SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance

Peter C. B. Phillips, Jun Yu

December 2006

Paper No. 30-2006

Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance*

Peter C. B. Phillips†, Jun Yu‡

April 2, 2006

Abstract

This paper overviews maximum likelihood and Gaussian methods of estimating continuous time models used in finance. Since the exact likelihood can be constructed only in special cases, much attention has been devoted to the development of methods designed to approximate the likelihood. These approaches range from crude Euler-type approximations and higher order stochastic Taylor series expansions to more complex polynomial-based expansions and infill approximations to the likelihood based on a continuous time data record. The methods are discussed, their properties are outlined and their relative finite sample performance compared in a simulation experiment with the nonlinear CIR diffusion model, which is popular in empirical finance. Bias correction methods are also considered and particular attention is given to jackknife and indirect inference estimators. The latter retains the good asymptotic properties of ML estimation while removing finite sample bias. This method demonstrates superior performance in finite samples.

Keywords: Maximum likelihood, Transition density, Discrete sampling, Continuous record, Realized volatility, Bias reduction, Jackknife, Indirect inference.

AMS Subject Classifications: 62M10
JEL Classifications: C22, C32

*Phillips gratefully acknowledges visiting support from the School of Economics and Social Sciences at Singapore Management University and support from a Kelly Fellowship at the University of Auckland Business School and from the NSF under Grant No. SES 04-142254. Yu gratefully acknowledge financial support from the Office of Research at Singapore Management University.

†Cowles Foundation for Research in Economics, Yale University, University of Auckland and University of York; email: peter.phillips@yale.edu

‡School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903; email: yujun@smu.edu.sg.

1 Introduction

Continuous time models have provided a convenient mathematical framework for the development of financial economic theory (e.g., Merton, 1990), asset pricing, and the modern field of mathematical finance that relies heavily on stochastic processes (Karatzas and Shreve, 2003). These models now dominate the option pricing literature, which has mushroomed over the last three decades from a single paper (Black and Scholes, 1973) to a vast subdiscipline with strong practical applications in the finance industry. Correspondingly, the econometric analysis of continuous time models has received a great deal attention in financial econometrics, providing a basis from which these models may be brought to data and used in practical applications. Much of the focus is on the econometric estimation of continuous time diffusion equations. Estimation not only provides parameter estimates which may be used directly in the pricing of financial assets and derivatives but also serves as a stage in the empirical analysis of specification and comparative diagnostics.

Many models that are used to describe financial time series are written in terms of a continuous time diffusion $X(t)$ that satisfies the stochastic differential equation

$$dX(t) = \mu(X(t);\theta)dt + \sigma(X(t);\theta)dB(t), \tag{1}$$

where $B(t)$ is a standard Brownian motion, $\sigma(X(t);\theta)$ is some specified diffusion function, $\mu(X(t);\theta)$ is a given drift function, and θ is a vector of unknown parameters. This class of parametric model has been widely used to characterize the temporal dynamics of financial variables, including stock prices, interest rates, and exchange rates.

It has been argued that when the model is correctly specified, the preferred choice of estimator and preferred basis for inference should be maximum likelihood (ML) – see, for example, Aït-Sahalia (2002) and Durham and Gallant (2002). Undoubtedly, the main justification for the use of the ML method lies in its desirable asymptotic properties, particularly its consistency and asymptotic efficiency under conditions of correct specification. In pursuit of this goal, various ML and Gaussian (that is, ML under Gaussian assumptions) methods have been proposed. Some of these methods involve discrete approximations, others are exact (or exact under certain limiting conditions on the approximation). Some are computationally inexpensive while others are computationally intensive. Some are limited to particular formulations, others have much wide applicability.

The purpose of the present chapter is to review this literature and overview the many different approaches to estimating continuous time models of the form given by (1) using ML and Gaussian methods. In the course of this overview, we shall discuss the existing methods of estimation and their merits and drawbacks. A simple Monte Carlo experiment is designed to

reveal the finite sample performance of some of the most commonly used estimation methods. The model chosen for the experiment is a simple example of (1) that involves a square root diffusion function. This model is popular in applied work for modeling short term interest rates and is known in the term structure literature as the Cox-Ingersoll-Ross or CIR model (see (8) below). One of the principal findings from this simulation experiment is that all ML methods, including "exact" methods, have serious *finite sample estimation bias* in the mean reversion parameter. This bias is significant even where the number of observations is as large as 500 or 1000. It is therefore important in ML/Gaussian estimation to take such bias effects into account. We therefore consider two estimation bias reduction techniques – the jackknife method and the indirect inference estimation – which may be used in conjunction with ML, Gaussian or various approximate ML methods. The indirect inference estimator demonstrates markedly superior results in terms of bias reduction and overall mean squared error in comparison with all other methods.

The chapter is organized as follows. Section 2 outlines the exact ML method, Section 3 and Section 4 review the literature on implementing ML/Gaussian methods in continuous time financial models and the practicalities of implementation. Section 5 reports a Monte Carlo study designed to investigate and compare the performance of some ML/Gaussian estimation methods for the CIR model. Section 6 reviews two bias reduction methods and examines their performance in the CIR model example. Section 7 briefly outlines some issues associated with extensions of ML/Gaussian procedures for multivariate models, and Section 8 concludes.

2 Exact ML Methods

2.1 ML based on the Transition Density

Assume the data $X(t)$ is recorded discretely at points $(h, 2h, \cdots, Nh(\equiv T))$ in the time interval $[0, T]$, where h is the discrete interval of observation of $X(t)$ and T is the time span of the data. The full sequence of N observations is $\{X_h, X_{2h}, \cdots, X_{Nh}\}$. If $X(t)$ is conceptualized for modeling purposes as annualized data which is observed discretely at monthly (weekly or daily) intervals, then $h = 1/12$ ($1/52$ or $1/252$). It is, of course, most convenient to assume that equi-spaced sampling observations are available and this assumption is most common in the literature, although it can be and sometimes is relaxed.

Many estimation methods are based on the construction of a likelihood function derived from the transition probability density of the discretely sampled data. This approach is explained as follows. Suppose $p(X_{ih}|X_{(i-1)h}, \theta)$ is the transition probability density. The Markov

property of model (1) implies the following log-likelihood function for the discrete sample[1]

$$\ell_{TD}(\theta) = \ln(p(X_{ih}|X_{(i-1)h}, \theta)). \tag{2}$$

The resulting estimator will be consistent, asymptotically normally distributed and asymptotically efficient under the usual regularity conditions for maximum likelihood estimation in (stationary) dynamic models (Hall and Heyde, 1980; Billingsley, 1961). In nonstationary, nonergodic cases, the limit theory is no longer asymptotically normal and there are several possibilities, including various unit root, local to unity, mildly explosive and explosive limit distributions (Phillips, 1987, Chan and Wei, 1988; Phillips, 1991; Phillips and Magdalinos, 2006).

To perform exact ML estimation, one needs a closed form expression for $\ell_{TD}(\theta)$ and hence $\ln(p(X_{ih}|X_{(i-1)h}, \theta))$. Unfortunately, only in rare cases, do the transition density and log likelihood component $\ln(p(X_{ih}|X_{(i-1)h}, \theta))$ have closed form analytical expressions. All other cases require numerical techniques or analytic or simulation-based approximants.

The following list reviews the continuous time models used in finance that have closed-form expressions for the transition density.

1. **Geometric Brownian Motion**:

$$dX(t) = \mu X(t)\, dt + \sigma X(t)\, dB(t). \tag{3}$$

Black and Scholes (1973) used this process to describe the movement of stock prices in their development of the stock option price formula. Since

$$d \ln X(t) = \frac{1}{X(t)} dX(t) - \frac{(dX(t))^2}{2X(t)^2} = \mu dt + \sigma dB(t) - \frac{1}{2}\sigma^2 dt,$$

the transformed process $\ln X(t)$ follows the linear diffusion

$$d \ln X(t) = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\, dB(t). \tag{4}$$

As a result, $X_{ih}|X_{(i-1)h} \sim \text{LN}((\mu - \frac{\sigma^2}{2})h + \ln(X_{(i-1)h}), \sigma^2 h)$, where LN denotes the log-normal distribution.

[1]Our focus in the present discussion is on the usefulness of the transition density for estimation purposes. But we note that the transition density is needed and used for many other applications, such as for pricing derivatives and for obtaining interval and density forecasts.

2. **Ornstein-Uhlenbeck (OU) process (or Vasicek model)**:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma\, dB(t). \tag{5}$$

Vasicek (1977) used this process to describe the movement of short term interest rates. Phillips (1972) showed that the exact discrete model corresponding to (5) is given by

$$X_{ih} = e^{-\kappa h} X_{(i-1)h} + \mu\left(1 - e^{-\kappa h}\right) + \sigma\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}\epsilon_i, \tag{6}$$

where $\epsilon_i \sim$ i.i.d. $N(0,1)$. Phillips (1972) also developed an asymptotic theory for nonlinear least squares/ML estimates of the parameters in a multivariate version of (5) using the exact discrete time model (6), showing consistency, asymptotic normality and efficiency under stationarity assumptions ($\kappa > 0$ in the univariate case here). The transition density for the Vasicek model follows directly from (6) and is

$$X_{ih}|X_{(i-1)h} \sim N\left(\mu(1 - e^{-\kappa h}) + e^{-\kappa h} X_{(i-1)h}, \sigma^2(1 - e^{-2\kappa h})/(2\kappa)\right). \tag{7}$$

3. **Square-root (or Cox-Ingersoll-Ross) model:**

$$dX(t) = \kappa(\mu - X(t))dt + \sigma\sqrt{X(t)}\, dB(t). \tag{8}$$

Cox, Ingersoll and Ross (1985, CIR hereafter) also used this process to describe movements in short term interest rates. The exact discrete model corresponding to (8) is given by

$$X_{ih} = e^{-\kappa h} X_{(i-1)h} + \mu\left(1 - e^{-\kappa h}\right) + \sigma\int_{(i-1)h}^{ih} e^{-\kappa(ih-s)}\sqrt{X(s)}dB\left(s\right). \tag{9}$$

When $2\kappa\mu/\sigma^2 \geq 1$, X is distributed over the positive half line. Feller (1952) showed that the transition density of the square root model is given by

$$X_{ih}|X_{(i-1)h} = ce^{-u-v}(v/u)^{q/2} I_q(2(uv)^{1/2}) \tag{10}$$

where $c = 2\kappa/(\sigma^2(1 - e^{-\kappa h}))$, $u = cX_{(i-1)h}e^{-\kappa h}$, $v = cX_{ih}$, $q = 2\kappa\mu/\sigma^2 - 1$, and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q.

