Journal of Econometrics 180 (2014) 73–80

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Maximum likelihood estimation of partially observed diffusion models☆

Tore Selland Kleppe [a], Jun Yu [b,*], Hans J. Skaug [a]

[a] University of Bergen, Norway

[b] Singapore Management University, Singapore

ARTICLE INFO

Article history:
Received 13 December 2012
Received in revised form
21 November 2013
Accepted 6 February 2014
Available online 26 February 2014

JEL classification:
C11
C15
G12

Keywords:
Closed-form approximation
Diffusion model
Efficient importance sampler

ABSTRACT

This paper develops a maximum likelihood (ML) method to estimate partially observed diffusion models based on data sampled at discrete times. The method combines two techniques recently proposed in the literature in two separate steps. In the first step, the closed form approach of Aït-Sahalia (2008) is used to obtain a highly accurate approximation to the joint transition probability density of the latent and the observed states. In the second step, the efficient importance sampling technique of Richard and Zhang (2007) is used to integrate out the latent states, thereby yielding the likelihood function. Using both simulated and real data, we show that the proposed ML method works better than alternative methods. The new method does not require the underlying diffusion to have an affine structure and does not involve infill simulations. Therefore, the method has a wide range of applicability and its computational cost is moderate.

1. Introduction

Continuous time diffusion models have long proven useful in economics and finance. For example, they provide a convenient mathematical framework for the development of financial economics and option pricing theory (Black and Scholes, 1973; Heston, 1993; Duffie and Kan, 1996). The separate treatment of stock and flow variables in macroeconomics (Bergstrom, 1984) and the formulation of continuous time inter-temporal optimization models (Turnovsky, 2000) represent additional usage of diffusion models in economics. Not surprisingly, fitting diffusion models based on data sampled at discrete times has received a great deal of attention in econometrics.

The case when some of the state variables are latent is often encountered in practical applications. One example of such partially observed diffusion models is the continuous time stochastic volatility models with the volatility being the latent state; see Hull and White (1987) and Heston (1993). Another example is the continuous time stochastic mean model of Balduzzi et al. (1998), in which the mean is a latent state. Obviously, the combination of a latent volatility and a latent mean also makes a partially observed diffusion model. This class of models has found a wide range of applications in the term structure literature Duffie and Kan (1996), Dai and Singleton (2000) and in the option pricing literature Duffie et al. (2000).

It has been argued, on the basis of asymptotic properties, that the preferred choice of estimation method for diffusion models should be maximum likelihood (ML) (Aït-Sahalia, 2002; Durham and Gallant, 2002). The ML estimation of partially observed diffusions necessitates the computation of the joint transition probability density (TPD) of the observed and the latent state variables as well as the marginalization of the latent variable from the joint density.

When the transition density of the state variables does not have a closed-form expression, it has to be approximated. Many approximation methods have been proposed in the literature which is reviewed in Jensen and Poulsen (2002), Hurn et al. (2007) and Phillips and Yu (2009). Broadly speaking, the methods can be classified

☆ Kleppe gratefully acknowledges the hospitality during his research visit to School of Economics and Sim Kee Boon Institute for Financial Economics at Singapore Management University. Yu thanks the Singapore Ministry of Education for Academic Research Fund under grant number MOE2011-T2-2-096. We are grateful to the co-Editor Yacine Aït-Sahalia and two anonymous referees for constructive comments that have greatly improved the paper. Moreover, we are indebted to Kenneth Lindsay for advise on implementing the method of Hurn et al. (2013) and to Roman Liesenfeld and Atle Øglend for invaluable comments.

* Correspondence to: School of Economics, 90 Stamford Road, Singapore 178903, Singapore. Tel.: +65 68280858; fax: +65 68280833.

E-mail address: yujun@smu.edu.sg (J. Yu).

http://dx.doi.org/10.1016/j.jeconom.2014.02.002
0304-4076/

into two classes. In the first class, the diffusion model is approximated by a discrete time model whose TPD is available in closed form. The well-known Euler–Maruyama (EM) approximation and the approximations of Bergstrom (1966), Nowman (1997), Milstein (1979) and Shoji and Ozaki (1998) all belong to this class. These methods in general lead to a bias in the calculation of the likelihood function that does not vanish asymptotically (see Aït-Sahalia, 2002 and references therein). We shall refer to such a bias as the discretization bias. In the second class, the TPD of the diffusion model is approximated directly, including the infill approximations and the closed-form approximations. To the best of our knowledge, the first contribution on the topic of infill approximations is in Pedersen (1995).[1] The closed-form approximation techniques include the Hermite approximation of Aït-Sahalia (2002), the polynomial approximation of Aït-Sahalia (2008), and the saddlepoint approximation of Aït-Sahalia and Yu (2006). Both the infill techniques and the closed-form techniques can reduce the discretization bias. Compared to the infill techniques, the closed-form techniques are computationally much cheaper and the approximation errors are smaller (Aït-Sahalia, 2002).

When the diffusion specifying the observed and latent states is not a linear Gaussian process, the marginalization of the latent state variable cannot be achieved analytically. Consequently, various importance sampling techniques have been proposed to integrate out the latent state variable from the joint density via simulations. The importance sampler of Shephard and Pitt (1997) and Durbin and Koopman (1997) (see also Sandmann and Koopman, 1998, Durham, 2006, and Skaug and Yu, forthcoming) used a global, multivariate Gaussian approximation to the joint density as the importance density, whereas Richard and Zhang (2007) (see also Liesenfeld and Richard, 2003, 2006) used a product of univariate Gaussian approximations to the conditional TPD as the importance density. While both methods work well for estimating discrete time stochastic volatility models, as shown in Lee and Koopman (2004), the importance sampler of Richard and Zhang (2007) tends to be more well-behaved than that of Shephard and Pitt (1997) and Durbin and Koopman (1997).

For most partially observed diffusion models, almost all of the afore-mentioned ML methods are not directly applicable. Arguably, the most widely used ML approach to estimating partially observed diffusion models is to use the EM approximation to discretize the diffusion models and then use an importance sampler to marginalize out the unobserved latent state variables. This approach naturally leads to the discretization bias when a sampling interval is fixed.

Several approaches have been proposed in the literature to provide ML estimation of partially observed diffusion models with the discretization bias being controlled. Bates (2006) proposed a frequency domain filtering method to compute the likelihood function of latent affine diffusion models via the conditional characteristic functions. This technique is not feasible for non-affine models for which the conditional characteristic functions are not available in closed form.

