delta}, \; \varepsilon_t \sim \text{i.i.d. } N(0,1), \; t = 1, \ldots, T \tag{22}$$

Here $\theta = [\alpha, \beta, \sigma, \gamma, \sigma_y]$ is the parameter vector to be determined. The reason why the interest rate data are contaminated may be the presence of microstructure effects. In the case of interest rates, one obvious reason why microstructure effects may be important is due to the discreteness in interest rates. For example, The Federal Reserve, the central bank of the United States, only changes the discount rate by multiples of 25 basis points. In the case of measuring volatility, microstructure effects have motivated Zhang et al. (2005), Aït-Sahalia et al. (2005) and many others to introduce methods to construct new realized volatility estimates for integrated volatility. In the case of measuring jump

14

intensity, microstructure effects have motivated Duan and Fulop (2007) to use a model of the same structure as in (21) and (22).

For computational convenience, we again transform the latent process $r_\tau$ to a process with constant volatility term by introducing the transformation $\tilde{x}(r) = (r^{1-\gamma} - 1)/(1 - \gamma)$ and its inverse $\tilde{r}(x) = (1 + x(1 - \gamma))^{1/(1-\gamma)}$. By Ito's lemma, the variance stabilized process $X_\tau = \tilde{x}(R_\tau)$ solves the stochastic differential equation

$$dX_\tau = \left[ \alpha \tilde{r}(X_\tau)^{-\gamma} + \beta \tilde{r}(X_\tau)^{1-\gamma} - \frac{1}{2} \sigma^2 \gamma \tilde{r}(X_\tau)^{\gamma-1} \right] d\tau + \sigma dB_\tau. \tag{23}$$

Correspondingly, the conditional PDF of the observations has the form

$$g(y_t|x_t) \propto \exp\left( -\frac{(y_t - \tilde{r}(x_t))^2}{2\sigma_y^2} \right). \tag{24}$$

Together, the latent model (23) at discrete times and the observation noise specification (24) constitute a state space system, which is a special case of Case 2. We start by considering a Monte Carlo study using simulated daily data ($\Delta = 1/252$).

In this simulation study we focus on the effect of failing to account for measurement errors when estimating diffusions using SML, as well as the effect of using different TPD approximations. We employ EUL, AS1 and AS2 as the latter appears to have sufficient precision for the model considered. We use $M = 16$ in all the replications.

Data are simulated from the model using the EM discretization with time-step $\Delta/256$. The computing times ranges from 11 seconds for EUL-based SML to 70 seconds for AS2-based SML to maximize a likelihood. The initial observation $z_0$ is set to $\tilde{z}(y_1)$. We use 4 EIS iterations, and the initial value of $\mathbf{a}$ is set to

$$a_{t,1}^{(0)} = \frac{y_t^{\gamma+1} - y_t^{2\gamma}}{\sigma_y^2(1 - \gamma)}, \ t = 1, \dots, T, \tag{25}$$

$$a_{t,2}^{(0)} = -\frac{y_t^{2\gamma}}{2\sigma_y^2}, \ t = 1, \dots, T, \tag{26}$$

so that $\psi_t(x_t; \mathbf{a}_t^{(0)})$ is proportional to Laplace approximation of $x_t \mapsto g(y_t|x_t)$.

Table 2 reports the bias $(\widehat{\theta} - \theta)$ and the standard error of each estimate across 1,000 replications under three scenarios. In the first scenario, we estimate the model given by (21) and (22) based on $y_t$. In the second scenario, we estimate the model given by (21) based on non-noisy observations of

15

| method | $\alpha$ | $\beta$ | $\sigma$ | $\gamma$ | $\sigma_y$ |
|---|---|---|---|---|---|
| true parameters | 9.6539e-03 | -1.6434e-01 | 4.2469e-01 | 1.2011e+00 | 5.2280e-04 |
| simulated maximum likelihood based on $y_t$ | | | | | |
| EUL | 2.2690e-02 | -4.1369e-01 | 1.5406e-02 | -3.6380e-03 | -7.6361e-07 |
| | (2.4675e-02) | (4.1536e-01) | (1.4114e-01) | (1.0211e-01) | (3.0504e-05) |
| AS1 | 2.2750e-02 | -4.1475e-01 | 1.4157e-02 | -4.9749e-03 | -7.6071e-07 |
| | (2.4746e-02) | (4.1672e-01) | (1.4037e-01) | (1.0206e-01) | (3.0515e-05) |
| AS2 | 2.2757e-02 | -4.1485e-01 | 1.4152e-02 | -4.9786e-03 | -7.6064e-07 |
| | (2.4764e-02) | (4.1682e-01) | (1.4037e-01) | (1.0206e-01) | (3.0515e-05) |
| direct maximum likelihood based on $r_t = R_{\Delta t}$ | | | | | |
| AS2 | 2.2499e-02 | -4.0999e-01 | 5.5807e-03 | -1.6990e-03 | |
| | (2.4664e-02) | (4.1483e-01) | (8.2960e-02) | (6.1734e-02) | |
| direct maximum likelihood based on $y_t$ | | | | | |
| AS2 | 5.8644e-02 | -1.1586e+00 | -2.9275e-01 | -5.5309e-01 | |
| | (4.2077e-02) | (1.0056e+00) | (7.0651e-02) | (1.8096e-01) | |