4. **Inverse square-root model:**

$$dX(t) = \kappa(\mu - X(t))X(t)dt + \sigma X^{1.5}(t)\, dB(t). \tag{11}$$

Ahn and Gao (1999) again used this process to model short term interest rates. When $\kappa, \mu > 0$, X is distributed over the positive half line. Ahn and Gao (1999) derived the transition density of the inverse square root model as

$$X_{ih}|X_{(i-1)h} = c^{-1}e^{-u-v}(v)^{q/2+2}u^{-q/2} I_q(2(uv)^{1/2}) \tag{12}$$

where $c = 2\kappa\mu/(\sigma^2(1 - e^{-\kappa\mu h}))$, $u = ce^{-\kappa\mu h}/X_{(i-1)h}$, $v = c/X_{ih}$, $q = 2(\kappa + \sigma^2)/\sigma^2 - 1$.

2.2 ML based on the Continuous Record Likelihood

If a continuous sample path of the process $X(t)$ were recorded over the interval $[0, T]$, direct ML estimation would be possible based on the continuous path likelihood. This likelihood is very useful in providing a basis for the so-called continuous record or infill likelihood function and infill asymptotics in which a discrete record becomes continuous by a process of infilling as the sampling interval $h \to 0$. Some of these infill techniques based on the continuous record likelihood are discussed later in Section 4. Since financial data are now being collected on a second by second and tick by tick basis, this construction is becoming much more important.

When $X(t)$ is observed continuously, a log-likelihood function for the continuous record $\{X(t)\}_{t=0}^{T}$ may be obtained directly from the Radon Nikodym (RN) derivative of the relevant probability measures. The RN derivative produces the relevant probability density and can be regarded as a change of measure among the absolutely continuous probability measures, the calculation being facilitated by the Girsanov theorem (e.g., Karatzas and Shreve, 2003). The approach is convenient and applies quite generally to continuous time models with flexible drift and diffusion functions.

In the stochastic process literature the quadratic variation or square bracket process is well known to play an important role in the study of stochastic differential equations. In the case of equation (1), the square bracket process of $X(t)$ has the explicit form

$$[X]_T = \int_0^T (dX(t))^2 = \int_0^T \sigma^2(X(t);\theta)dt, \tag{13}$$

which is a continuously differentiable increasing function. In fact, we have $d[X]_t = \sigma(X(t);\theta)^2 dt$. In consequence, when a continuous sample path of the process $X(t)$ is available, the quadratic variation of X provides a perfect estimate of the diffusion function and hence the parameters on which it depends, provided these are identifiable in $\sigma^2(X(t);\theta)$. Thus, with the availability of a continuous record, we can effectively assume the diffusion term (i.e., $\sigma(X(t);\theta) = \sigma(X(t))$ is known and so this component does not involve any unknown parameters. It follows that the exact continuous record or infill log-likelihood can be constructed via the Girsanov theorem (e.g., Liptser and Shiryaev, 2000) as

$$\ell_{IF}(\theta) = \int_0^T \frac{\mu(X(t);\theta)}{\sigma^2(X(t))} dX(t) - \frac{1}{2}\int_0^T \frac{\mu^2(X(t);\theta)}{\sigma^2(X(t))} dt. \tag{14}$$

In this likelihood, the parameter θ enters via the drift function $\mu(X(t);\theta)$. Lánska (1979) established the consistency and asymptotic normality of the continuous record ML estimator of θ when $T \to \infty$ under certain regularity conditions.

To illustrate the approach, consider the following OU process,

$$dX(t) = \kappa X(t)dt + \sigma_0 dB(t),$$

where σ_0 is known and κ is the only unknown parameter. The exact log-likelihood in this case is given by

$$\ell_{IF}(\kappa) = \int_0^T \frac{\kappa X(t)}{\sigma_0^2} dX(t) - \frac{1}{2}\int_0^T \frac{\kappa^2 X^2(t)}{\sigma_0^2} dt,$$

and maximizing the log-likelihood function immediately gives rise to the following ML estimator of κ:

$$\hat{\kappa} = \left(\int_0^T X^2(t)dt\right)^{-1} \int_0^T X(t)dX(t) \tag{15}$$

This estimator is analogous in form to the ML/OLS estimator of the autoregressive coefficient in the discrete time Gaussian autoregression

$$X_t = \phi X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{i.i.d. } N(0,1) \tag{16}$$

viz., $\hat{\phi} = \left(\sum_{t=1}^n X_{t-1}^2\right)^{-1} \sum_{t=1}^n X_t X_{t-1}$. It is also interesting to observe that when $\kappa = 0$ (15) has the same form as the limit distribution of the (discrete time) autoregressive coefficient estimator when $\phi = 1$ in (16). These connections with unit root limit theory are explored in Phillips (1987).

In practice, of course, a continuous record of $\{X(t)\}_{t=0}^T$ is not available and estimators such as (15) are infeasible. On the other hand, as the sampling interval h shrinks, discrete data may be used to produce increasingly good approximations to the quadratic variation (13), the continuous record likelihood (14) and estimators such as (15). These procedures may be interpreted as infill likelihood methods in that they replicate continuous record methods by infilling the sample record as $h \to 0$.

3 Approximate ML Methods Based on Transition Densities

Except for a few special cases such as those discussed earlier, the transition density does not have a closed-form analytic expression. As a result, the exact ML method discussed in Section 2.1 is not generally applicable. To address this complication, many alternative approaches have been developed. The methods involve approximating the transition densities, the model itself or the likelihood function. This section reviews these methods.

3.1 The Euler Approximation and Refinements

The Euler scheme approximates a general diffusion process such as equation (1) by the following discrete time model

$$X_{ih} = X_{(i-1)h} + \mu(X_{(i-1)h}, \theta)h + \sigma(X_{(i-1)h}, \theta)\sqrt{h}\epsilon_i, \tag{17}$$

where $\epsilon_i \sim$ i.i.d. $N(0,1)$. The transition density for the Euler discrete time model has the following closed form expression:

$$X_{ih}|X_{(i-1)h} \sim N\left(X_{(i-1)h} + \mu(X_{(i-1)h}, \theta)h, \sigma^2(X_{(i-1)h}, \theta)h\right). \tag{18}$$

For the Vasicek model, the Euler discrete approximation is of the form

$$X_{ih} = \kappa\mu h + (1 - \kappa h)X_{(i-1)h} + \sigma N(0, h). \tag{19}$$

Comparing the approximation (19) with the exact discrete time model (6), we see that $\kappa\mu h$, $1-\kappa h$ and $\sigma^2 h$ replace $\mu(1-e^{-\kappa h})$, $e^{-\kappa h}$, and $\sigma^2(1-e^{-2\kappa h})/(2\kappa)$, respectively. These replacements may be motivated by considering the first order term in the following Taylor expansions:

$$\mu(1 - e^{-\kappa h}) = \kappa\mu h + O(h^2), \tag{20}$$

$$e^{-\kappa h} = 1 - \kappa h + O(h^2), \tag{21}$$

$$\sigma^2(1 - e^{-2\kappa h})/(2\kappa) = \sigma^2 h + O(h^2). \tag{22}$$

Obviously, when h is small, the Euler scheme should provide a good approximation to the exact discrete time model. However, when h is large, the Euler approximation can be poor. To illustrate magnitude of the approximation error, first consider the case where $\kappa = 1$ and $h = 1/12$, in which case $e^{-\kappa h}$ is 0.92 whereas $1 - \kappa h$ is 0.9167 and the approximation is good. But if $\kappa = 1$ and $h = 1$, then $e^{-\kappa h}$ is 0.3679 whereas $1 - \kappa h$ is 0. These comparisons suggest that the Euler discretization offers a good approximation to the exact discrete time model for daily or higher frequencies but not for annual or lower frequencies. The bias introduced by this discrete time approximation is called the *discretization bias.*

The advantages of the Euler method include the ease with which the likelihood function is obtained, the low computational cost, and the wide range of its applicability. The biggest problem with the procedure is that when h is fixed the estimator is inconsistent (Merton, 1980; Lo, 1988). The magnitude of the inconsistency can be analyzed, using the methods of Sargan (1974), in terms of the observation interval h. Lo (1988) illustrated the size of inconsistency in the context of model (3).

A closely related discretization method, suggested by Bergstrom (1966) and Houthakker and Taylor (1966), is based on integrating the stochastic differential equation and using the following trapezoidal rule approximation

$$\int_{(i-1)h}^{ih} \mu(X(t);\theta)dt = \frac{h}{2}\left\{\mu(X_{ih};\theta) + \mu(X_{(i-1)h};\theta)\right\}.$$

For the OU process the corresponding discrete approximate model is given by

$$X_{ih} - X_{(i-1)h} = \kappa\mu - \frac{\kappa h}{2}\left(X_{ih} + X_{(i-1)h}\right) + \sigma N(0,h), \tag{23}$$

which involves the current period observation X_{ih} on both sides of the equation. Solving (23) we obtain

$$\begin{aligned} X_{ih} &= \frac{\kappa\mu h}{\left(1+\frac{\kappa h}{2}\right)} + \frac{1-\frac{\kappa h}{2}}{1+\frac{\kappa h}{2}}X_{(i-1)h} + \frac{\sigma}{\left(1+\frac{\kappa h}{2}\right)}N(0,h) \\ &= \kappa\mu h + (1-\kappa h)X_{(i-1)h} + \sigma N(0,h) + O\left(h^{3/2}\right), \end{aligned}$$

so that the Bergstrom approximation is equivalent to the Euler approximation to $O(h)$. In the multivariate case, the Bergstrom approximation leads to a non-recursive simultaneous equations model approximation to a system of recursive stochastic differential equations. The resulting system may be estimated by a variety of simultaneous equations estimators, such as instrumental variables, for example by using lagged X values as instruments. Again, the magnitude of the inconsistency may be analyzed in terms of the observation interval h, as in Sargan (1974) who showed the asymptotic bias in the estimates to be typically of $O\left(h^2\right)$.

There are a number of ways to reduce the discretization bias induced by the Euler approximation. Before we review these refinements, it is important to emphasize that the aim of these refinements is simply bias reduction.