Aït-Sahalia and Kimmel (2007) proposed to approximate the volatility (latent state) using the implied volatility computed from the underlying options. Consequently, no state variable is latent in the continuous time stochastic volatility model and the closed form approximation of Aït-Sahalia (2008) is directly applicable. However, it is well-known that option prices are derived from the risk neutral measure (Heston, 1993). As a result, using data from both the spot market and the options market jointly, one can simultaneously learn about both the physical measure and the risk-neutral measure. Naturally, this benefit is achieved at a cost. To connect the physical measure to the risk-neutral measure, 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 a stochastic mean, it is not clear how to extract latent variables from derivative prices. Perhaps most importantly, when the latent volatility is approximated by the implied volatility, approximation errors are introduced. How these errors influence the estimated price dynamics remains to be answered.

More recently, a quasi-ML (QML) approach was proposed by Hurn et al. (2013) to estimate partially observed diffusion models. However, the discretization bias cannot be completely removed by this method. While the infill technique combined with the importance sampler via the global approximation has been introduced to provide the ML estimation (see, for example, Durham and Gallant, 2002), it is computationally very expensive.

In this paper, we introduce a new ML method to estimate partially observed diffusion models. Our ML method combines the closed-form approach of Aït-Sahalia (2008) for approximating the joint TPD of the observed and the latent state variables and the efficient importance sampler (EIS) of Richard and Zhang (2007) for integrating out latent states from the joint density. Our method inherits two nice features of the closed-form approximation techniques of Aït-Sahalia (2002) and Aït-Sahalia (2008). First, it can practically remove the discretization bias and, hence, leads to more accurate likelihood values than the QML and the EM methods. Second, it is computationally inexpensive, especially relative to the infill methods. Moreover, our method is very general in the sense that only weak assumptions regarding the structure of the underlying diffusion must be made. Most notably, an affine structure does not need to be assumed.

The paper is organized as follows. Section 2 proposes the new estimation method. Section 3 illustrates our method using the GARCH diffusion model of Nelson (1990) and investigates the performance of the proposed method relative to alternative methods, including the EM method and the QML of Hurn et al. (2013), using simulated data. In Section 4, we fit the GARCH diffusion to real data. Finally, Section 5 concludes and outlines some further applications and implications of the approach.

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_τ and $a(X_\tau;\theta)$ are q-vectors, and $b(X_\tau;\theta)$ is a $q \times q$ matrix, with B_τ being a q-dimensional uncorrelated Brownian motion. θ is the vector of parameters to be estimated. We assume that (1) admits a unique solution and that $b(X_\tau;\theta)b(X_\tau;\theta)'$ is positive definite for all admissible values of X_τ and θ. Moreover, we assume that $a(\cdot;\theta)$ and $b(\cdot;\theta)$ are infinitely differentiable. Let $x_t = X_{t\Delta}$ $(t = 1,\ldots,T)$ be the value of X_τ sampled at frequency $1/\Delta$ and $\mathbf{x} = (x_1,\ldots,x_T)$ be the collection of such values.

Due to the Markovian property of X_τ, the joint probability density function (PDF) of $\mathbf{x}$ may be written as

$$p(\mathbf{x};\theta) = p(x_1;\theta)\prod_{t=2}^{T} p_t(x_t|x_{t-1};\theta). \tag{2}$$

Here $p_t(x_t|x_{t-1};\theta)$ is the TPD associated with (1) and $p(x_1;\theta)$ is the, possibly degenerate, density of the initial state x_1. We assume that the first q_y elements of x_t, denoted by y_t, are observed at frequency $1/\Delta$ and $t = 1,\ldots,T$. The remaining $q_z = q - q_y$ elements of

[1] Elerian et al. (2001) and Eraker (2001) use the infill methods to conduct Bayesian inference of continuous time models.

x_t, denoted by z_t, are assumed to be latent.[2] Thus, $x_t = [y_t' \; z_t']'$, $p_t(x_t|x_{t-1};\theta) = p_t(y_t, z_t|y_{t-1}, z_{t-1};\theta)$ and $p(\mathbf{x};\theta) = p(\mathbf{y},\mathbf{z};\theta)$ with $\mathbf{y} = (y_1,\ldots,y_T), \mathbf{z} = (z_1,\ldots,z_T)$. Without loss of generality, we focus on the case where $q_z = 1$ in this paper. Models with multiple latent states can be estimated in the same spirit.

The likelihood function based on observed data is obtained by marginalizing out the latent variable from the joint density, namely,

$$L(\theta|\mathbf{y}) = p(\mathbf{y};\theta) = \int p(\mathbf{x};\theta)d\mathbf{z} = \int p_1(x_1;\theta)\prod_{t=2}^{T} p_t(x_t|x_{t-1};\theta)d\mathbf{z}. \quad (3)$$

Denote the log-likelihood function by $l(\theta|\mathbf{y})$. From now on, the dependence on the parameter vector θ of all quantities is made implicit in the notation.

2.2. Transition density function approximations

As the TPD of (1) has a closed-form expression only for a few special cases, approximations are inevitably unavoidable in general. Denote the generic TPD approximation by $\bar{p}_t = \bar{p}_t(x_t|x_{t-1})$. In the present paper, we focus primarily on the closed-form technique of Aït-Sahalia (2008), but also consider two other alternatives for references. We choose these three classes of approximate TPDs based on two criteria. First, they should have a closed-form expression and be cheap to compute. Consequently, computationally expensive methods, such as the infill techniques, are not considered (see Aït-Sahalia, 2002, Fig. 1). Second, the approximate TPDs should be applicable to general multivariate diffusions (1). In particular, they should not be restricted to a narrow range of functional form for a and b, such as the affine diffusions or the reducible diffusions (Aït-Sahalia, 2008).

2.2.1. Euler Maruyama approximation

The first approximate TPD we consider is the Euler–Maruyama (EM) approximation, defined by

$$\bar{p}_t^{(\mathrm{EM})}(x_t|x_{t-1}) = \mathcal{N}\left(x_t; x_{t-1} + \Delta a(x_{t-1}), \Delta b(x_{t-1})b(x_{t-1})'\right), \quad (4)$$

where $\mathcal{N}(x; m, \Sigma)$ is the PDF of $N(m, \Sigma)$ evaluated at x. The EM-TPD is easy to implement and fast to evaluate. However, for a fixed sampling interval Δ, the EM approximation has fixed accuracy and, hence, may lead to an unacceptable discretization bias.