Table 2: Results from the Monte Carlo experiment for the CEV diffusion. All results are based on 1,000 simulated data sets using the parameters given in the "true parameters" row. Estimated bias (no parenthesis) is calculated as the average of the estimates minus the true parameters. Statistical standard errors are given in parenthesizes.

the diffusion $r_t$. This scenario is empirically infeasible in high frequencies because $r_t$ is not observed. In the third scenario, we estimate the model given by (21) based on noisy observations $y_t$ without accounting for the noise in model specification. This is mis-specified model which enables us to examine the impact of the microstructure noise on the estimated dynamics.

From Table 2 we see that the results from all the estimation procedures are fairly similar in scenario 1, suggesting that low-order expansions are adequate for this model. Also, very little loss of statistical accuracy incurs by adjusting for noise, when comparing scenario 1 to scenario 2. As in the previous example, the bias in the drift structure is quite large and points towards a faster mean reversion.

Comparing scenario 1 to scenario 3, we see that large biases occur in all the estimated parameters, even when the noise is very small. The most striking finding is that the average of the estimates of $\gamma$ is close to half of the true parameter. Interestingly, this is the case for the empirical application considered below.

| method | $\alpha$ | $\beta$ | $\sigma$ | $\rho$ | $a$ | log-likelihood |
|---|---|---|---|---|---|---|
| EUL | 7.8793e-02 | -1.6783e+00 | 2.7119e+00 | -7.6605e-01 | 1.3710e-02 | 6529.3 |
| | (1.9159e-02) | (1.1675e+00) | (2.4401e-01) | (4.6443e-02) | (4.3687e-02) | |
| | [4.1061e-04] | [1.3876e-02] | [6.3125e-03] | [9.4707e-04] | [2.5174e-04] | [1.1699e-01] |
| AS1 | 9.0781e-02 | -9.9313e-01 | 3.2343e+00 | -8.5153e-01 | -1.9521e-02 | 6544.2 |
| | (1.8587e-02) | (9.4814e-01) | (2.5800e-01) | (3.6043e-02) | (4.4929e-02) | |
| | [3.8931e-04] | [9.5332e-03] | [7.4181e-03] | [5.2175e-04] | [2.2926e-04] | [1.2579e-01] |
| AS2 | 9.4753e-02 | -1.1754e+00 | 3.2607e+00 | -8.4668e-01 | -1.8307e-02 | 6544.4 |
| | (1.9204e-02) | (1.0473e+00) | (2.6269e-01) | (3.6248e-02) | (4.4012e-02) | |
| | [5.0388e-04] | [1.1149e-02] | [8.7445e-03] | [5.2126e-04] | [2.1651e-04] | [1.2594e-01] |
| AS3 | 9.4624e-02 | -1.1833e+00 | 3.2542e+00 | -8.4561e-01 | -1.8226e-02 | 6544.4 |
| | (1.9189e-02) | (1.0458e+00) | (2.6225e-01) | (3.5774e-02) | (4.4464e-02) | |
| | [4.9865e-04] | [1.1067e-02] | [8.5580e-03] | [5.3222e-04] | [2.3434e-04] | [1.2573e-01] |

Table 3: Parameter estimates and log-likelihood values for the GARCH diffusion fitted to S&P500 data using the four different estimation procedures. The parameter estimates are taken as the mean over 100 random seeds in the importance sampler. Statistical standard errors taken from Table 1 with 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.

# 4 Empirical Applications

To illustrate the proposed procedure in practice, we fit the two diffusion models to real data.

## 4.1 GARCH diffusion

In the first empirical application, we fit the GARCH diffusion (17) to the daily Standard & Poor 500 sampled from January 3, 2003 to January 13, 2011. The time-series of daily (i.e $\Delta = 1/252$) log-returns consists of a total of $T = 2022$ observations.