Elerian (1998) suggests using the scheme proposed by Milstein (1978). The idea is to take a second order term in a stochastic Taylor series expansion to refine the Euler approximation (17). We proceed as follows. Integrating (1) we have

$$\int_{(i-1)h}^{ih} dX(t) = \int_{(i-1)h}^{ih} \mu(X(t);\theta)dt + \int_{(i-1)h}^{ih} \sigma(X(t);\theta)dB(t), \tag{24}$$

and by stochastic differentiation we have

$$\begin{aligned} d\mu(X(t);\theta) &= \mu'(X(t);\theta)dX(t) + \frac{1}{2}\mu''(X(t);\theta)\left(dX(t)\right)^2 \\ &= \mu'(X(t);\theta)dX(t) + \frac{1}{2}\mu''(X(t);\theta)\sigma^2(X(t);\theta)dt, \end{aligned}$$

and

$$d\sigma(X(t);\theta) = \sigma'(X(t);\theta)dX\,(t) + \frac{1}{2}\sigma''(X(t);\theta)\sigma^2(X(t);\theta)dt,$$

so that

$$\begin{aligned}
\mu(X\,(t)\,;\theta) &= \mu(X_{(i-1)h};\theta) + \int_{(i-1)h}^{t} \mu'(X(s);\theta)dX\,(s) + \frac{1}{2}\int_{(i-1)h}^{t} \mu''(X(s);\theta)\sigma^2(X(s);\theta)ds \\
&= \mu(X_{(i-1)h};\theta) + \int_{(i-1)h}^{t} \mu'(X(s);\theta)\mu(X(s);\theta)ds + \frac{1}{2}\int_{(i-1)h}^{t} \mu''(X(s);\theta)\sigma^2(X(s);\theta)ds \\
&\quad + \int_{(i-1)h}^{t} \mu'(X(s);\theta)\sigma(X(s);\theta)dB(s),
\end{aligned}$$

and

$$\begin{aligned}
\sigma(X\,(t)\,;\theta) &= \sigma(X_{(i-1)h};\theta) + \int_{(i-1)h}^{t} \sigma'(X(s);\theta)\mu(X(s);\theta)ds + \frac{1}{2}\int_{(i-1)h}^{t} \sigma''(X(s);\theta)\sigma^2(X(s);\theta)ds \\
&\quad + \int_{(i-1)h}^{t} \sigma'(X(s);\theta)\sigma(X(s);\theta)dB(s),
\end{aligned}$$

with $\sigma'(X_{(i-1)h};\theta) = [\partial\sigma(X;\theta)/\partial X]_{X=X_{(i-1)h}}$. Substituting these expressions into (24) we obtain

$$\begin{aligned}
X_{ih} - X_{(i-1)h} &= \mu(X_{(i-1)h};\theta)h + \sigma(X_{(i-1)h};\theta)\int_{(i-1)h}^{ih} dB\,(t) \\
&\quad + \int_{(i-1)h}^{ih}\int_{(i-1)h}^{t} \sigma'(X(s);\theta)\sigma(X(s);\theta)dB(s)dB\,(t) + R, \qquad (25)
\end{aligned}$$

where R is a remainder of smaller order. Upon further use of the Ito formula on the penultimate term of (25), we obtain the following refinement of the Euler approximation

$$\begin{aligned}
X_{ih} - X_{(i-1)h} &\simeq \mu(X_{(i-1)h};\theta)h + \sigma(X_{(i-1)h};\theta)\int_{(i-1)h}^{ih} dB\,(t) + \\
&\quad \sigma'(X_{(i-1)h};\theta)\sigma(X_{(i-1)h};\theta)\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t} dB(s)dB\,(t)\,,
\end{aligned}$$

The multiple stochastic integral has the following reduction

$$\begin{aligned}\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t} dB(s)dB(t) &= \int_{(i-1)h}^{ih}\left(B(t)-B_{(i-1)h}\right)dB(t)\\ &= \int_{(i-1)h}^{ih} B(t)dB(t) - B_{(i-1)h}\left(B_{ih}-B_{(i-1)h}\right)\\ &= \frac{1}{2}\left\{\left(B_{ih}^2-B_{(i-1)h}^2\right)-h\right\} - B_{(i-1)h}\left(B_{ih}-B_{(i-1)h}\right)\\ &= \frac{1}{2}\left\{\left(B_{ih}-B_{(i-1)h}\right)^2-h\right\},\end{aligned}$$

Then the refined Euler approximation can be written as

$$\begin{aligned}X_{ih}-X_{(i-1)h} &\simeq \mu(X_{(i-1)h};\theta)h+\sigma(X_{(i-1)h};\theta)\left(B_{ih}-B_{(i-1)h}\right)\\ &\quad+\sigma'(X_{(i-1)h};\theta)\sigma(X_{(i-1)h};\theta)\frac{1}{2}\left\{\left(B_{ih}-B_{(i-1)h}\right)^2-h\right\}\\ &= \left\{\mu(X_{(i-1)h};\theta)-\frac{1}{2}\sigma'(X_{(i-1)h};\theta)\sigma(X_{(i-1)h};\theta)\right\}h+\sigma(X_{(i-1)h};\theta)\left(B_{ih}-B_{(i-1)h}\right)\\ &\quad+\frac{1}{2}\sigma'(X_{(i-1)h};\theta)\sigma(X_{(i-1)h};\theta)\left(B_{ih}-B_{(i-1)h}\right)^2\end{aligned}$$

The approach to such refinements is now very well developed in the numerical analysis literature and higher order developments are possible - see Kloeden and Platen (1999) for an extensive review.

It is convenient to write $B_{ih}-B_{(i-1)h}=\sqrt{h}\epsilon_i$ where ϵ_i is standard Gaussian. Then, the Milstein approximation to model (1) produces the following discrete time model:

$$X_{ih}=X_{(i-1)h}+\mu(X_{(i-1)h},\theta)h-g(X_{(i-1)h},\theta)h+\sigma(X_{(i-1)h},\theta)\sqrt{h}\epsilon_i+g(X_{(i-1)h},\theta)h\epsilon_i^2, \quad (26)$$

where

$$g(X_{(i-1)h},\theta)=\frac{1}{2}\sigma'(X_{(i-1)h};\theta)\sigma(X_{(i-1)h};\theta).$$

While Elerian (1998) used the Milstein scheme in connection with a simulation based approach, Tse, Zhang and Yu (2004) used the Milstein scheme in a Bayesian context. Both papers document some improvement from the Milstein scheme over the Euler scheme.

Kessler (1997) advocated approximating the transition density using a Gaussian density whose conditional mean and variance are obtained using higher order Taylor expansions. For example, the second-order approximation leads to the following discrete time model:

$$X_{ih}=\hat{\mu}(X_{(i-1)h};\theta)+\hat{\sigma}(X_{(i-1)h};\theta)\epsilon_i, \quad (27)$$

where

$$\begin{aligned}\hat{\mu}(X_{(i-1)h};\theta) &= X_{(i-1)h} + \mu(X_{(i-1)h};\theta)h + \\ &\left(\mu(X_{(i-1)h};\theta)\mu'(X_{(i-1)h};\theta) + \frac{\sigma^2(X_{(i-1)h};\theta)\mu''(X_{(i-1)h};\theta)}{2}\right)\frac{h}{2}\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}^2(X_{(i-1)h};\theta) &= X^2_{(i-1)h} + \left(2\mu(X_{(i-1)h};\theta)X_{(i-1)h} + \sigma^2(X_{(i-1)h};\theta)\right)h \\ &= \{2\mu(X_{(i-1)h};\theta)(2\mu'(X_{(i-1)h};\theta)X_{(i-1)h} + \mu(X_{(i-1)h};\theta) \\ &\quad +\sigma(X_{(i-1)h};\theta)\sigma'(X_{(i-1)h};\theta)) + \sigma^2(X_{(i-1)h};\theta) \times \\ &\quad [\mu''(X_{(i-1)h};\theta)X_{(i-1)h} + 2\mu(X_{(i-1)h};\theta) + (\sigma'(X_{(i-1)h};\theta))^2 \\ &\quad +\sigma(X_{(i-1)h};\theta)\sigma'(X_{(i-1)h};\theta)]\}\frac{h^2}{2} - \hat{\mu}^2(X_{(i-1)h};\theta).\end{aligned}$$

Nowman (1997) suggested an approach which assumes that the conditional volatility remains unchanged over the unit intervals, $[(i-1)h, ih)$, $i = 1, 2..., N$. In particular, he approximates the model:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma(X(t),\theta)dB(t) \tag{28}$$

by

$$dX(t) = \kappa(\mu - X(t))dt + \sigma(X_{(i-1)h};\theta)dB(t), \quad (i-1)h \le t < ih. \tag{29}$$

It is known from Phillips (1972) and Bergstrom (1984) that the exact discrete model of (29) has the form

$$X_{ih} = e^{-\kappa h}X_{(i-1)h} + \mu\left(1 - e^{-\kappa h}\right) + \sigma(X_{(i-1)h};\theta)\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}\epsilon_i, \tag{30}$$

where $\epsilon_i \sim$ i.i.d. $N(0,1)$. With this approximation, the Gaussian ML method can be used to estimate equation (30) directly. This method also extends in a straightforward way to multivariate systems. The Nowman procedure can be understood as applying the Euler scheme to the diffusion term over the unit interval. Compared with the Euler scheme where the approximation is introduced to both the drift function and the diffusion function, the Nowman method can be expected to reduce some of the discretization bias, as the treatment of the drift term does not involve an approximation at least in systems with linear drift.

Nowman's method is related to the local linearization method proposed by Shoji and Ozaki (1997, 1998) for estimating diffusion processes with a constant diffusion function and a possible nonlinear drift function, that is

$$dX(t) = \mu(X(t);\theta)dt + \sigma dB(t). \tag{31}$$

While Nowman approximates the nonlinear diffusion term, Shoji and Ozaki (1998) approximate the drift term, both by a locally linear function. The local linearization method can be used to estimate a diffusion process with a nonlinear diffusion function, provided that the process can be first transformed to make the diffusion function constant. This is achieved by the so-called Lamperti transform which will be explained in detailed below.

While all these refinements offer some improvements over the Euler method, with a fixed h, all the estimators remain inconsistent. As indicated, the magnitude of the inconsistency or bias may analyzed in terms of its order of magnitude as $h \to 0$. This appears only to have been done by Sargan (1974), Phillips (1974) and Lo (1988) for linear systems and some special cases.