2.2.2. The approximation of Hurn et al. (2013)

The second approximate TPD we consider is due to Hurn et al. (2013), which we denote by HLM. The HLM-TPD takes the form of a multivariate Gaussian, namely

$$\bar{p}_t^{(\mathrm{HLM})}(x_t|x_{t-1}) = \mathcal{N}\left(x_t; M(x_{t-1},\Delta), S(x_{t-1},\Delta)\right)$$

where $M(x_{t-1},\Delta)$ and $S(x_{t-1},\Delta)$ approximate $E(x_t|x_{t-1})$ and $\mathrm{Var}(x_t|x_{t-1})$, respectively, and the elements of $M(x_{t-1},\Delta)$ and $S(x_{t-1},\Delta)$ are obtained as a solution of a system of ordinary differential equations (see Hurn et al., 2013, for details). For a general diffusion model (1), this system of differential equations may involve expectations of non-linear functions of Gaussian random variables. Hurn et al. (2013) suggest to approximate these expectations using low-order Gaussian quadrature. Moreover, the system of differential equations may have to be solved numerically. Owing to the Gaussian structure of $\bar{p}_t^{(\mathrm{HLM})}(x_t|x_{t-1})$, the HLM-TPDs have fixed accuracy for a fixed sampling interval Δ. Nevertheless, the HLM-TPDs are relatively easy to implement and fast to evaluate relative to infill methods.

2.2.3. Aït-Sahalia (2008) expansions

The third approximate TPD we consider is the closed-form expansion of Aït-Sahalia (2008). Though more cumbersome to derive than the former two methods, the Aït-Sahalia expansions are attractive in that they have computationally tractable closed-form expressions while retaining adjustable accuracy for fixed Δ. This nice feature enables us to study the incurred discretization bias by considering a sequence of increasingly accurate Aït-Sahalia expansions.

The Aït-Sahalia expansion of order K, denoted by ASK, has the form

$$\log \bar{p}_t^{(\mathrm{ASK})}(x_t|x_{t-1}) \equiv -\frac{q}{2}\log(2\pi\Delta) - D_v(x_t) + \frac{C_{-1}^{j_{-1}}(x_t|x_{t-1})}{\Delta} + \sum_{k=0}^{K} C_k^{j_k}(x_t|x_{t-1})\frac{\Delta^k}{k!}, \quad (5)$$

where

$$D_v(x) = \frac{1}{2}\log(\mathrm{Det}(b(x)b(x)')). \quad (6)$$

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

$$C_k^{j_k}(r|s) = \sum_{|i|\le j_k} c_i^{(k)}(r_1 - s_1)^{i_1}(r_2 - s_2)^{i_2}\cdots(r_q - s_q)^{i_q}, \quad (7)$$

where $i = (i_1,\ldots,i_m)$ is a multi-index with $|i| = i_1 + \cdots + i_m$ and $j_k = 2(K-k)$.[3] The form of the coefficients $c_i^{(k)}$ is found by solving both the forward and the backward Kolmogorov partial differential equations to the appropriate orders in Δ using the algorithms outlined in Aït-Sahalia (2008). The model-specific expressions $C_k^{j_k}$ are in general complicated, and we obtain 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 do not lead to proper densities as they do not integrate exactly to one in general. To make a proper density, one may normalize the approximate TPDs. However, in our experience the expansions are very accurate and a re-normalization is unnecessary. Moreover, since the terms $C_k^{j_k}$ in (5) are polynomials, there is no guarantee that $\bar{p}_t^{(\mathrm{ASK})}(x_t|x_{t-1})$ stays bounded in the very far tails. However, we do not encounter this problem in practice.

2.3. Efficient importance sampling

The joint density of all the state variables (both the latent and the observed) can be obtained from the joint TPDs after they are approximated. To obtain the likelihood function of the model, however, the latent variables have to be integrated out. Various Monte Carlo methods, all based on the importance sampler, have been proposed to integrate out the latent state variables from the joint density. In the present paper, we propose to use the EIS method of Richard and Zhang (2007) because, unlike the Laplace importance sampler of Shephard and Pitt (1997) and Durbin and Koopman (1997), it does not rely on a global near-Gaussian kernel assumption on the integrand $\mathbf{z} \mapsto \tilde{p}(\mathbf{y},\mathbf{z})$.

[2] Note that the assumption that the first q_y elements of x_t are observed comes without loss of generality as the ordering of the elements of x_t is arbitrary.

[3] 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.

Before fixing the idea, let us introduce some new notations. First, we introduce the approximate joint density $\bar{p}(\mathbf{y},\mathbf{z}) = \bar{p}(\mathbf{x})$, which is obtained by plugging in the approximate TPDs into Eq. (2):

$$\bar{p}(\mathbf{x}) = \bar{p}(\mathbf{y},\mathbf{z}) = p(y_1,z_1)\prod_{t=2}^{T}\bar{p}_t(y_t,z_t|y_{t-1},z_{t-1}). \tag{8}$$

Second, the approximate likelihood function based on $\bar{p}(\mathbf{y},\mathbf{z})$ is denoted by $\bar{L}(\theta;\mathbf{y}) = \int \bar{p}(\mathbf{y},\mathbf{z})d\mathbf{z}$. This is the target for the Monte Carlo integration.

Next, we introduce the Gaussian approximate conditional kernel $k_t,\ t = 1,\ldots,T$. When $q_y = 1$, $\log k_t$ is a second-order Taylor series expansion of $\log \bar{p}_t$ in z_t around $\lambda_t(z_{t-1})$, namely

$$\log k_t(z_t|z_{t-1}) = \sum_{r=0}^{2}\frac{F_t^{(r)}(z_{t-1})}{r!}(z_t-\lambda_t(z_{t-1}))^r, \tag{9}$$

where

$$F_t^{(r)}(z_{t-1}) = \frac{\partial^r}{\partial z_t^r}\log\bar{p}_t(y_t,z_t|y_{t-1},z_{t-1})|_{z_t=\lambda_t(z_{t-1})},$$
$$r = 0,1,2, \tag{10}$$
$$\lambda_t(z_{t-1}) = E_{\bar{p}_t^{(EM)}}[z_t|y_t,y_{t-1},z_{t-1}], \tag{11}$$

with $\bar{p}_t^{(\mathrm{EM})}$ defined in Eq. (4), and the dependence on y_t, y_{t-1} made implicit in the notation. The expansion point $\lambda_t(z_{t-1})$, which has closed form, is intended to be in a high probability region of $z_t|y_t,y_{t-1},z_{t-1}$ under $\bar{p}_t$. Consequently, the normalized version of k_t, namely $k_t(z_t|z_{t-1})/\int k(z_t|z_{t-1})dz_t$, may be regarded as a Gaussian approximation to the density of $z_t|y_t,y_{t-1},z_{t-1}$ under $\bar{p}_t$. For $t = 1$, let $k_1(z_1)$ denote some unspecified Gaussian approximation to $p(z_1|y_1)$, with associated $F_1^{(r)},\ r = 0,1,2$, so that $\log k_1(z_1) = \sum_{r=0}^{2}F_1^{(r)}(z_1-\lambda_1)^r/r!$ for an arbitrary expansion point λ_1.