Parameter estimates using the four different estimation procedures are presented in Table 3. The estimates are calculated as the mean across 100 estimates with different random seeds in the importance sampler. In addition to parameter estimates and statistical standard errors taken from Table 1, we also present standard errors induced by the EIS-Monte Carlo methods and also known as the Monte Carlo standard errors in the literature. As seen, the Monte Carlo standard errors are small comparing with statistical standard errors.

From Table 3, we see that the parameter estimates for three expansion-based SML methods are

fairly consistent. However, there are substantial differences between the EUL and the expansion-based SML methods in the estimates of the parameters and the log-likelihood values. These findings reinforce the simulation study. The SML estimate of $\rho$ is much larger in magnitude than what has been found in the literature using data from earlier periods. The estimated $\rho$ is around -0.85 in the AS1-AS3 while it is only -0.32 when Yu (2005) fitted the log-normal stochastic volatility model to S&P 500 data between 1980 and 1987. However, the estimated $\rho$ is similar to what has been found in Aït-Sahalia et al (2011) based on options data collected in a similar sample period. Moreover, we see that the difference in the parameter estimates between AS2 and AS3 is less than that between AS1 and AS2. This suggests that $K = 2$ is sufficient in the closed-form expansions.

In addition to estimating parameters, we have also considered some tests for a finite variance of importance weights in the SML procedures. Recall that to have $\sqrt{M}$ convergence and asymptotic normality of the integral estimate (5), a finite variance of the importance weights $w_j = \varphi(\lambda_{(T)}^{(j)})/m(\lambda_{(T)}^{(j)})$ is required. Recently, Koopman et al. (2009) proposed several tests for finite variance based on extreme value theory. In the present paper we apply some of their methods. Throughout this section, we consider the scaled importance weights $w' = \exp(\log w - 6545)$ for AS1-AS3 and $w' = \exp(\log w - 6526)$ for EUL 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 $N = 1000 \times M = 16,000$ importance weights obtained by evaluating the EIS procedures 1000 times on the real data at the parameter estimates obtained from AS2. The 100 largest scaled weights, along with a histogram of the scaled weights are presented in Figures 1 and 2 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. 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