3.2 Closed-form Approximations

The approaches reviewed above seek to approximate continuous time models by discrete time models, the accuracy of the approximations depending on the sampling interval h. Alternatively, one can use closed-form sequences to approximate the transition density itself, thereby developing an approximation to the likelihood function. Two different approximation mechanisms have been proposed in the literature. One is based on Hermite polynomial expansions whereas the other is based on the saddlepoint approximation.

3.2.1 Hermite Expansions

This approach was developed in Aït-Sahalia (2002) and illustrated in Aït-Sahalia (1999). Before obtaining the closed-form expansions, a Lamperti transform (mentioned earlier) is performed on the continuous time model so that the diffusion function becomes a constant. The transformation has the form $Y(t) = G(X(t))$, where $G'(x) = 1/\sigma(x;\cdot)$. The transformation is variance stabilizing and leads to another diffusion $Y(t)$, which by Ito's lemma can be shown to satisfy the stochastic differential equation

$$dY(t) = \mu_Y(Y(t);\theta)dt + dB(t), \tag{32}$$

where

$$\mu_Y(Y(t);\theta) = \frac{\mu(G^{-1}(Y);\theta)}{\sigma(G^{-1}(Y);\theta)} - \frac{1}{2}\sigma'(G^{-1}(Y);\theta).$$

Based on a Hermite polynomial expansion of the transition density $p(Y_{ih}|Y_{(i-1)h},\theta)$ around

the normal distribution, one gets

$$p(Y_{ih}|Y_{(i-1)h},\theta) \approx h^{-1/2}\phi\left(\frac{Y_{ih}-Y_{(i-1)h}}{h^{1/2}}\right)\exp\left(\int_{Y_{(i-1)h}}^{Y_{ih}}\mu_Y(\omega;\theta)d\omega\right)\sum_{k=0}^{K}c_k(Y_{ih}|Y_{(i-1)h};\theta)\frac{h^k}{k!}, \tag{33}$$

where $\phi(\cdot)$ is the standard normal density function, $c_0(Y_{ih}|Y_{(i-1)h}) = 1$,

$$\begin{aligned} c_j(Y_{ih}|Y_{(i-1)h}) &= j(Y_{ih}-Y_{(i-1)h})^{-j}\int_{Y_{(i-1)h}}^{Y_{ih}}(\omega-Y_{(i-1)h})^{j-1}\{\lambda_{Y_{ih}}(\omega;\theta)c_{j-1}(\omega|Y_{(i-1)h};\theta) \\ &\quad +\frac{1}{2}\partial^2 c_{j-1}(\omega|Y_{(i-1)h};\theta)/\partial\omega^2\}d\omega,\ \forall j\geq 1, \end{aligned}$$

and

$$\lambda_Y(y;\theta) = -\frac{1}{2}\left(\mu_Y^2(y;\theta)+\partial\mu_Y(y;\theta)/\partial y\right).$$

Under some regular conditions, Aït-Sahalia (2002) showed that when $K\to\infty$, the Hermite expansions (i.e., the right hand right in Equation (33)) approaches the true transition density. When applied to various interest rate models, Aït-Sahalia (1999) has found negligible approximation errors even for small values of K. Another advantage of this approach is that it is in closed-form and hence numerically tractable.

As noted in Durham and Gallant (2002), there are some drawbacks in this method. First, when the Lamperti transform is not feasible, the Hermite expansions are not possible. As a result, some interesting continuous time models cannot be estimated by this approach. Second, it is not clear how to apply the method to latent variable models.

3.2.2 Saddlepoint Approximations

The leading term in the Hermite expansions is normal whose tails may be too thin and the shape too symmetric relative to the true transition density. When this is the case, a moderately large value of K may be needed to ensure a good approximation of the Hermite expansion. An alternative approach is to choose a better approximating distribution as the leading term. One way to achieve this is to use a saddlepoint approximation.

The idea of the saddlepoint approximations is to approximate the conditional cumulant generating function of the transition density by means of a suitable expansion, followed by a careful choice of integration path in the integral that defines the transition density so that most of the contribution to the integral comes from integrating in the immediate neighborhood of a saddlepoint. The method was originally explored in statistics by Daniels (1953). Phillips (1978) developed a saddlepoint approximation to the distribution of ML estimator of the coefficient in discrete time first order autoregression, while Holly and Phillips (1979) proposed

saddlepoint approximations for the distributions of k-class estimators of structural coefficients in simultaneous equation systems. There has since been a great deal of interest in the method in statistics - see Reid (1988) and Field and Ronchetti (1990) for partial overviews of the field. Aït-Sahalia and Yu (2006) proposed the use of saddlepoint approximations to the transition density of continuous time models, which we now consider.

Let $\varphi_{X_{(i-1)h}}(u;\theta)$ be the conditional characteristic function corresponding to the transition density, viz.,

$$\varphi_{X_{(i-1)h}}(u;\theta) = E[\exp(uX_{ih}|X_{(i-1)h})].$$

The conditional cumulant generating function is

$$K_{X_{(i-1)h}}(u;\theta) = \ln(\varphi_{X_{(i-1)h}}(u;\theta)). \tag{34}$$

The transition density has the following integral representation by Fourier inversion:

$$\begin{aligned} p(X_{ih}|X_{(i-1)h},\theta) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp(-iX_{ih}u)\varphi_{X_{(i-1)h}}(iu;\theta)du \\ &= \frac{1}{2\pi}\int_{\hat{u}-i\infty}^{\hat{u}+i\infty}\exp(-uX_{ih})\varphi_{X_{(i-1)h}}(u;\theta)du \\ &= \frac{1}{2\pi}\int_{\hat{u}-i\infty}^{\hat{u}+i\infty}\exp(K_{X_{(i-1)h}}(u;\theta)-uX_{ih})du \end{aligned} \tag{35}$$

Applying a Taylor expansion to $K_{X_{(i-1)h}}(u;\theta) - uX_{ih}$ around the saddlepoint $\hat{u}$, one gets

$$\begin{aligned} K_{X_{(i-1)h}}(u;\theta) - uX_{ih} &= K_{X_{(i-1)h}}(\hat{u};\theta) - \hat{u}X_{ih} - \frac{1}{2}\frac{\partial^2 K_{X_{(i-1)h}}(\hat{u};\theta)}{\partial u^2}\nu \\ &\quad -\frac{1}{6}\frac{\partial^3 K_{X_{(i-1)h}}(\hat{u};\theta)}{\partial u^3}i\nu^3 + O(\nu^4). \end{aligned}$$

Substituting this expansion to (34), one obtains a saddlepoint approximation to the integral, which involves the single leading term of the form

$$\frac{\exp(K_{X_{(i-1)h}}(\hat{u};\theta) - uX_{ih})}{\sqrt{2\pi}\left(\frac{\partial^2 K_{X_{(i-1)h}}(\hat{u};\theta)}{\partial u^2}\right)^{1/2}},$$

and higher order terms of small order. As shown in Daniels (1954), the method has the advantage of producing a smaller relative error than Edgeworth and Hermite expansions.

When applying this method to transition densities for some continuous time models that are widely used in finance, Aït-Sahalia and Yu (2006) have found very small approximation

errors. The method requires the saddlepoint to be analytically available or at least numerically calculable, an approach considered in Phillips (1984) that widens the arena of potential application. The saddlepoint method also requires the moment generating function of the transition density to exist, so that all moments of the distribution must be finite and heavy tailed transition distributions are therefore excluded. Multivariate extensions are possible using extensions of the saddlepoint method to this case - see Phillips (1980,1984), Tierney and Kadane (1986) and McCullagh (1987).

3.3 Simulated Infill ML Methods

As explained above, the Euler scheme introduces discretization bias. The magnitude of the bias is determined by h. When the sampling interval is arbitrarily small, the bias becomes negligible. One way of making the sampling interval arbitrarily small is to partition the original interval, say $[(i-1)h, ih]$, so that the new subintervals are sufficiently fine for the discretization bias to be negligible. By making the subintervals smaller, one inevitably introduces latent (that is, unobserved) variables between $X_{(i-1)h}$ and X_{ih}. To obtain the required transition density $p(X_{ih}|X_{(i-1)h}, \theta)$, these latent observations must be integrated out. When the partition becomes finer, the discretization bias is closer to 0 but the required integration becomes high dimensional. We call this approach to bias reduction the simulated infill ML method.

To fix ideas, suppose $M-1$ auxiliary points are introduced between $(i-1)h$ and ih, i.e.,

$$((i-1)h \equiv)\tau_0, \tau_1, \cdots, \tau_{M-1}, \tau_M (\equiv ih).$$

The Markov property implies that

$$\begin{aligned} p(X_{ih}|X_{(i-1)h};\theta) &= \int \cdots \int p(X_{\tau_M}, X_{\tau_{M-1}}, \cdots, X_{\tau_1}|X_{\tau_0};\theta) dX_{\tau_1} \cdots dX_{\tau_{M-1}} \\ &= \int \cdots \int \prod_{m=1}^{M} p(X_{\tau_m}|X_{\tau_{m-1}};\theta) dX_{\tau_1} \cdots dX_{\tau_{M-1}}. \end{aligned} \tag{36}$$

The idea behind the simulated infill ML method is to approximate the densities $p(X_{\tau_m}|X_{\tau_{m-1}};\theta)$ (step 1) and then evaluate the multidimensional integral using importance sampling techniques (step 2). Among the class of simulated infill ML methods that have been suggested, Pedersen (1995) is one of the earliest contributions.

Pedersen suggested approximating the latent transition densities $p(X_{\tau_m}|X_{\tau_{m-1}};\theta)$ based on the Euler scheme and approximating the integral by drawing samples of $(X_{\tau_{M-1}}, \cdots, X_{\tau_1})$ via simulations from the Euler scheme. That is, the importance sampling function is the mapping from $(\epsilon_1, \epsilon_2, \cdots, \epsilon_{M-1}) \mapsto (X_{\tau_1}, X_{\tau_2}, \cdots, X_{\tau_{M-1}})$ given by the Euler scheme:

$$X_{\tau_{m+1}} = X_{\tau_m} + \mu(X_{\tau_m};\theta)h/M + \sigma(X_{\tau_m},\theta)\sqrt{h/M}\epsilon_{m+1},\ m = 0, \cdots, M-2,$$

where $(\epsilon_1, \epsilon_2, \cdots, \epsilon_{M-1})$ is a multivariate standard normal.

As noted in Durham and Gallant (2002), there are two sources of approximation error in Pedersen's method. One is the (albeit reduced) discretization bias in the Euler scheme. The second is due to the Monte Carlo integration. These two errors can be further reduced by increasing the number of latent infill points and the number of simulated paths, respectively. However, the corresponding computational cost will inevitably be higher.