The final piece of notation that we introduce is the reminder term associated with the second-order Taylor expansion defining k_t, namely

$$Q_t(z_t|z_{t-1}) = \log\bar{p}_t(y_t,z_t|y_{t-1},z_{t-1}) - \log k_t(z_t|z_{t-1}),$$
$$t = 2,\ldots,T, \tag{12}$$
$$Q_1(z_1) = \log p(y_1,z_1) - \log k_1(z_1). \tag{13}$$

For the EM and the HLM methods, which are both Gaussian, k_t is exact (when keeping y_t, y_{t-1}, z_{t-1} fixed) in that $\bar{p}_t(y_t, z_t|y_{t-1},z_{t-1}) = k_t(z_t|z_{t-1})$ and, therefore, $Q_t(z_t|z_{t-1}) = 0$. For the ASK-TPDs, Q_t is generally non-zero as the polynomials $Q_k^{j_k}$ defined in (7) are of order greater than 2 for $K > 1$. Moreover, D_v defined in (6) is typically a non-linear function.

In general importance sampling techniques rely on introducing an importance density $m(\mathbf{z})$ and approximating the target integral, in our case $\bar{L}(\theta|\mathbf{y})$ defined in Eq. (8), via an application of the law of large numbers as

$$\bar{L}(\theta|\mathbf{y}) \approx \frac{1}{M}\sum_{j=1}^{M}\frac{\bar{p}(\mathbf{y},\mathbf{z}^{(j)})}{m(\mathbf{z}^{(j)})}, \quad \text{where } \mathbf{z}^{(j)} \sim \text{i.i.d. } m(\mathbf{z}),$$
$$j = 1,\ldots,M. \tag{14}$$

The fractions $w^{(j)} = \bar{p}(\mathbf{y},\mathbf{z}^{(j)})/m(\mathbf{z}^{(j)}),\ j = 1,\ldots,M$, are known as the importance weights (Geweke, 1989). The EIS algorithm aims at choosing $m(\mathbf{z})$ within a particular class of importance densities so that the variance of the right hand side of (14) is minimized for a fixed number of draws M. The class of importance densities considered here has a Markovian structure

$$m(\mathbf{z};\mathbf{A}) = m_1(z_1;\mathbf{A}_1)\prod_{t=2}^{T}m_t(z_t|z_{t-1};\mathbf{A}_t),$$
$$\mathbf{A} = (\mathbf{A}_1,\ldots,\mathbf{A}_T), \tag{15}$$

where $\{m_t\}_{t=1}^{T}$ are all univariate Gaussian densities. Moreover, the mean and the variance of each m_t depend on an auxiliary parameter $\mathbf{A}_t = (A_{t,1},A_{t,2})',\ t = 1,\ldots,T$, that is chosen using the EIS algorithm. Due to the Markovian structure of (15), generating random draws $\mathbf{z}^{(j)} \sim m(\mathbf{z};\mathbf{A})$ is conceptually simple and computationally tractable. The specific forms of $m_t,\ t = 1,\ldots,T$, have the interpretation of exponentially tilted and normalized versions of the Gaussian approximate conditional kernels k_t, and are given by

$$m_t(z_t|z_{t-1};\mathbf{A}_t) = \frac{k_t(z_t|z_{t-1})\psi_t(z_t;\mathbf{A}_t)}{\chi_t(z_{t-1};\mathbf{A}_t)},$$
$$m_1(z_1;\mathbf{A}_1) = \frac{k_1(z_1)\psi_1(z_1;\mathbf{A}_1)}{\chi_1(\mathbf{A}_1)},$$

where

$$\log\psi_t(z_t;\mathbf{A}_t) = A_{t,1}z_t + A_{t,2}z_t^2, \quad t = 1,\ldots,T,$$
$$\chi_t(z_{t-1};\mathbf{A}_t) = \int k_t(z_t|z_{t-1})\psi_t(z_t;\mathbf{A}_t)dz_t, \quad t = 2,\ldots,T,$$
$$\chi_1(\mathbf{A}_1) = \int k_1(z_1)\psi_1(z_1;\mathbf{A}_1)dz_1.$$

Since both $\log k_t$ and $\log\psi_t$ are the second-order polynomials in z_t, it is clear that m_t is a Gaussian density with mean and variance given by

$$\mu_t = -\frac{F_t^{(1)} - F_t^{(2)}\lambda_t + A_{t,1}}{F_t^{(2)} + 2A_{t,2}}, \qquad \Sigma_t = -\frac{1}{F_t^{(2)} + 2A_{t,2}},$$

provided that $A_{t,2} < F^{(2)}/2$.

With the above notation, we may rewrite each importance weight as

$$w^{(j)} = \frac{p(y_1,z_1^{(j)})\prod_{t=2}^{T}\bar{p}_t(y_t,z_t^{(j)}|y_{t-1},z_{t-1}^{(j)})}{m(z_1^{(j)};\mathbf{A}_1)\prod_{t=2}^{T}m_t(z_t^{(j)}|z_{t-1}^{(j)};\mathbf{A}_t)}$$
$$= \chi_1(\mathbf{A}_1)\left\{\frac{\exp(Q_1(z_1^{(j)}))\chi_2(z_1^{(j)},\mathbf{A}_2)}{\psi_1(z_1^{(j)};\mathbf{A}_1)}\right\}$$
$$\times\prod_{t=2}^{T}\left\{\frac{\exp(Q_t(z_t^{(j)}|z_{t-1}^{(j)}))\chi_{t+1}(z_t^{(j)};\mathbf{A}_{t+1})}{\psi_t(z_t^{(j)};\mathbf{A}_t)}\right\},$$

where we define $\chi_{T+1} = 1$ for notational convenience. Note that each factor in curly brackets above depends only on a single z_t except for the reminder factors $\exp(Q_t),\ t = 2,\ldots,T$, that also depends on z_{t-1}. The EIS algorithm proceeds by choosing $\mathbf{A}_t$ to minimize the observed variance (over j) of the logarithm of each factor in curly brackets. Due to this particular parameterization of m_t, the choice of optimal importance sampler amounts to solving a sequence of least squares problems

$$\hat{\mathbf{A}}_t,\hat{C}_t = \arg\min_{\mathbf{A}_t,C_t}\sum_{j=1}^{M}\Big(Q_t(z_t^{(j)}|z_{t-1}^{(j)}) + \log\chi_{t+1}(z_t^{(j)};\mathbf{A}_{t+1})$$
$$- C_t - A_{t,1}z_t^{(j)} - A_{t,2}\left(z_t^{(j)}\right)^2\Big)^2, \quad t = T,\ldots,2, \tag{16}$$

$$\hat{\mathbf{A}}_1,\hat{C}_1 = \arg\min_{\mathbf{A}_1,C_1}\sum_{j=1}^{M}\Big(Q_1(z_1^{(j)}) + \log\chi_2(z_1^{(j)},\mathbf{A}_2) - C_1$$
$$- A_{1,1}z_1^{(j)} - A_{1,2}\left(z_1^{(j)}\right)^2\Big)^2, \tag{17}$$

which can be implemented using computationally tractable linear regression routines.