$$f(u; \xi, b) = \frac{1}{b} \left( 1 - \xi \frac{u}{b} \right)^{-\frac{1}{\xi} - 1}, \ u \geq 0 \tag{27}$$
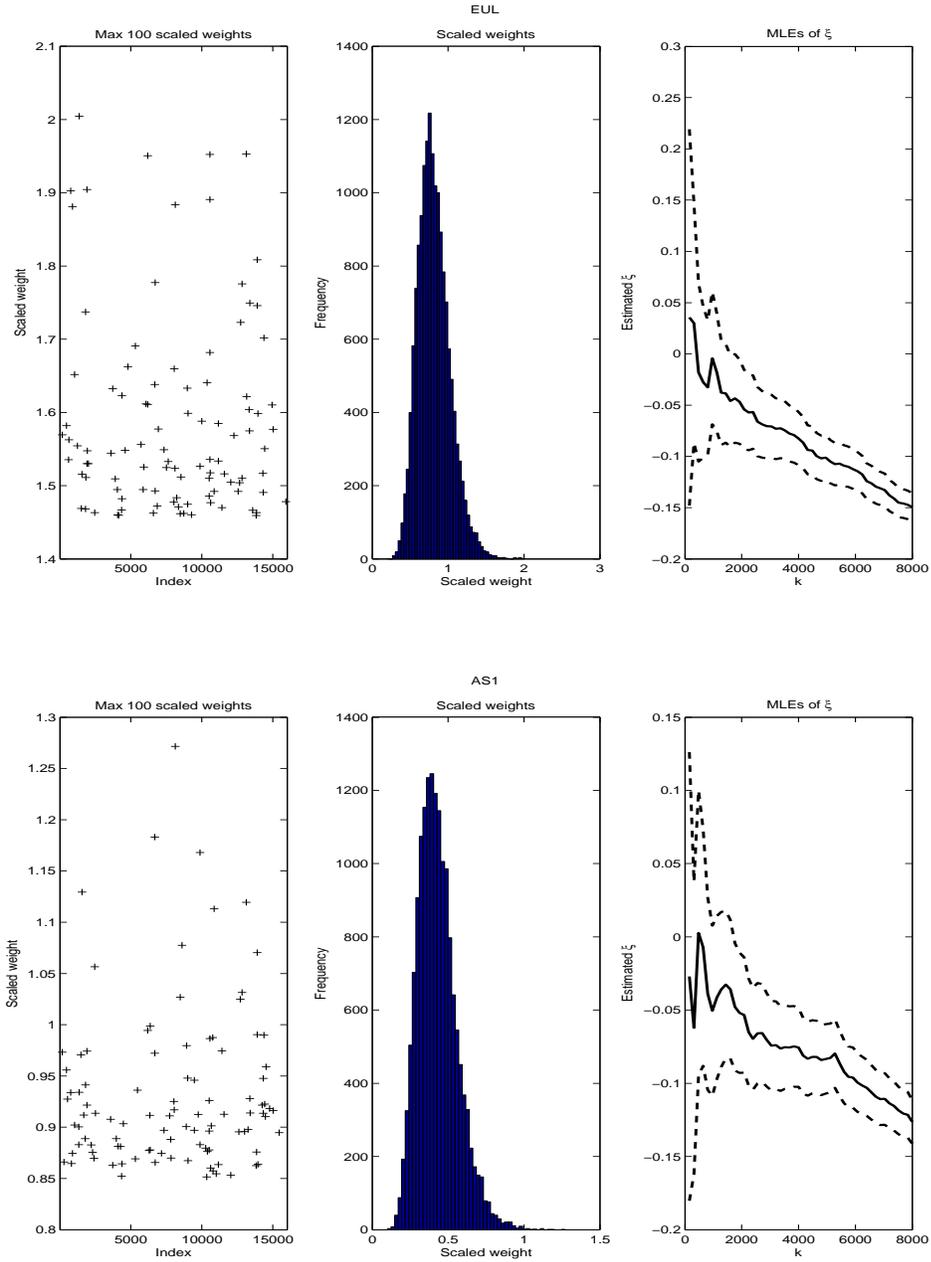
Figure 1: Finite variance diagnostics for EUL and AS1 for the GARCH diffusion model. The left hand side panels present the 100 largest scaled weights. The middle panels are histograms of all the scaled weights. The right hand side panel plots the maximum likelihood estimates of $\xi$ (solid) along with 95% confidence bands (dashed) for different values of the truncation parameter $k$.
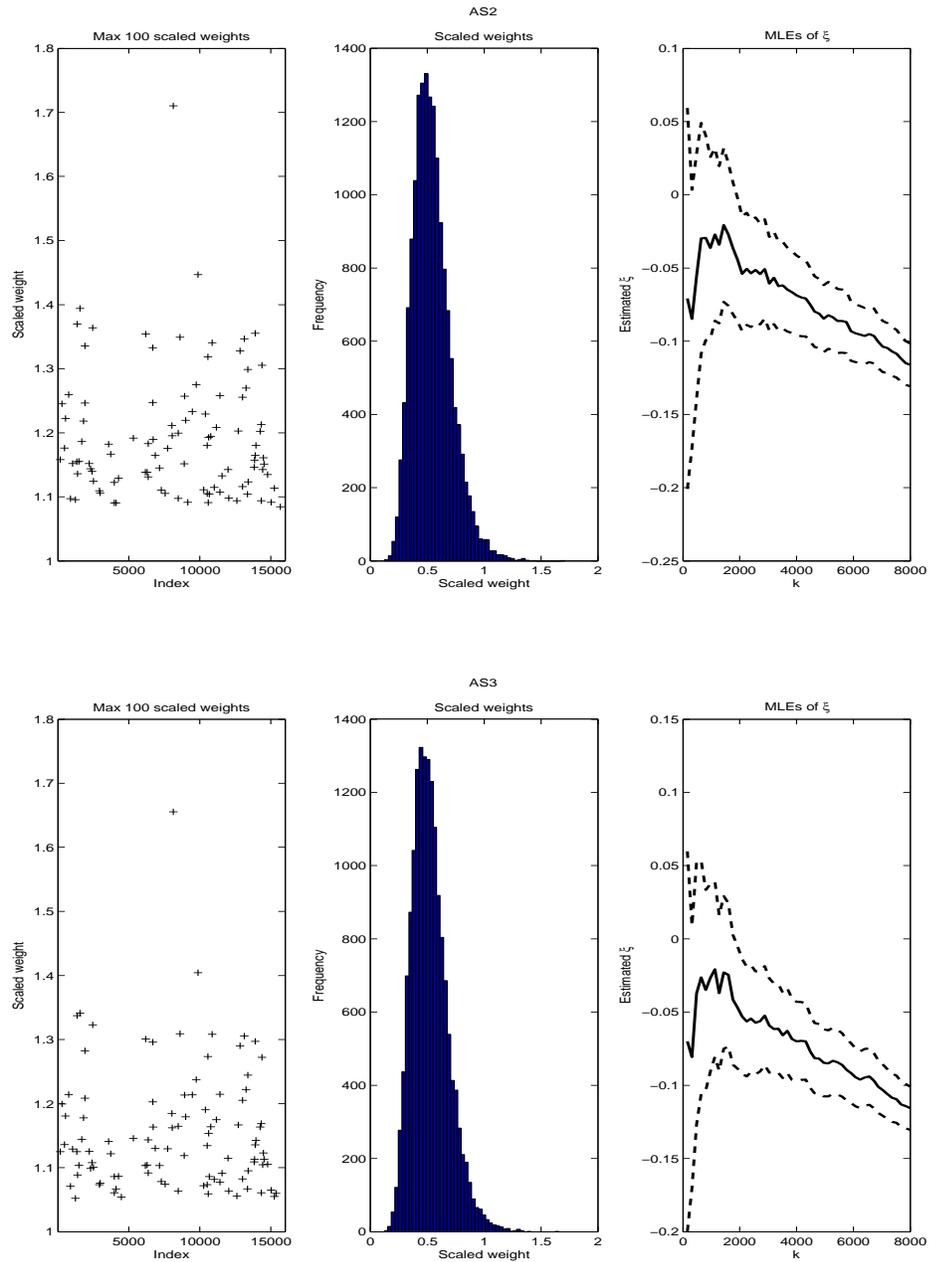
Figure 2: Finite variance diagnostics for AS2 and AS3 for the GARCH diffusion model. The left hand side panels present the 100 largest scaled weights. The middle panels are histograms of all the scaled weights. The right hand side panel plots the maximum likelihood estimates of $\xi$ (solid) along with 95% confidence bands (dashed) for different values of the truncation parameter $k$.