In order to reduce the discretization bias in step 1, Elerian (1998) suggested replacing the Euler scheme with the Milstein scheme while Durham and Gallant advocated using a variance stablization transformation, i.e., applying the Lamperti transform to the continuous time model. Certainly, any method that reduces the discretization bias can be used. Regarding step 2, Elerian et al (2001) argued that the importance sampling function of Pedersen ignores the end-point information, X_{τ_M}, and Durham and Gallant (2002) showed that Pedersen's importance function draws most samples from regions where the integrand has little mass. Consequently, Pedersen's method is simulation-inefficient.

To improve the efficiency of the importance sampler, Durham and Gallant (2002) considered the following importance sampling function

$$X_{\tau_{m+1}} = X_{\tau_m} + \frac{X_{ih} - X_{\tau_m}}{ih - \tau_m} h/M + \sigma(X_{\tau_m}, \theta)\sqrt{h/M}\epsilon_{m+1},\ m = 0, \cdots, M-2,$$

where $(\epsilon_1, \epsilon_2, \cdots, \epsilon_{M-1})$ is a multivariate standard normal. Loosing speaking, this is a Brownian bridge because it starts from $X_{(i-1)h}$ at $(i-1)h$ and is conditioned to terminate with X_{ih} at ih.

Another importance sampling function proposed by Durham and Gallant (2002) is to draw $X_{\tau_{m+1}}$ from the density $N(X_{\tau_m} + \tilde{\mu}_m h/M, \tilde{\sigma}_m^2 h/M)$ where $\tilde{\mu}_m = (X_{\tau_M} - X_{\tau_m})/(ih - \tau_m)$, $\tilde{\sigma}_m^2 = \sigma^2(X_{\tau_m})(M-m-1)/(M-m)$.

Elerian et al. (2001) proposed a more efficient importance function which is based on the following tied-down process:

$$p(X_{\tau_1}, \cdots, X_{\tau_{M-1}} | X_{\tau_0}, X_{\tau_M}).$$

In particular, they proposed using the Laplace approximation (c.f., Phillips, 1984; Tierney and Kadane, 1986) to the tied-down process. That is, they used the distributional approximation $(X_{\tau_1}, \cdots, X_{\tau_{M-1}}) \sim N(\mathbf{x}^*, \Sigma^*)$ where

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \ln p(X_{\tau_1}, \cdots, X_{\tau_{M-1}} | X_{\tau_0}, X_{\tau_M})$$

$$\Sigma^2 = -\left[\frac{\partial^2 \ln p(X^*_{\tau_1}, \cdots, X^*_{\tau_{M-1}} | X_{\tau_0}, X_{\tau_M})}{\partial \mathbf{x}' \partial \mathbf{x}}\right]^{-1},$$

where $\mathbf{x} = (X_{\tau_1}, \cdots, X_{\tau_{M-1}})'$.

Durham and Gallant (2002) compared the performance of these three importance functions relative to Pedersen (1995) and found that all these methods deliver substantial improvements.

3.4 Other Approaches

3.4.1 Numerical ML

While the transition density may not have a closed-form expression for a continuous time model, it must satisfy the Fokker-Planck-Komogorov (also known as "forward") equation. That is,

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial y^2}.$$

where $p(y, t|x, s)$ is the transition density. Solving the partial differential equation numerically at $y = X_{ih}, x = X_{(i-1)h}$ yields the transition density. This is approach proposed by Lo (1988). Similarly, one can numerically solve the "backward" equation

$$\frac{\partial p}{\partial s} = -\frac{1}{2}\frac{\partial^2 p}{\partial x^2}.$$

Obviously, solving these two partial differential equations numerically can be computationally demanding. Consequently, this approach has been little used in practical work.

3.4.2 An Exact Gaussian Method based on Time Changes

Yu and Phillips (2001) developed an exact Gaussian method to estimate continuous time models with a linear drift function of the following form:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma(X(t); \theta)dB(t), \tag{37}$$

The approach is based on the idea that any continuous time martingale can be written as a Brownian motion after a suitable time change. That is, when we adjust from chronological time in a local martingale M_t to time based on the evolution of the quadratic variation process $[M]_t$ of M, we have the time change given by $T_t = \inf\{s|[M]_s > t\}$ and the process transforms to a Brownian motion (called DDS Brownian motion) so that $M_t = W_{[M]_t}$, where W is standard Brownian motion.

To see how this approach can be used to estimate equation (37), first write (37) as

$$X(t+\delta) = e^{-\kappa h}X(t) + \mu\left(1 - e^{-\kappa h}\right) + \int_0^\delta \sigma e^{-\kappa(\delta-\tau)}\sigma(t+\tau)dB(\tau), \forall\delta > 0. \tag{38}$$

Define $M(\delta) = \sigma \int_0^\delta e^{-\kappa(\delta-\tau)}\sigma(t+\tau)dB(\tau)$, which is a continuous martingale with quadratic variation process

$$[M]_\delta = \sigma^2 \int_0^\delta e^{-2\kappa(\delta-\tau)}\sigma^2(t+\tau)d\tau. \tag{39}$$

To construct a DDS Brownian motion to represent $M(\delta)$, one can construct a sequence of positive numbers $\{\delta_j\}$ which deliver the required time changes. For any fixed constant $a > 0$, let

$$\delta_{j+1} = \inf\{s|[M_j]_s \geq a\} = \inf\{s|\sigma^2 \int_0^s e^{-2\kappa(s-\tau)}\sigma^2(t_j+\tau)d\tau \geq a\}. \tag{40}$$

Next, construct a sequence of time points $\{t_j\}$ using the iterations $t_{j+1} = t_j + \delta_{j+1}$ with t_1 assumed to be 0. Evaluating equation (38) at $\{t_j\}$, we have

$$X_{t_{j+1}} = \mu\left(1 - e^{-\kappa\delta_{j+1}}\right) + e^{-\kappa\delta_{j+1}}X_{t_j} + M(\delta_{j+1}). \tag{41}$$

where $M(\delta_{j+1}) = W_{[M]_{\delta_{j+1}}} = W_a \equiv N(0,a)$ is the DDS Brownian motion. Hence, equation (41) is an exact discrete model with Gaussian disturbances and can be estimated directly by ML conditional on the sequence of time changes. Of course, since the new sequence of time points $\{t_j\}$ is path dependent, this approach does not deliver the true likelihood. Also, since a continuous record of observations is not available, the time points $\{t_j\}$ must be approximated.

4 Approximate ML Methods Based on the Continuous Record Likelihood and Realized Volatility

While (1) is formulated in continuous time, the sample data are always collected at discrete points in time or over discrete intervals in the case of flow data. One may argue that for highly liquid financial assets, the sampled data are so frequently observed as to be nearly continuously available. This is especially true for some tick-by-tick data. Unfortunately, at the highest frequencies, continuous time models such as that given by (1) are often bad descriptions of reality. One reason for the discrepancy is the presence of market microstructure noise, due to trading frictions, bid-ask bounces, recording errors and other anomalies. As a result of these noise effects, the exact ML method based on the continuous record likelihood that was reviewed in Section 2.2 is not applicable.

An alternative approach that is available in such situations was developed in Phillips and Yu (2005c) and involves a two-step procedure to estimate the underlying continuous time model that makes use of the empirical quadratic variation process. To explain the method, suppose the model has the form

$$dX(t) = \mu(X(t);\theta_1)dt + \sigma(X(t);\theta_2)dB(t), \tag{42}$$

Note that in this specification the vector of parameters θ_2 in the diffusion function is separated from the parameter vector, θ_1, that appears in the drift function. The reason for this distinction will become clear below.

In the first step, Phillips and Yu (2005c) propose to estimate parameters in the diffusion function from the empirical quadratic variation process or so-called realized volatility. The approach is justified by the fact that realized volatility is a natural consistent estimate of quadratic variation and, with certain modifications, can be made consistent even in the presence of microstructure noise effects. Also, realized volatility has convenient distributional characteristics that are determined asymptotically by (functional) central limit theory, as derived by Jacod (1994) and Barndorff-Nielsen and Shephard (2002).

To proceed, assume that X_t is observed at the following times

$$t = \underbrace{h, 2h, \cdots, Mh(=\frac{T}{K})}, \underbrace{(M+1)h, \cdots, 2Mh(=\frac{2T}{K})}, \cdots, nh(=T),$$

where $n = KM$ with K a fixed and positive integer, T is the time span of the data, h is the sampling frequency, and $M = O(n)$. Phillips and Yu constructed the non-overlapping K subsamples

$$((k-1)M+1)h, \cdots, kMh, \quad \text{where } k = 1, \cdots, K,$$

so that each sub-sample has M observations over the interval $((k-1)\frac{T}{K}, k\frac{T}{K}]$. For example, if ten years of weekly observed data are available and we split the data into ten blocks, then $T = 10$, $h = 1/52$, $M = 52$, $K = 10$. The total number of observations is 520 and the number of observations contained in each block is 52.

As $h \to 0$, $n = \frac{T}{h} \to \infty$ and $M \to \infty$,

$$\sum_{i=2}^{M_h} (X_{(k-1)M_h+ih} - X_{(k-1)M_h+(i-1)h})^2 \xrightarrow{p} [X]_{k\frac{T}{K}} - [X]_{(k-1)\frac{T}{K}}, \tag{43}$$

and

$$\frac{\ln(\sum_{i=2}^{M}(X_{(k-1)M+ih} - X_{(k-1)M+(i-1)h})^2 - \ln([X]_{k\frac{T}{K}} - [X]_{(k-1)\frac{T}{K}})}{s_k} \xrightarrow{d} N(0,1), \tag{44}$$

where

$$s_k = \min\left\{\sqrt{\frac{r_k^2}{(\sum_{i=2}^{M_h}(X_{(k-1)M+ih} - X_{(k-1)M+(i-1)h})^2)^2}}, \frac{2}{M}\right\}, \tag{45}$$

for $k = 1, \cdots, K$, and $[X]_T$ is the quadratic variation of X which can be consistently estimated by the empirical counterpart $[X_h]_T$ defined as

$$[X_h]_T = \sum_{i=2}^{n} (X_{ih} - X_{(i-1)h})^2.$$

The limit (43) follows by virtue of the definition of quadratic variation. The central limit theorem (CLT) (44) is based on the asymptotic theory of Barndorff-Nielsen and Shephard (2005), which involves a finite sample correction (45) on some important earlier limit theory contributions made by Jacod (1994) and Barndorff-Nielsen and Shephard (2002).