It is clear from (16)–(17) that $\mathbf{A}_t$ depends on $\mathbf{A}_{t+1}$. Hence, the sequence of regression problems has to be solved backwards. Moreover, since the draws $\mathbf{z}^{(j)} \sim m(\mathbf{z}; \mathbf{A})$ generally depend on $\mathbf{A}$, the solution for $\mathbf{A}$ using (16)–(17) is obtained iteratively. This is implemented by first choosing a suitable initial $\mathbf{A}$, and then alternating between sampling from $m(\mathbf{z}; \mathbf{A})$ using the latest available $\mathbf{A}$ and obtaining a new $\mathbf{A}$ from (16)–(17). We refer to these iterations as EIS-iterations.

It is well-known (see e.g. Geweke, 1989; Koopman et al., 2009) that the variance (over j) of the importance weights need to be finite in order to get reliable importance sampling estimates. A rule of thumb is that the tails of the importance density have to be heavier than those of the target kernel (8) so that the importance weights are bounded as functions of the draws $\mathbf{z}^{(j)}$. On the other hand, we rely on a joint importance density that is constructed from relatively thin-tailed Gaussian transition densities m_t as this is desirable from a computational perspective. Despite that we use thin-tailed Gaussian transition densities, we have not encountered problems with infinite weight variance, and there are several explanations for this. First, since the mean and the variance of the normalized version of k_t are generally non-linear functions of z_{t-1}, globally $m(\mathbf{z}; \mathbf{A})$ can generate highly non-Gaussian and heavy tailed behavior when integrating $\bar{p}(\mathbf{y}, \mathbf{z})$ with respect to $\mathbf{z}$. This property is supported by what is well-known for diffusions, namely that the joint behavior over longer time spans can be highly non-Gaussian, although for short time steps the TPD is approximately Gaussian. Second, the integrand $\bar{p}(\mathbf{y}, \mathbf{z}) \propto \bar{p}(\mathbf{z}|\mathbf{y})$ is conditioned on the observations and therefore often has thinner tails than those of the marginal $\bar{p}(\mathbf{z})$. Finally, in the example considered below, we work with a transformed latent process with constant volatility, and it is known (Aït-Sahalia, 2002) that the tails of the TPD associated with the latent process in this case are no heavier than those of a Gaussian density.

3. Example model and simulation studies

In this section we illustrate the proposed method using the GARCH diffusion model. After that, we carry out some simulation studies to check the performance of the proposed method.

3.1. The GARCH diffusion

Let Y_τ denote the log-price of an asset, and V_τ the volatility of this asset. The GARCH diffusion model is given by

$$d\begin{bmatrix} Y_\tau \\ V_\tau \end{bmatrix} = \begin{bmatrix} a \\ \alpha + \beta V_\tau \end{bmatrix} d\tau + \begin{bmatrix} \sqrt{(1-\rho^2)V_\tau} & \rho\sqrt{V_\tau} \\ 0 & \sigma V_\tau \end{bmatrix} \times \begin{bmatrix} dB_{\tau,1} \\ dB_{\tau,2} \end{bmatrix}, \quad (18)$$

where $B_{\tau,1}$ and $B_{\tau,2}$ denote a pair of independent canonical Brownian motions. The parameters to be determined are $\theta = (\alpha, \beta, \sigma, \rho, a)'$.

Provided that $\alpha > 0$, $\beta < 0$, $\sigma > 0$, the volatility process V_t is strictly stationary and mean reverts to the long run mean, $-\alpha/\beta$. The marginal 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). The parameter ρ, when it is less than 0, represents the so-called leverage effect (Harvey et al., 1994; Yu, 2005). The model was first obtained by Nelson (1990) as a continuous time limit of the discrete time GARCH(1,1) model of Bollerslev (1986). Christoffersen et al. (2010), Kleppe et al. (2010) and 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). However, the improvement in model fit comes at the cost of losing the affine structure of Heston's model. Hence, the model does not have a closed-form expression for the conditional characteristic function and cannot be estimated by the procedure of Bates (2006). For our method, on the other hand, the loss of affine structure causes no difficulty.

We follow Nelson (1990), Aït-Sahalia (2002) and Durham and Gallant (2002) by applying the variance stabilizing transformation to V_τ, namely $Z_\tau = \log(V_\tau)$. It is our experience that $p(z_t|y_t, z_{t-1}, y_{t-1})$ is better approximated by a Gaussian importance distribution than $p(v_t|y_t, v_{t-1}, y_{t-1})$ is, and that the Aït-Sahalia expansions converge faster when the domain of the diffusion is unbounded, as it is for Z_τ. The joint dynamics of Y_τ and Z_τ are obtained via Ito's lemma as

$$d\begin{bmatrix} Y_\tau \\ Z_\tau \end{bmatrix} = \begin{bmatrix} a \\ \alpha \exp(-Z_\tau) + \beta - \frac{1}{2}\sigma^2 \end{bmatrix} d\tau + \begin{bmatrix} \sqrt{(1-\rho^2)} \exp\left(\frac{1}{2}Z_\tau\right) & \rho \exp\left(\frac{1}{2}Z_\tau\right) \\ 0 & \sigma \end{bmatrix} \times \begin{bmatrix} dB_{\tau,1} \\ dB_{\tau,2} \end{bmatrix}. \quad (19)$$

The derivation of the TPD-approximations, and subsequent integration over $\mathbf{z}$ are carried out under the representation (19). For the HLM-TPDs, the differential equations governing the approximate moments of $x_t|x_{t-1}$ have a closed-form expression and are provided in the supplementary material available at http://folk.uib.no/tkl083/diffusion/.

To specify the initial density $p(y_1, z_1)$ in (8), we start by defining

$$p(z_1) = k(z_1) = \mathcal{N}\left(z_1; -\log\left(\frac{\sigma^2 - 2\beta}{2\alpha}\right), \left(\frac{\sigma^2}{\sigma^2 - 2\beta}\right)^2\right), \quad (20)$$

which represents a Gaussian Laplace approximation to $p(Z_1)$.[4] Next, we condition the likelihood on y_1, and hence we do not need to specify $p(y_1|z_1)$.