for which we fit to $\{u_i\}_{i=1}^k$ using two different methods. The parameter $\xi$ determines the tail thickness, and in particular $\xi < 1/2$ corresponds to a finite variance. For $\xi < 0$, the Generalized Pareto distribution has finite support and, hence, trivially finite variance. The parameter $b$ is a scale parameter, whose actual value is of little interest for our application. ML estimates of $\xi$ are plotted in the rightmost plots of Figures 1 and 2 along with 95% confidence bands for values of $k$ ranging from $[0.01N]$ to $[0.5N]$ where $[\cdot]$ denotes the integer part.[3] From the Figures, we see that the MLEs of $\xi$ stay consistently below $1/2$ for any reasonable truncation parameter $k$.

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

$$\xi_H = \frac{1}{k}\sum_{j=1}^k \log w'_{(N-j+1)} - \log w'_{(N-k)}, \tag{28}$$

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 test is asymptotically standard normal under the 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 $\xi$ and finite variance. The Monahan test statistics are smaller than $-5.8$ for $k = [2N^{1/3}]$ and $-8.1$ for $k = [4N^{1/3}]$ for all four SML methods considered. We again see strong evidence against the null hypothesis. All in all, the tests for finite variance of the importance weights conclusively points towards finite variance.

## 4.2 CEV diffusion observed with noise

In the second empirical application, we fit the CEV diffusion plus noise (23-24) to the Eurodollar interest rate data between 1983 and 1995 ($T = 3082$). The same data were used previously by Aït-Sahalia (1996).

A result obtained from the simulation study is that the form of the estimated volatility is highly altered when the noise is accounted for. This is also seen in the estimates presented in Table 4 for the real data. We see that the estimates for $\gamma$ is nearly doubled when the noise is accounted for, suggesting that the noise is highly material. Since the models are nested, it is also reasonable to compare the log-likelihoods, and we find a difference of around 83 which gives a strong rejection of

---

[3]ML estimates and confidence intervals where obtained using the `gpfit`-function in MATLAB.