Based on the CLT (44), θ_2 can be estimated in the first stage by running a (nonlinear) least squares regression of the standardized realized volatility

$$\frac{\ln\left(\sum_{i=2}^{M}(X_{(k-1)M+ih} - X_{(k-1)M+(i-1)h})^2\right)}{s_k} \tag{46}$$

on the standardized diffusion function

$$\frac{\ln\left([X]_{k\frac{T}{K}} - [X]_{(k-1)\frac{T}{K}}\right)}{s_k} = \frac{\ln\left(\int_{(k-1)\frac{T}{K}}^{k\frac{T}{K}} \sigma^2\left(X_t;\theta_2\right) dt\right)}{r_k} \tag{47}$$

$$\simeq \frac{\ln\left(\sum_{i=2}^{M} \sigma^2\left(X_{(k-1)M_h+(i-1)h};\theta_2\right) h\right)}{r_k} \tag{48}$$

for $k = 1, \cdots, K$. This produces a consistent estimate $\hat{\theta}_2$ of θ_2. In the second stage, the approximate continuous record or infill log-likelihood function (AIF) is maximized with respect to θ_1

$$\ell_{AIF}(\theta_1) = \sum_{i=2}^{n} \frac{\mu(X_{(i-1)h};\theta_1)}{\sigma^2(X_{(i-1)h};\hat{\theta}_2)}(X_{ih} - X_{(i-1)h}) - \frac{h}{2}\sum_{i=2}^{n} \frac{\mu^2(X_{(i-1)h};\theta_1)}{\sigma^2(X_{(i-1)h};\hat{\theta}_2)}. \tag{49}$$

The procedure is discussed more fully in Phillips and Yu (2005c).

To illustrate the two-stage method, we consider the following specific models.

1. **Vasicek model (5):** Since there is only one parameter in the diffusion function, one could choose $M = 1$. As a result, the first stage estimation gives the following estimator for σ,

$$\hat{\sigma} = \sqrt{\frac{[X_h]_T}{T}}, \tag{50}$$

and the approximate infill log-likelihood function is given by

$$\ell_{AIF}(\kappa,\mu)=\sum_{i=2}^{n}\kappa(\mu-X_{(i-1)h})(X_{ih}-X_{(i-1)h})-\frac{h}{2}\sum_{i=2}^{n}\kappa^2(\mu-X_{(i-1)h})^2. \tag{51}$$

2. **Square root model (8):** With $M=1$, the first stage estimation gives the following estimator for σ.

$$\hat{\sigma}=\sqrt{\frac{[X_h]_T}{h\sum_{i=1}^{n_h}X_{(i-1)h}}}. \tag{52}$$

The approximate infill log-likelihood function is given by

$$\ell_{AIF}(\kappa,\mu)=\sum_{i=2}^{n}\frac{\kappa(\mu-X_{(i-1)h})}{\hat{\sigma}^2X_{(i-1)h}}(X_{ih}-X_{(i-1)h})-\frac{h}{2}\sum_{i=2}^{n}\frac{\kappa^2(\mu-X_{(i-1)h})^2}{\hat{\sigma}^2X_{(i-1)h}}. \tag{53}$$

5 Monte Carlo Simulations

This section reports the results of a Monte Carlo experiment designed to compare the performance of the various ML estimation methods reviewed in the previous sections. In the experiment, the true generating process is assumed to be the CIR model of short term interest rates of the form

$$dX(t)=\kappa(\mu-X(t))dt+\sigma\sqrt{X(t)}\,dB(t),$$

where $\kappa=0.1,\mu=0.1,\sigma=0.1$. Replications involving 1000 samples, each with 120 monthly observations (ie $h=1/12$), are simulated from the true model. The parameter settings are realistic to those in many financial applications and the sample period covers 10 years.

It is well-known that κ is difficult to estimate with accuracy whereas the other two parameters, especially σ, are much easier to estimate (Phillips and Yu, 2005a, b) and extensive results are already in the literature. Consequently, we only report estimates of κ in the present Monte Carlo study. In total, we employ six estimation methods, namely, exact ML, the Euler scheme, the Milstein scheme, the Nowman method, the infill method, and the Hermite expansion (with $K=1$).

Table 1 reports the means, standard errors, and root mean square errors (RMSEs) for all these cases. The exact ML estimator is calculated for comparison purposes. Since the other estimators are designed to approach to the exact ML estimator, we also report the means and the standard errors of the differences between the exact ML estimator and the alternative estimators.

Table 1

Exact and Approximate ML Estimation and Bias Reduced Estimation of κ									
True Value $\kappa = 0.1$									
Method	Exact	Euler	Milstein	Nowman	In-fill	Hermite	Jackk (m=2)	Jackk (m=3)	Ind Inf
Mean	.2403	.2419	.2444	.2386	.2419	.2413	.1465	.1845	.1026
Std error	.2777	.2867	.2867	.2771	.2867	.2870	.3718	.3023	.2593
RMSE	.3112	.3199	.3210	.3098	.3199	.3199	.3747	.3139	.2594
Mean of diff	NA	.0016	.0041	-.0017	.0016	.0010	NA	NA	NA
Std error of diff	NA	.0500	.0453	.0162	.0500	.0503	NA	NA	NA

Note: A square-root model with $\kappa = 0.1, \mu = 0.1, \sigma = 0.1$ is used to simulate 120 monthly observations for each of the 1,000 replications. Various methods are used to estimate κ.

Several conclusions can be drawn from the table (Note the true value of $\kappa = 0.1$.) First, the ML estimator of κ is upward biased by more than 140%, consistent with earlier results reported in Phillips and Yu (2005a, b). This result is also consistent with what is known about dynamic bias in local-to-unity discrete time autoregressive models. Second, all the approximation-based ML methods perform very similar to the exact ML method, and hence, all inherit substantial estimation bias from the exact ML method that these methods seek to imitate. Indeed, compared to the estimation bias in exact ML, the bias that is induced purely by the approximations is almost negligible. Third, relative to the Euler scheme, the Milstein scheme fail to offer any improvements in terms of both mean and variation while Nowman's method offers slight improvements in terms of variation and root mean squared error (RMSE). Although the Hermite expansions do not perform well among the approximation-based methods examined here, it is important to recognize that the estimator is nonetheless very close to the exact ML estimator. Further improvements in the quality of this approximation of ML can be achieved by increasing the value of K, as pointed out in Aït-Sahalia (2002), although such improvements do not help to remove the finite sample bias of the ML procedure.

6 Estimation Bias Reduction Techniques

It has frequently been argued in the continuous time finance literature that ML should be the preferred choice of estimation method. The statistical justification for this choice is the generality of the ML approach and its good asymptotic properties of consistency and efficiency.

Moreover, since sample sizes in financial data applications are typically large[2], it is often expected that these good asymptotic properties will be realized in finite samples. However, for many financial time series, the asymptotic distribution of the ML estimator often turns out to be a poor approximation to the finite sample distribution, which may be badly biased even when the sample size is large. This is especially the case in the commonly occurring situation of drift parameter estimation in models where the process is nearly a martingale. From the practical viewpoint, this is an important shortcoming of the ML method. The problem of estimation bias turns out to be of even greater importance in the practical use of econometric estimates in asset and option pricing, where there is nonlinear dependence of the pricing functional on the parameter estimates, as shown in Phillips and Yu (2005a). This nonlinearity seems to exacerbate bias and makes good bias correction more subtle.

In the following sections we describe two different approaches to bias correction. The first of these is a simple procedure based on Quenouille's (1956) jackknife. To improve the finite sample properties of the ML estimator in continuous time estimation and in option pricing applications, Phillips and Yu (2005a) proposed a general and computationally inexpensive method of bias reduction based on this approach. The second approach is simulation-based and involves the indirect inference estimation idea of Gourieroux et al (1993). Monfort (1996) proposed this method of bias corrected estimation in the context of nonlinear diffusion estimation.

6.1 Jackknife estimation

Quenouille (1956) proposed the jackknife as a solution to finite sample bias problems in parametric estimation contexts such as discrete time autoregressions. The method involves the systematic use of subsample estimates. To fix ideas, let N be the number of observations in the whole sample and decompose the sample into m consecutive subsamples each with ℓ observations, so that $N = m \times \ell$. The jackknife estimator of a certain parameter, θ, then utilizes the subsample estimates of θ to assist in the bias reduction process giving the jackknife estimator

$$\hat{\theta}_{jack} = \frac{m}{m-1}\hat{\theta}_N - \frac{\sum_{i=1}^{m}\hat{\theta}_{li}}{m^2 - m}, \tag{54}$$

where $\hat{\theta}_N$ and $\hat{\theta}_{li}$ are the estimates of θ obtained by application of a given method like the exact ML or approximate ML to the whole sample and the i'th sub-sample, respectively. Under quite general conditions which ensure that the bias of the estimates $(\hat{\theta}_N, \hat{\theta}_{li})$ can be expanded asymptotically in a series of increasing powers of N^{-1}, it can be shown that the bias in the jackknife estimate $\hat{\theta}_{jack}$ is of order $O(N^{-2})$ rather than $O(N^{-1})$.

[2]Time series samples of weekly data often exceed 500 and sample sizes are very much larger for daily and intradaily data.

The jackknife has several appealing properties. The first advantage is its generality. Unlike other bias reduction methods, such as those based on corrections obtained by estimating higher order terms in an asymptotic expansion of the bias, the jackknife technique does not rely (at least explicitly) on the explicit form of an asymptotic expansion. This means that it is applicable in a broad range of model specifications and that it is unnecessary to develop explicit higher order representations of the bias. A second advantage of the jackknife is that this approach to bias reduction can be used with many different estimation methods, including general methods like the exact ML method whenever it is feasible or approximate ML methods when the exact ML is not feasible. Finally, unlike many other bias correction methods, the jackknife is computationally much cheaper to implement. In fact, the method is not much more time consuming than the initial estimation itself. A drawback with jackknife is that it cannot completely remove the bias as it is only designed to decrease the order of magnitude of the bias.

Table 1 reports the results of the jackknife method applied with $m = 2, 3$ based on the same experimental design above. It is clear that the jackknife makes substantial reductions in the bias but this bias reduction comes with an increase in variance. However, a carefully designed jackknife method can reduce the RMSE.