3.2. Simulation study

The proposed methodology for the GARCH diffusion was implemented in Fortran 90, with the routines for evaluating μ_t, Σ_t and Q_t being distributed on the four cores of a Dell PowerEdge R200 with 1 Quad core Intel Xeon X3330 2.66 GHz CPU and 8 GB of memory, running Linux. A total of 12 EIS iterations were employed, with initial auxiliary parameters given by

$$A_{t,1} = \frac{1}{2}\log(\max((y_{t+1} - y_t)^2, 0.00001)/\Delta), \quad t = 1, \ldots, T-1,$$
$$A_{t,2} = -\frac{1}{4}, \quad t = 1, \ldots, T-1,$$
$$A_{T,1} = a_{T,2} = 0,$$

so that $\psi_t(z_t; \mathbf{A}_t)$ is close to being proportional to $\bar{p}^{(\mathrm{EM})}(y_{t+1}|y_t, z_t)$ for $t = 1, \ldots, T-1$. When applying ASK-TPDs, the 6 first iterations were based on the EM approximation to ensure greater stability and faster convergence of the algorithm. $M = 16$ draws were used in the importance sampler. The approximate log-likelihood was maximized using a line-searching BFGS quasi-Newton optimizer (Nocedal and Wright, 1999) and forward mode automatic differentiation (Griewank, 2000) was used to calculate the gradients.

[4] That is, the Gaussian density with the same mode and the same second derivative as $p(Z_\tau)$ at the mode.

The setup of the simulation study is as follows. A yearly timescale with daily sampling interval ($\Delta = 1/252$) was used. In total, 1000 synthetic datasets, each with the sample size of $T = 2023$, were simulated using the EM scheme with time step $\Delta/256$. The sample size corresponds to approximately 8 years of data and matches that of the real data set considered later in the empirical application. The parameters used for generating synthetic data match those obtained from applying the proposed methodology to the real data.

Two experiments were designed in this Monte Carlo study. In the first experiment, we simulated Y_τ and Z_τ from Model (19) and assumed to be observed log-volatility, Z_τ. We then estimated θ using the EM-TPD, HLM-TPD, AS1-TPD, AS2-TPD and AS3-TPD ML methods. Although the assumption of observed Z_τ is not realistic in practice, the experiment allows us to check the discretization bias. In the second experiment, we estimated θ from Y_τ only, as we no longer assume Z_τ to be observed. The ML techniques used include EM-EIS, HLM-EIS, AS1-EIS, AS2-EIS and AS3-EIS.

As a benchmark in the case of unobserved log-volatility, we also considered the filtering methodology provided in Section 7 of Hurn et al. (2013) for handling latent states, which we denote as HLM-FILTER. The HLM-FILTER is obtained by alternating between the following operations in each time step t:

1. Calculate a Gaussian approximation to $p(z_t, y_t|y_1, \ldots, y_{t-1})$ based on the first two moments of $p(z_{t-1}|y_1, \ldots, y_{t-1})$. This step is carried out by setting the mean and variance at time $t-1$ equal to the mean and variance of $p(z_{t-1}|y_1, \ldots, y_{t-1})$ in the differential equations governing the evolution of moments from time $t-1$ to t.
2. Calculate an approximation to $p(y_t|y_1, \ldots, y_{t-1})$ from the Gaussian approximation to $p(z_t, y_t|y_1, \ldots, y_{t-1})$ and the observed y_t.
3. Calculate approximations to the mean and variance of $p(z_t|y_1, \ldots, y_t)$ from the Gaussian approximation to $p(z_t, y_t|y_1, \ldots, y_{t-1})$ and the observed y_t.

Then the filter-based conditional likelihood function is obtained since $p(y_2, \ldots, y_T|y_1) = \prod_{t=2}^{T} p(y_t|y_1, \ldots, y_{t-1})$, and the approximations to each factor in the latter product are obtained in step 2.

Biases and standard deviations associated with the different estimators are presented in Table 1. In the case where log-volatility is observed, the HLM- and ASK-based estimators perform similarly. This result is in line with what was found by Hurn et al. (2013). The biases are at least an order of magnitude smaller than statistical standard errors. On the other hand, the estimator EM-TPD introduces a much bigger bias in the volatility-of-volatility parameter σ than the other methods.

In the case where volatility is latent, the differences among alternative methods are much bigger. There are considerable biases in β and σ for the EM-EIS and HLM-EIS. For the ASK-EIS, the bias is much smaller. This observation suggests that in the latent volatility case, important information concerning V_t is contained in the higher order moments of the true TPD. Therefore, there is a need for using non-Gaussian approximate TPDs, such as the ASK. Moreover, we observe that the results from AS2 and AS3 are very similar, suggesting that AS2 provides sufficient accuracy in this experiment.

Comparing AS2-EIS with the benchmark HLM-FILTER method, we see that our proposed method produces smaller biases in all cases, and for α, σ and ρ, substantially smaller standard errors. However, it should be noted that the increase in statistical efficiency for AS2-EIS is associated with a considerably higher computational cost relative to HLM-FILTER.

4. Empirical application

In the empirical application, we fit the GARCH diffusion (19) to the logarithm of the Standard & Poor 500 index, sampled daily from January 3, 2003, to January 13, 2011 (hence $\Delta = 1/252$). In total we have $T = 2023$ observations. Parameter estimates obtained using the alternative estimation procedures are presented in Table 2. The estimates based on EIS-integration are calculated as the averaged estimates across 100 sets of estimates with different random seeds in the importance sampler. In addition, we also calculate Monte Carlo standard errors (MC Std.err.) induced by the EIS-integration methods as the standard deviation across the 100 sets of estimates. As is expected, the Monte Carlo standard errors are smaller than the statistical standard errors (Std.dev). As a further check of reliability of the proposed EIS-integration procedure, we carry out a battery of tests in line with Koopman et al. (2009) to check for infinite variance in the importance weights $w^{(j)}$. No rejection of the finite variance was found and the detailed results are reported to supplementary material that can be found at http://folk.uib.no/tkl083/diffusion/.

The estimates of ρ are much larger in magnitude than what has been found in the literature using data from earlier periods. The estimated ρ 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 ρ is similar to what has been found in Yu (2012), Aït-Sahalia et al. (2013) and Hurn et al. (2013). For the benchmark HLM-FILTER we obtain estimates that are rather different. Most notably, the estimate of ρ is very close to -1. This, along with evidence from the simulation study, leads us to conclude that a very accurate numerical integration method, such as the EIS, is needed for obtaining MLE from price data only.