| method | $\alpha$ | $\beta$ | $\sigma$ | $\gamma$ | $\sigma_y$ | log-likelihood |
|--------|----------|---------|----------|----------|------------|----------------|
| EUL | 9.7251e-03 | -1.6562e-01 | 4.2497e-01 | 1.2014e+00 | 5.2279e-04 | 16146.12 |
| | (2.4675e-02) | (4.1536e-01) | (1.4114e-01) | (1.0211e-01) | (3.0504e-05) | |
| | [3.9971e-06] | [6.4930e-05] | [3.2958e-04] | [3.1883e-04] | [1.0117e-07] | [9.7087e-02] |
| AS1 | 9.6332e-03 | -1.6396e-01 | 4.2470e-01 | 1.2011e+00 | 5.2281e-04 | 16146.11 |
| | (2.4746e-02) | (4.1672e-01) | (1.4037e-01) | (1.0206e-01) | (3.0515e-05) | |
| | [4.6649e-06] | [7.7315e-05] | [3.2980e-04] | [3.1930e-04] | [1.0122e-07] | [9.7043e-02] |
| AS2 | 9.6539e-03 | -1.6434e-01 | 4.2469e-01 | 1.2011e+00 | 5.2280e-04 | 16146.11 |
| | (2.4764e-02) | (4.1682e-01) | (1.4037e-01) | (1.0206e-01) | (3.0515e-05) | |
| | [4.6262e-06] | [7.5769e-05] | [3.2996e-04] | [3.1945e-04] | [1.0131e-07] | [9.7051e-02] |
| AS2 | 2.3050e-02 | -3.6689e-01 | 1.3655e-01 | 6.8788e-01 | 0 | 16063.34 |
| | (4.2077e-02) | (1.0056e+00) | (7.0651e-02) | (1.8096e-01) | | |

Table 4: Parameter estimates and log-likelihood values for the Eurodollar interest rate data using the three different estimation SML procedures, along with direct estimation with $\sigma_n = 0$. The parameter estimates are taken as the mean over 100 replications using different random seeds in the importance sampler. Statistical standard errors taken from Table 2 and are presented in parenthesizes. The estimates of the standard errors due to the EIS MC variation are included in brackets.

the model without noise.

The estimates for AS1 and AS2 are quite similar, suggesting that AS1 should be sufficient for these ranges of parameters and $\Delta$. Looking at the values for the MC variation, we see that the algorithm performs very well, even with the modest $M = 16$ simulation paths.

As for the previous model, we also perform some diagnostics on the variance of the importance weights. The scaled weights are given as $w' = \exp(\log w - 16146)$, and the tests are based on $N = 1000 \times M = 16,000$ draws obtained based on the real data and parameters equal to those in Table 4. Diagnostics plots are given in Figures 3 and 4 and show no signs of infinite variances, with estimated $\xi$ safely below 1/2. For the Monahan tests we get $-6.3$ for the test statistic when $k = [2N^{1/3}]$ and $-8.6$ when $k = [4N^{1/3}]$ for all three methods considered. Thus there is no evidence of infinite variance from any of the tests.

# 5    Concluding Remarks

This paper extends the closed-form method of Aït-Sahalia (2008) to estimate the diffusion models with latent variables. The method synthesizes the closed form method and the efficient importance sampler
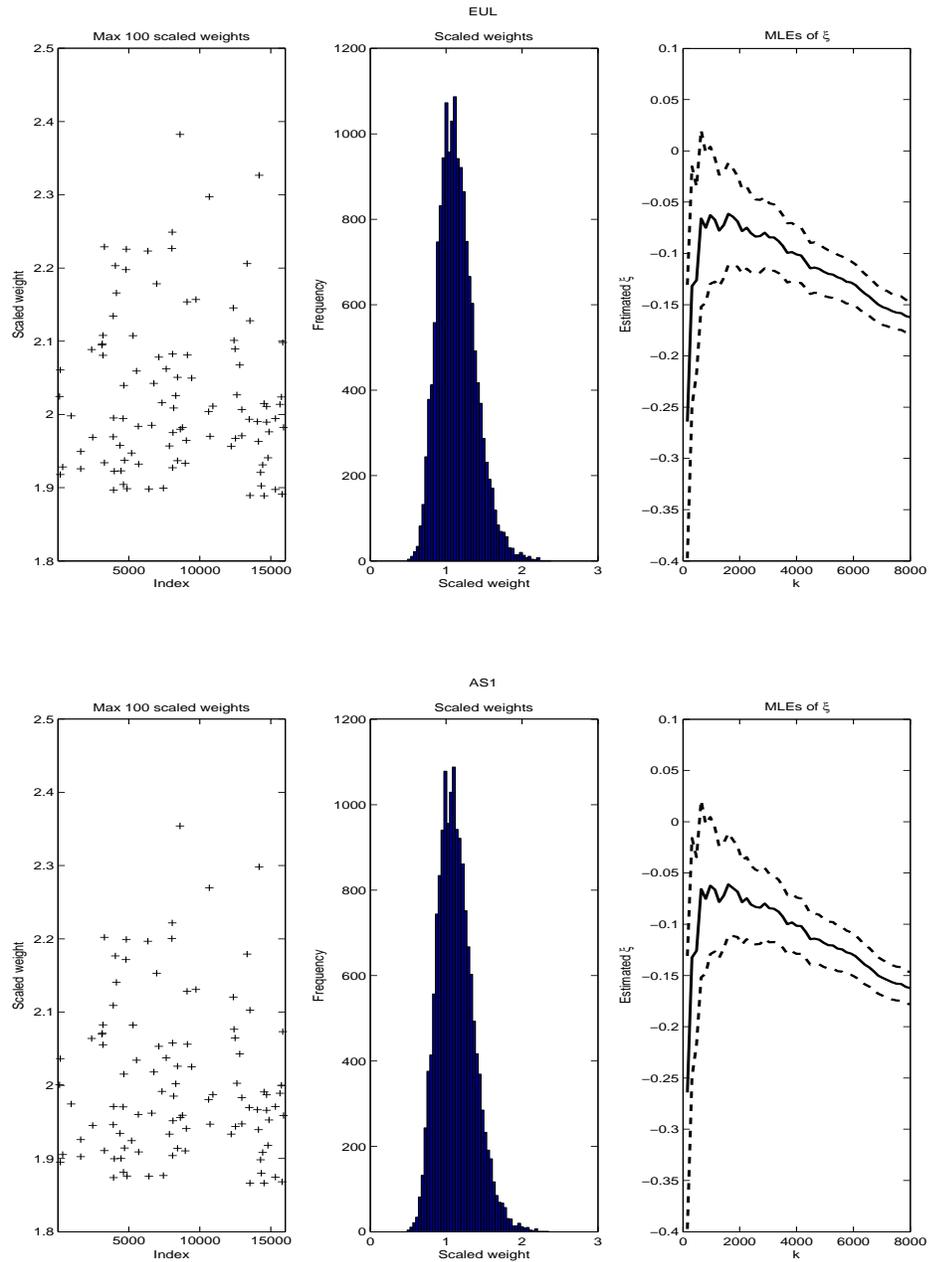
Figure 3: Finite variance diagnostics for EUL and AS1 for the CEV model. The left hand side panels present the 100 largest scaled weights. The middle panels are histograms of all the scaled weights. The right hand side panel plots the maximum likelihood estimates of $\xi$ (solid) along with 95% confidence bands (dashed) for different values of the truncation parameter $k$.