6.2 Indirect inference estimation

The indirect inference (II) procedure, first introduced by Gouriéroux, Monfort, and Renault (1993), and independently proposed by Smith (1993) and Gallant and Tauchen (1996), can be understood as a generalization of the simulated method of moments approach of Duffie and Singleton (1993). It has been found to be a highly useful procedure when the moments and the likelihood function of the true model are difficult to deal with, but the true model is amenable to data simulation. Since many continuous time models are easy to simulate but present difficulties in the analytic derivation of moment functions and likelihood, the indirect inference procedure has some convenient advantages in working with continuous time models in finance. A carefully designed indirect inference estimator can also have good small sample properties, as shown by Gouriéroux, et al (2000) in the time series context and by Gouriéroux, Phillips and Yu (2005) in the panel context. The method therefore offers some interesting opportunities for bias correction and the improvement of finite sample properties in continuous time estimation.

Without loss of generality, we focus on the OU process. Suppose we need to estimate the parameter κ in the model

$$dX(t) = \kappa(\mu - X(t))dt + \sigma\, dB(t).$$

from observations $\mathbf{x} = \{X_h, \cdots, X_{Nh}\}$. An initial estimator of κ can be obtained, for example, by applying the Euler scheme to $\{X_h, \cdots, X_{Nh}\}$ (call it $\hat{\kappa}_N$). Such an estimator is inconsistent (due to the discretization error) and may be seriously biased (due to the poor finite sample property of ML in the low κ or near-unit-root case).

The indirect inference method makes use of simulations to remove the discretization bias. It also makes use of simulations to calibrate the bias function and hence requires neither the explicit form of the bias, nor the bias expansion. This advantage seems important when the computation of the bias expression is analytically involved, and it becomes vital when the bias and the first term of the bias asymptotic expansions are too difficult to compute explicitly.

The idea of indirect inference here is as follows. Given a parameter choice κ, we apply the Euler scheme with a much smaller step size than h (say $\delta = h/10$), which leads to

$$\tilde{X}^k_{t+\delta} = \kappa(\mu - \tilde{X}^k_t)h + \tilde{X}^k_t + \sigma\sqrt{\delta}\epsilon_{t+\delta},$$

where

$$t = \underbrace{0, \delta, \cdots, h(=10\delta)}, \underbrace{h+\delta, \cdots, 2h(=20\delta)}, 2h+\delta, \cdots, Nh.$$

This sequence may be regarded as a nearly exact simulation from the continuous time OU model for small δ. We then choose every $(h/\delta)^{th}$ observation to form the sequence of $\{\tilde{X}^k_{ih}\}_{i=1}^N$, which can be regarded as data simulated directly from the OU model with the (observationally relevant) step size h.

Let $\tilde{\mathbf{x}}^k(\kappa) = \{\tilde{X}^k_h, \cdots, \tilde{X}^k_{Nh}\}$ be data simulated from the true model, where $k = 1, \cdots, K$ with K being the number of simulated paths. It should be emphasized that it is important to choose the number of observations in $\tilde{\mathbf{x}}^k(\kappa)$ to be the same as the number of observations in the observed sequence $\mathbf{x}$ for the purpose of the bias calibration. Another estimator of κ can be obtained by applying the Euler scheme to $\{X^k_h, \cdots, X^k_{Nh}\}$ (call it $\tilde{\kappa}^k_N$). Such an estimator and hence the expected value of them across simulated paths is naturally dependent on the given parameter choice κ.

The central idea in II estimation is to match the parameter obtained from the actual data with that obtained from the simulated data. In particular, the II estimator of κ is defined as

$$\hat{\kappa}^{II}_{N,K} = \text{argmin}_\kappa \parallel \hat{\kappa}_N - \frac{1}{K}\sum_{h=1}^{K} \tilde{\kappa}^k_N(\kappa) \parallel, \tag{55}$$

where $\parallel \cdot \parallel$ is some finite dimensional distance metric. In the case where K tends to infinity, the II estimator is the solution of the limiting extremum problem

$$\hat{\kappa}^{II}_N = \text{argmin}_\kappa \parallel \hat{\kappa}_N - E(\tilde{\kappa}^k_N(\kappa)) \parallel . \tag{56}$$

This limiting extremum problem involves the so-called binding function

$$b_N(\kappa) = E(\tilde{\kappa}_N^k(\kappa)),$$

which is a finite sample functional relating the bias to κ. In the case where b_N is invertible, the indirect inference estimator is given by

$$\hat{\kappa}_N^{II} = b_N^{-1}(\hat{\kappa}_N).$$

The II estimation procedure essentially builds in a small-sample bias correction to parameter estimation, with the bias (in the base estimate, like ML) being computed directly by simulation.

Indirect inference has several advantages for estimating continuous time models. First, it overcomes the inconsistency problem that is common in many approximate ML methods. Second, the indirect inference technique calibrates the bias function via simulation and hence does not require, just like the jackknife method, an explicit form for the bias function or its expansion. Consequently, the method is applicable in a broad range of model specifications. Thirdly, indirect inference can be used with many different estimation methods, including the exact ML method or approximate ML methods, and in doing so will inherit the good asymptotic properties of these base estimators. For instance, it is well known that the Euler scheme offers an estimator which has very small dispersion relative to many consistent estimators and indirect inference applied to it should preserve its good dispersion characteristic while at the same time achieving substantial bias reductions. Accordingly, we expect indirect inference to perform very well in practice and in simulations on the basis of criteria such as RMSE, which take into account central tendency and variation. A drawback with indirect inference is that it is a simulation-based method and can be computationally expensive. However, with the continuing explosive growth in computing power, such a drawback is obviously of less concern.

Table 1 reports the results of the indirect inference method with $K = 1000$ based on the same experiment discussed earlier. Clearly, indirect inference is very successful in removing bias and the bias reduction is achieved without increasing the variance. As a result, the RMSE is greatly reduced.

7 Multivariate Continuous Time Models

Multivariate systems of stochastic differential equations may be treated in essentially the same manner as univariate models such as (1) and methods such as Euler-approximation-based ML methods and transition density-approximation-based ML methods continue to be applicable. The literature on such extensions is smaller, however, and there are more and more financial

data applications of multivariate systems at present; see, for example, Ghysels et al (1996) and Shephard (2005) for reviews of the stochastic volatility literature and Dai and Singleton (2002) for a review of the term structure literature.

One field where the literature on multivariate continuous time econometrics is well developed is macroeconomic modeling of aggregative behavior. These models have been found to provide a convenient mechanism for embodying economic ideas of cyclical growth, market disequilibrium and dynamic adjustment mechanisms. The models are often constructed so that they are stochastic analogues (in terms of systems of stochastic differential equations) of the differential equations that are used to develop the models in economic theory. The Bergstrom (1966) approximation, discussed in Section 3.1 above, was developed specifically to deal with such multiple equation systems of stochastic equations. Also, the exact discrete time model corresponding to a system of linear diffusions, extending the Vasicek model in Section 2.1, was developed in Phillips (1972, 1974) as the basis for consistent and efficient estimation of structural systems of linear diffusion equations using nonlinear systems estimation and Gaussian ML estimation.

One notable characteristic of such continuous time systems of equations is that there are many across-equation parameter restrictions. These restrictions are typically induced by the manner in which the underlying economic theory (for example, the theory of production involving a parametric production function) affects the formulation of other equations in the model, so that the parameters of one relation (the production relation) become manifest elsewhere in the model (such as wage and price determination, because of the effect of labor productivity on wages). The presence of these across-equation restrictions indicates that there are great advantages to the use of systems procedures, including ML estimation, in the statistical treatment of systems of stochastic differential equations.

While many of the statistical issues already addressed in the treatment of univariate diffusions apply in systems of equations, some new issues do arise. A primary complication is that of aliasing, which in systems of equations leads to an identification problem when a continuous system in estimated by a sequence of discrete observations at sampling interval h. The manifestation of this problem is evident in a system of linear diffusions for an $n-$ vector process $X(t)$ of the form

$$dX(t) = A(\theta_2) X(t) dt + \Sigma(\theta_2) dW(t),$$

where $A = A(\theta)$ is an $n \times n$ coefficient matrix whose elements are dependent on the parameter vector θ_1, $\Sigma = \Sigma(\theta_2)$ is a matrix of diffusion coefficients dependent on the parameter vector θ_2, and $W(t)$ is $n-$ vector standard Brownian motion. The exact discrete model corresponding

to this system has the form

$$X_{ih} = e^{hA(\theta_2)} X_{ih} + N\left(0, \int_0^h e^{sA(\theta_2)} \Sigma\left(\theta_2\right) e^{sA(\theta_2)'} ds\right),$$

and the coefficient matrix in this discrete time model involves the matrix exponential function $e^{hA(\theta_2)}$. However, there are in general, an infinite number of solutions (A) to the matrix exponential equation

$$e^{hA} = B^0 \tag{57}$$

where $B^0 = e^{hA^0} = e^{hA\left(\theta_2^0\right)}$ and θ_2^0 is the true value of θ_2. In fact, the solutions of the matrix equation (57) all have the form

$$A = A^0 + TQT^{-1},$$

where T is a matrix that diagonalizes A^0 (so that $T^{-1}AT = \text{diag}(\lambda_1, ..., \lambda_n)$, assuming that A^0 has distinct characteristics roots $\{\lambda_i : i = 1, ..., n\}$), Q is a matrix of the form

$$Q = \frac{2\pi i}{h} \begin{bmatrix} 0 & 0 & 0 \\ 0 & P & 0 \\ 0 & 0 & -P \end{bmatrix},$$

and P is a diagonal matrix with integers on the diagonal. The multiple solutions of (57) effectively correspond to aliases of A^0.

Fortunately, in this simple system the aliasing problem is not consequential because there are enough restrictions on the form of the system to ensure identifiability. The problem was originally considered in Phillips (1973). In particular, the coefficient matrix $A = A\left(\theta\right)$ is real and is further restricted by its dependence on the parameter vector θ. Also, the covariance matrix of the error process $\int_0^h e^{sA(\theta_2)} \Sigma\left(\theta_2\right) e^{sA(\theta_2)'} ds$ in the discrete system is real and necessarily positive semi-definite. These restrictions suffice to ensure the identifiability of A^0 in (57), removing the aliasing problem. Discussion and resolution of these issues is given in Phillips (1973) and Hansen and Sargent (1984). Of course, further restrictions may be needed to ensure that θ_1 and θ_2 are identified in $A\left(\theta_1^0\right)$ and $\Sigma\left(\theta_2^0\right)$.