Focusing on the EIS-based methods, we see substantial differences in parameters α, β, σ and ρ, suggesting that discretization bias is important, and, hence, a refined TPD approximation, such as AS2 and AS3, is needed. In particular, the estimates of the leverage effect parameter ρ and the volatility of volatility parameter σ change by approximately two standard deviations between the EM-TPDs and HLM-TPDs estimates and the AS3-TPDs estimate. Moreover, in line with the simulation study we see that the parameter estimates for the three ASK-based methods are close to each others and that the differences in the parameter estimates between AS2 and AS3 are less than those between AS1 and AS2. This leads us to conclude that AS2-TPD along with EIS is our preferred method for carrying out ML estimation in this empirical study.

5. Concluding remarks

This paper presents a new and computationally efficient method for ML estimation of partially observed diffusions. To compute the likelihood function, we propose to use the EIS method of Richard and Zhang (2007) to integrate out the latent states from the joint density of the observed and the latent state variables, which is in turn approximated using the closed-form approach of Aït-Sahalia (2008). The use of the closed-form TPD approximations is important as the discretization bias is controlled. The proposed method does not require any infill observations and hence is computationally appealing. In the context of GARCH diffusions the proposed method performs better than the methods that do not control the discretization bias, including the EM method and the QML method.

An additional advantage of the proposed methodology is that it is very easy to adapt to the case where the diffusion is completely latent, and we only have noisy observations of (some of) the states.

Table 1
Results from the Monte Carlo experiment for the GARCH diffusion. All results are obtained based on 1000 synthetic data sets, each with $T = 2023$ and θ equal to those in the "True parameters" row. When volatility is observed, EM, HLM and ASK, $K = 1, 2, 3$ correspond to ML based on the EM-TPD, Hurn et al. (2013)-TPD and Aït-Sahalia (2008)-TPD. When volatility is unobserved, EM-EIS, HLM-EIS and ASK-EIS correspond to ML based on the EM-EIS, Hurn et al. (2013)-EIS and Aït-Sahalia (2008)-EIS. HLM-FILTER is ML based on the HLM-TPDs with volatility being integrated out using the filtering method of Hurn et al. (2013). Mean computing time for the case of the latent volatility ranges from 43 (EM) to 103 (AS3) seconds.

Method		α	β	σ	ρ	a
	True parameters	0.0948	−1.1754	3.2607	−0.8467	−0.0183
Observed log-volatility						
EM	Bias	−0.0114	0.1321	−0.0534	3.5e−3	0.0213
	Std.dev.	0.0326	0.9862	0.0362	5.2e−3	0.0821
HLM	Bias	3.4e−3	0.0280	5.4e−3	−4.9e−4	−8.6e−3
	Std.dev.	0.0136	1.0341	0.0372	5.1e−3	0.0457
AS1	Bias	−3.6e−3	0.1224	−2.2e−3	−4.4e−5	4.6e−3
	Std.dev.	0.0229	0.9492	0.0362	5.1e−3	0.0597
AS2	Bias	−1.7e−3	0.0417	−2.3e−3	−1.4e−4	3.8e−3
	Std.dev.	0.0235	0.9897	0.0361	5.1e−3	0.0610
AS3	Bias	−1.4e−3	0.0351	−2.7e−3	−6.9e−5	3.3e−3
	Std.dev.	0.0230	0.9818	0.0359	5.1e−3	0.0606
Unobserved log-volatility						
EM-EIS	Bias	1.8e−3	−0.5580	−0.2073	0.0285	0.0144
	Std.dev.	0.0192	1.1675	0.2440	0.0464	0.0437
HLM-EIS	Bias	1.1e−3	−0.4705	−0.1499	0.0257	0.0153
	Std.dev.	0.0191	1.1403	0.2538	0.0465	0.0436
HLM-FILTER	Bias	0.0324	0.4694	0.0493	−0.0376	−0.0844
	Std.dev.	0.0562	0.9608	0.8564	0.1460	0.1556
AS1-EIS	Bias	4.5e−4	0.0551	−0.0460	−4.3e−3	−6.4e−3
	Std.dev.	0.0186	0.9481	0.2580	0.0360	0.0449
AS2-EIS	Bias	4.7e−3	−0.1085	−0.0129	−6.7e−3	−6.6e−3
	Std.dev.	0.0192	1.0473	0.2627	0.0363	0.0440
AS3-EIS	Bias	4.4e−3	−0.1107	−0.0149	−5.9e−3	−6.0e−3
	Std.dev.	0.0192	1.0458	0.2623	0.0358	0.0445

Table 2
Parameter estimates and log-likelihood values for the GARCH diffusion fitted to S&P500 data using different estimation procedures. The parameter estimates based on EIS-integration are calculated as the mean over 100 random seeds in the importance sampler. "MC Std.err" denotes the Monte Carlo standard errors across the 100 sets of estimates. Statistical standard errors (Std.dev.) are taken from Table 1.

Method		α	β	σ	ρ	a	Log-likelihood
EM-EIS	Estimate	0.0788	−1.6783	2.7119	−0.7661	0.0137	6529.3
	Std.dev.	0.0192	1.1675	0.2440	0.0464	0.0437	
	MC Std.err.	4.1e−4	0.0139	0.0063	9.5e−4	2.5e−4	0.1170
HLM-EIS	Estimate	0.0772	−1.5715	2.7409	−0.7676	0.0147	6529.3
	Std.dev.	0.0191	1.1403	0.2538	0.0465	0.0436	
	MC Std.err.	4.0e−4	0.0134	0.0064	9.3e−4	2.5e−4	0.1167
HLM-FILTER	Estimate	0.0745	−0.1875	2.0267	−0.9878	−0.1268	6449.3
	Std.dev.	0.0562	0.9608	0.8564	0.1460	0.1556	
AS1-EIS	Estimate	0.0908	−0.9931	3.2343	−0.8515	−0.0195	6544.2
	Std.dev.	0.0186	0.9481	0.2580	0.0360	0.0450	
	MC Std.err.	3.9e−4	0.0095	0.0074	5.2e−4	2.3e−4	0.1258
AS2-EIS	Estimate	0.0948	−1.1754	3.2607	−0.8467	−0.0183	6544.4
	Std.dev.	0.0192	1.0473	0.2627	0.0363	0.0440	
	MC Std.err.	5.0e−4	0.0111	0.0087	5.2e−4	2.2e−4	0.1259
AS3-EIS	Estimate	0.0946	−1.1833	3.2542	−0.8456	−0.0182	6544.4
	Std.dev.	0.0192	1.0458	0.2623	0.0358	0.0445	
	MC Std.err.	5.0e−4	0.0111	0.0086	5.3e−4	2.3e−4	0.1257

Using this adaptation, we have successfully estimated the one-factor short-term interest rate model of Chan et al. (1992) to data contaminated with micro structure noise. For that model, it is found that disregarding the contamination of data can lead to significantly different estimates. The method is also applicable to the structural credit risk models of Duan and Fulop (2009) and Huang and Yu (2010). Results for these two classes of models may be found in an earlier version of the present paper.