of Richard and Zhang (2007). It does not require any infill observations and hence is computationally appealing. The method was illustrated using two classes of models. In the first class, the state variable is only partially observed. This class includes the stochastic volatility models and the stochastic mean models. It also includes the combinations of the two specifications as advocated by Duffie and Kan (1996) and Dai and Singleton (2000). The second class assumes that the variable, which follows a diffusion, is observed with a noise. In the simulation study and the empirical study, the method was used to estimate the GARCH diffusion and the CEV diffusion with noise. The Monte Carlo study shows that the method works well in finite sample. For the progressively precise TPD-approximations, we see that there is a decreasing difference in the resulting SML estimates, suggesting that arbitrarily accurate approximations to the exact continuous time likelihood based on discrete data can be produced. Of course, there is a trade-off between decreasing the discretization bias and containing the computational cost. As a reference, the AS2 expansion for the GARCH diffusion requires about 200 lines of machine generated FORTRAN90 code to be evaluated, whereas the corresponding figure for AS3 is about 700. The empirical applications illustrate usefulness of the method and there is no evidence of infinite variance in the importance sampler.

In the literature, Durham and Gallant (2002) has introduced a SML method to estimate continuous time stochastic volatility models using spot prices only. While both their method and ours are based on simulations, there are several key differences between the two methods. First, while our method employs the closed-form method to control the discretization bias, the method of Durham and Gallant (2002) uses infill observations which are in turn integrated out using importance sampler. When there is no latent state in the diffusion, however, Aït-Sahalia (2002a) showed that the closed-form method is not only more accurate but also computationally faster than the infill method. Second, Durham and Gallant (2002) only applied the SML method to estimate continuous time stochastic volatility models. In this paper, we implement our method in two classes of models.

To control the discretization bias in latent diffusions, in addition to the SML approach of Durham and Gallant (2002), Bayesian Markov Chain Monte Carlo (MCMC) may also be used. See Eraker (2001) and Stramer et al (2010) for the studies. The later reference is especially relevant to the present paper because it uses the closed form method in connection to MCMC in the context of latent factors to alleviate the need for infill simulations. However, their method is Bayesian and the sampling algorithms are unrelated.