A second complication that arises in the statistical treatment of systems of stochastic differential equations is that higher order systems involve exact discrete systems of the vector autoregressive and moving average type, which have more complicated likelihood functions. A third complication is that the discrete data often involves both stock and flow variables, so that some variables are instantaneously observed (like interest rates) while other variables (like consumption expenditure) are observed as flows (or integrals) over the sampling interval. Derivation of the exact discrete model and the likelihood function in such cases presents further

difficulties - see Phillips (1978) and Bergstrom (1984) - and involves complicated submatrix formulations of matrix exponential series. Most of these computational difficulties have now been resolved and Gaussian ML methods have been regularly used in applied research with these continuous time macroeconometric systems. Bergstrom (1996) provides a survey of the subject area and much of the empirical work. A more recent discussion is contained in Bergstrom and Nowman (2006).

8 Conclusions

Research on ML estimation of continuous time systems has been ongoing in the econometric and statistical literatures for more than three decades. But the subject has received its greatest attention in the last decade, as researchers in empirical finance have sought to use these models in practical applications of importance in the financial industry. Among the more significant of these applications have been the analysis of the term structure of interest rates and the pricing of options and other financial derivatives which depend on parameters that occur in the dynamic equations of motion of variables that are most relevant for financial asset prices, such as interest rates. The equations of motion of such variables are typically formulated in terms of stochastic differential equations and so the econometric estimation of such equations has become of critical importance in these applications. We can expect the need for these methods and for improvements in the statistical machinery that is available to practitioners to grow further as the financial industry continues to expand and data sets become richer. The field is therefore of growing importance for both theorists and practitioners.

REFERENCES

Ahn, D. and B. Gao, 1999, A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, 12, 721–762.

Aït-Sahalia, Y., 1999, Transition Densities for Interest Rate and Other Nonlinear Diffusions, *Journal of Finance*, 54, 1361–1395.

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

Aït-Sahalia, Y. and J. Yu, 2006, Saddlepoint approximation for continuous-time Markov Processes. *Journal of Econometrics*, forthcoming.

Barndorff-Nielsen, O. and N. Shephard, 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64, 253-280.

Barndorff-Nielsen, O. and N. Shephard, 2005, How accurate is the asymptotic approximation to the distribution of realized volatility?. In *Identification and Inference for Econometric Models*, eds by D.W.K. Andrews, J. Powell, P. Ruud and J. Stock, Cambridge University Press.

Bergstrom, A. R. 1966, Nonrecursive models as discrete approximation to systems of stochastic differential equations, *Econometrica*, 34, 173-82.

Bergstrom, A. R. 1984. Continuous time stochastic models and issues of aggregation over time. In Z. Griliches and M.D. Intriligator, editors, *Handbook of Econometrics*. Vol. II (Elsevier Science, Amsterdam).

Bergstrom, A. R. 1996. Survey of continuous time econometrics, in *Dynamic Disequilibrium Modeling* edited by W. A. Barnett, G. Gandolfo and ?C. Hillinger, Cambridge: Cambridge University Press, pp. 3-25.

Bergstrom, A. R. and B. Nowman 2006. *A Continuous Time Econometric Model of the United Kingdom with Stochastic Trends.* Cambridge: Cambridge University Press.

Billingsley, P., 1961, *Statistical Inference for Markov Processes*. Chicago, US: University of Chicago Press.

Black, F. and Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economics*, 81, 637-54.

Chan, N. H. and C.Z. Wei (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics*, 16, 367-401.

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

Dai, Q. and K. Singleton, 2003, Term Structure Dynamics in Theory and Reality, *Review of Financial Studies*, 16, 631-678.

Daniels, H. E. 1954. Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, 25, 631-650.

Durham, G., and A. R. Gallant, 2002, Numerical Techniques for Maximum Likelihood Estimation of Continuous-time Diffusion Processes, *Journal of Business and Economic Statistics*, 20, 297–316.

Duffie, D., and K. J. Singleton, 1993, Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, 61, 929–952.

Elerian, O., 1998, A Note on the Existence of a Closed-form Conditional Transition Density for the Milstein Scheme, Economics discussion paper 1998-W18, Nuffield College, Oxford.

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

Feller, W. 1951, Two Singular Diffusion Problems, *Annals of Mathematics*, 54, 173-182.

Field, C. and E. Ronchetti. 1990, *Small Sample Asymptotics*. IMS Lecture Notes, 13. Hayward, California.

Gallant, A. R., and G. Tauchen, 1996, Which moments to match? *Econometric Theory*, 12, 657–681.

Ghysels, E., Harvey, A.C., and E. Renault 1996. Stochastic volatility. In *Statistical Models in Finance*, eds. Rao, C.R. and Maddala, G. S., North-Holland, Amsterdam, pp. 119–191.

Gouriéroux, C., A. Monfort, and E. Renault, 1993, Indirect Inference, *Journal of Applied Econometrics*, 8, S85–S118.

Gouriéroux, C., Phillips, P.C.B. and J. Yu, 2005, Indirect inference for dynamic panel models. Working paper, Cowles Foundation for Research in Economics, Yale University.

Gouriéroux, C., E. Renault, and N. Touzi, 2000, Calibration by simulation for small sample bias correction, in R.S. Mariano, T. Schuermann and M. Weeks (eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*, Cambridge University Press, p328-358.

Hall, P. and Heyde, 1980, Martingale Limit Theory and Its Application, Academic Press.

Hansen, L. P. and T. J. Sargent 1983. The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica*, 51, 377-388.

Holly, A. and P.C.B. Phillips, 1979, An saddlepoint approximation to the distribution to the k-class estimator in a simultaneous system, *Econometrica*, 47, pp. 1527-1548.

Houthakker, H. S. and L. D. Taylor 1966. *Consumer demand in the United States 1929-1970, Analysis and Projections*. Cambridge: Harvard University Press.

Jacod, J., 1994, Limit of random measures associated with the increments of a Brownian semiartingal, Working paper, Laboratoire de Probabilities, Universite Pierre et Marie Curie, Paris.

Karatzas, I. and Shreve, S.E., 1988, Brownian Motion and Stochastic Calculus. Springer-Verlag, New York.

Kessler, M., 1997, Estimation of an ergordic diffusion from discrete observations, *Scandinavian Journal of Statistics*, 24, 211–229.

Kloeden, P.E. and E. Platen, 1999, *Numerical Solution of Stochastic Differential Equations*, Springer, New York.

Lánska, V., 1979, Minimum contrast estimation in diffusion processes. *Journal of Applied Probability*, 16, 65–75.

Liptser, R.S. and A.N. Shiryaev, 2000, *Statistics of Random Processes*, Springer-Verlag, New York.

Lo, A. W., 1988, Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data, *Econometric Theory*, 4, 231–247.

McCullagh, P., 1987 *Tensor Methods in Statistics*. London: Chapman and Hall.

Merton, R. C., 1980, On Estimating the Expected Return on the Market: An Exploratory Investigation, *Journal of Financial Economics*, 8, 323–361.

Merton, R. C. 1990. Continuous-time Finance. Blackwell, Massachusetts.

Milstein, G.N. 1978, A Method of Second-Order Accuracy Integration of Stochastic Differential Equations, *Theory of Probability and its Applications*, 23, 396-401.

Monfort, A. 1996, A reappraisal of misspecified econometric models. *Econometric Theory*, 12, 597–619.

Nowman, K. B., 1997, Gaussian Estimation of Single-factor Continuous Time Models of the Term Structure of Interest Rates, *Journal of Finance*, 52, 1695–1703.

Pedersen, A., 1995, A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observation. *Scandinavian Journal of Statistics*, 22, 55-71.

Phillips, P.C.B., 1972, The Structural Estimation of a Stochastic Differential Equation System, *Econometrica*, 40, pp. 1021-1041.

Phillips, P.C.B., 1973, The problem of identification in finite parameter continuous time models, *Journal of Econometrics*, 1, 351-362

Phillips, P.C.B., 1974, The Estimation of Some Continuous Time Models, *Econometrica*, 42, 803-823.

Phillips, P.C.B., 1978, Edgeworth and saddlepoint approximations in a first order non-circular autoregression, *Biometrika*, 65, pp. 91-98.

Phillips, P.C.B., 1980, The exact finite sample density of instrumental variable estimators in an equation with n+1 endogenous variables. *Econometrica*, 48, 861-878.

Phillips, P.C.B., 1984, Marginal densities of instrumental variable estimators in the general single equation case , *Advances in Econometrics*, 2, 1-24.

Phillips, P.C.B., 1987, Time series regression with a unit root, *Econometrica*, 55, 277-301.

Phillips, P.C.B., 1991, Error correction and long run equilibrium in continuous time, *Econometrica,* 59, 967-980.

Phillips, P.C.B., and T. Magdalinos (2006). Limit theory for moderate deviations from unity. *Journal of Econometrics* (forthcoming).

Phillips, P.C.B. and J. Yu, 2005a, Jackknifing bond option prices. *Review of Financial Studies*, 18, 707-742.

Phillips, P.C.B. and J. Yu, 2005b, Comments: A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, 20, 338–343.

Phillips, P.C.B. and J. Yu, 2005c, A Two-Stage Realized Volatility Approach to Estimation of Diffusion Processes with Discrete Data. Working paper, Cowles Foundation for Research in Economics, Yale University.

Quenouille, M. H., 1956, Notes on Bias in Estimation, *Biometrika*, 43, 353-360.

Reid, N. 1988. Saddlepoint methods and statistical inference. *Statistical Science*, 3, 213-238.

Sargan, J. D. 1974. Some discrete approximations to continuous time stochastic models. *Journal of the Royal Statistical Society, Series B*, 36, 74-90.

Shoji, I. and T. Ozaki 1997. Comparative study of estimation methods for continuous time stochastic processes. *Journal of Time Series Analysis* 18(5), 485–506.

Shoji, I. and T. Ozaki, 1998. Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications*, 16(4), 733–752.

Smith, A.A., 1993, Estimating nonlinear time-series models using simulated vector autoregressions, *Journal of Applied Econometrics*, 8, S63–S84.

Shephard, N.,2005, *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford.

Tierney, L. and J. Kadane, 1986, Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81, 82-87.

Tse, Y., X. Zhang, and J. Yu, 2004. Estimation of Hyperbolic Diffusion using MCMC Method. *Quantitative Finance*, 4, 158–169.

Vasicek, O., 1977, An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5, 177–186.

Yu, J., and P. C. B. Phillips, 2001, A Gaussion Approach for Estimating Continuous Time Models of Short Term Interest Rates, *The Econometrics Journal*, 4, 211-225.