As an alternative to ML estimation, Bayesian Markov Chain Monte Carlo (MCMC) methods may also be used to estimate partially observed diffusion models; see Eraker (2001) and Stramer et al. (2010). To control the discretization bias, Eraker (2001) proposed to use an infill technique, which is computationally expensive. Stramer et al. (2010) is especially relevant to the present paper because it uses the closed-form method in connection to MCMC.

There should be scope for applying the current methodology to a broader class of models, including models with multiple latent states. 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 μ_t, Σ_t and Q_t. As in the univariate case, the sampling interval Δ and the degree of

deviation from the normality for the latent process are important factors to determine the effectiveness of the proposed method relative to a more crude approximation.

A possible direction for further research is to allow for jumps in one or more of the states. To deal with diffusion models with jumps, Yu (2007) provides polynomial TPD-expansions for jump-diffusions. Relative to the TPD of pure diffusion models, the TPD of jump-diffusion is expected to deviate more from the Gaussian distribution and, hence, it is more challenging for the EIS to perform well. To cope with jumps in the EIS framework, the mixture EIS framework of Kleppe and Liesenfeld (forthcoming) can be employed and we leave the implementation of this new method for future research.

References

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

Aït-Sahalia, Y., 2008. Closed-form likelihood expansions for multivariate diffusions. Ann. Statist. 36 (2), 906–937.

Aït-Sahalia, Y., Fan, J., Li, Y., 2013. The leverage effect puzzle: disentangling sources of bias at high frequency. J. Financ. Econ. 109 (1), 224–249.

Aït-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. J. Financ. Econ. 134, 507–551.

Aït-Sahalia, Y., Yu, J., 2006. Saddlepoint approximations for continuous-time Markov processes. J. Econometrics 134 (2), 507–551.

Balduzzi, P., Das, S.R., Foresi, S., 1998. The central tendency: a second factor in bond yields. Rev. Econ. Stat. 80 (1), 62–72.

Bates, D.S., 2006. Maximum likelihood estimation of latent affine processes. Rev. Financ. Stud. 19 (3), 909–965.

Bergstrom, A.R., 1966. Nonrecursive models as discrete approximations to systems of stochastic differential equations. Econometrica 34 (1), 173–182.

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

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. J. Polit. Econ. 81 (3), 637–654.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307–327.

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

Christoffersen, P., Jacobs, K., Mimouni, K., 2010. Volatility dynamics for the S&P500: evidence from realized volatility, daily returns, and option prices. Rev. Financ. Stud. 23 (8), 3141–3189.

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

Duan, J.-C., Fulop, A., 2009. Estimating the structural credit risk model when equity prices are contaminated by trading noises. J. Econometrics 150 (2), 288–296.

Duan, J., Yeh, C.-Y., 2011. Price and volatility dynamics implied by the VIX term structure. Working Paper.

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

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

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

Durham, G.B., 2006. Monte Carlo methods for estimating, smoothing, and filtering one and two-factor stochastic volatility models. J. Econometrics 133, 273–305.

Durham, G.B., Gallant, A.R., 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). J. Bus. Econom. Statist. 20 (3), 297–338.

Elerian, O., Chib, S., Shephard, N., 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. J. Bus. Econom. Statist. 19, 177–191.

Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57 (6), 1317–1339.

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

Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Rev. Econom. Stud. 61, 247–264.

Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bonds and currency options. Rev. Financ. Stud. 6, 327–343.

Huang, S.J., Yu, J., 2010. Bayesian analysis of structural credit risk models with microstructure noises. J. Econom. Dynam. Control 34 (11), 2259–2272.

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

Hurn, A.S., Jeisman, J.I., Lindsay, K.A., 2007. Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. J. Financ. Econ. 5 (3), 390–455.

Hurn, A., Lindsay, K., McClelland, A., 2013. A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions. J. Econometrics 172 (1), 106–126.

Jensen, B., Poulsen, R., 2002. Transition densities of diffusion processes: numerical comparison of approximation techniques. J. Derivatives 9 (4), 18–32.

Kleppe, T.S., Liesenfeld, R., 2013. Efficient high-dimensional importance sampling in mixture frameworks. Comput. Statist. Data Anal. (forthcoming).

Kleppe, T.S., Yu, J., Skaug, H.J., 2010. Simulated maximum likelihood estimation of continuous time stochastic volatility models. In: Greene, W., Hill, R.C. (Eds.), Maximum Simulated Likelihood Methods and Applications. In: Advances in Econometrics, vol. 26. Emerald Group Publishing Limited, pp. 137–161.

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

Lee, K., Koopman, S., 2004. Estimating stochastic volatility models: a comparison of two importance samplers. Stud. Nonlinear Dyn. Econom. 8 (2).

Liesenfeld, R., Richard, J.-F., 2003. Univariate and multivariate stochastic volatility models: estimation and diagnostics. J. Empir. Finance 10, 505–531.

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

Milstein, G., 1979. A method of second-order accuracy integration of stochastic differential equations. Theory Probab. Appl. 23 (2), 396–401.

Nelson, D.B., 1990. ARCH models as diffusion approximations. J. Econometrics 45, 7–38.

Nocedal, J., Wright, S.J., 1999. Numerical Optimization. Springer.

Nowman, K.B., 1997. Gaussian estimation of single-factor continuous time models of the term structure of interest rates. J. Finance 52 (4), 1695–1706.

Pedersen, A.R., 1995. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scand. J. Statist. 22 (1), 55–71.

Phillips, P.C.B., Yu, J., 2009. Maximum likelihood and Gaussian estimation of continuous time models in finance. In: Torben Andersen, E.A. (Ed.), Handbook of Financial Time Series. Springer, pp. 497–530.

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

Sandmann, G., Koopman, S., 1998. Maximum likelihood estimation of stochastic volatility models. J. Econometrics 63, 289–306.

Shephard, N., Pitt, M.K., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653–667.

Shoji, I., Ozaki, T., 1998. Estimation for nonlinear stochastic differential equations by a local linearization method 1. Stoch. Anal. Appl. 16 (4), 733–752.

Skaug, H.J., Yu, J., 2013. A flexible and automated likelihood based framework for inference in stochastic volatility models. Comput. Statist. Data Anal. (forthcoming).

Stramer, O., Bognar, M., Schneider, P., 2010. Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions. J. Financ. Econ. 8 (4), 450–480.

Turnovsky, S.J., 2000. Methods of Macroeconomic Dynamics, second ed. MIT Press.

Yu, J., 2005. On leverage in a stochastic volatility model. J. 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. J. Econometrics 141 (2), 1245–1280.

Yu, J., 2012. A semiparametric stochastic volatility model. J. Econometrics 167 (2), 473–482.