There should be scopes for applying the current methodology to a broader class of models, including
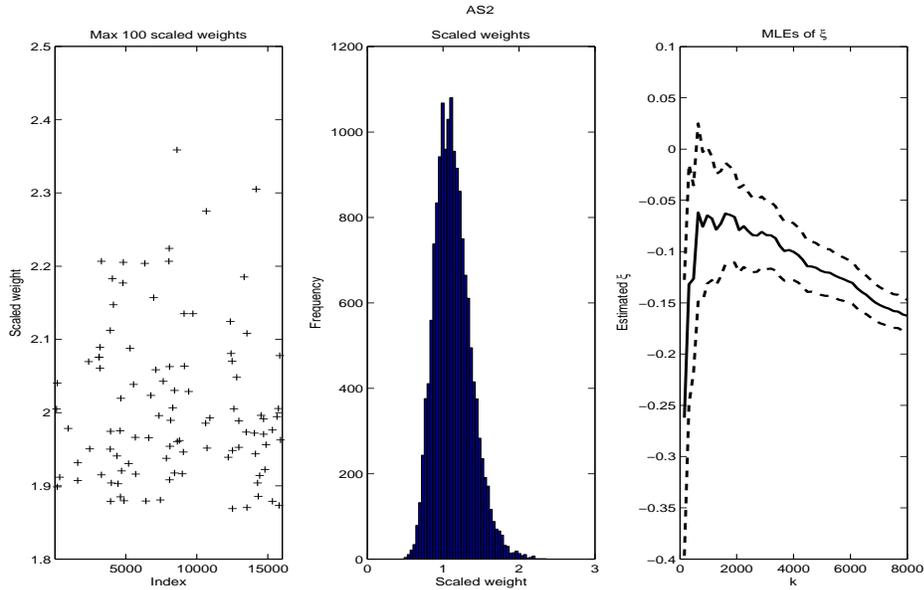
Figure 4: Finite variance diagnostics for AS2 for the CEV model. The left hand side panels present the 100 largest scaled weights. The middle panels are histograms of all the scaled weights. The right hand side panel plots the maximum likelihood estimates of $\xi$ (solid) along with 95% confidence bands (dashed) for different values of the truncation parameter $k$.

multiple latent state variables. Once the EIS framework is implemented, it is relatively easy to adapt to new models using symbolic manipulation software to generate code for the model specific $\mu_t$, $\Sigma_t$ and $Q_t$. However, the sampling interval $\Delta$ 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 the above proposed methodology should produce precise results for sufficiently small $\Delta$. However, this limit argument may not be of practical interest as data may be available only for large $\Delta$. 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.

Another possible direction for future research may be to employ the EIS importance density as the proposal for updating blocks of latent states in an Independent Metropolis-Hastings algorithm. This approach may be applied either in a Gibbs sampler-based MCMC algorithm for estimating parameters, or for providing smoothed estimates of latent states. See Liesenfeld and Richard (2006) for a discussion of MCMC based on the EIS.

Finally, one may wish to allow for jumps either in the volatility process 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 using the Mixture EIS framework of Kleppe and Liesenfeld (2011).

# References

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies 9* (2), 385–426.

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 and 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., Fan, J., Y. Li (2011). The leverage effect puzzle: Disentangling sources of bias in high frequency inference. work in progress.

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

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies 18* (2), 351–416.

Andersen, T. G. and J. Lund (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics 77* (2), 343 – 377.

Balduzzi, P., S. R. Das, and S. Foresi (1998). The central tendency: A second factor in bond yields. *The Review of Economics and Statistics 80* (1), pp. 62–72.

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.

Bergstrom, A. (1984). Continuous time stochastic models and issues of aggregation over time. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2 of *Handbook of Econometrics*, Chapter 20, pp. 1145–1212. Elsevier.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy 81*(3), pp. 637–654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance 47*(3), pp. 1209–1227.

Dai, Q. and K. J. Singleton (2000). Specification analysis of affine term structure models. *The Journal of Finance 55*(5), 1943–1978.

Duan, J. and A. Fulop (2007). How frequently does the stock price jump? - an analysis of high-frequency data with microstructure noises. working paper.

Duan, J. and C.-Y. Yeh (2011). Price and volatility dynamics implied by the vix term structure. working paper.

Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance 6*(4), 379–406.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica 68*(6), 1343–1376.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika 84*(3), 669–684.

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

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

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.

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

Kleppe, T. S. and R. Liesenfeld (2011). Effcient high-dimensional importance sampling in mixture frameworks. working paper.

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

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.

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.

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. 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.

Phillips, P. C. B. and J. Yu (2009). Simulation-based estimation of contingent-claims prices. *Review of Financial Studies 22*(9), 3669–3705.

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

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

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

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.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), pp. 1394–1411.