Maximum Likelihood Estimation of Generalized Ito Processes With Discretely-Sampled Data

Article *in* Econometric Theory · August 1986

DOI: 10.1017/S0266466600012044 · Source: RePEc

1 author:

Andrew W Lo
Massachusetts Institute of Technology

Econometric Theory, **4**, 1988, 231–247. Printed in the United States of America.

MAXIMUM LIKELIHOOD ESTIMATION OF GENERALIZED ITÔ PROCESSES WITH DISCRETELY SAMPLED DATA

ANDREW W. LO
University of Pennsylvania

This paper considers the parametric estimation problem for continuous-time stochastic processes described by first-order nonlinear stochastic differential equations of the generalized Itô type (containing both jump and diffusion components). We derive a particular functional partial differential equation which characterizes the exact likelihood function of a discretely sampled Itô process. In addition, we show by a simple counterexample that the common approach of estimating parameters of an Itô process by applying maximum likelihood to a discretization of the stochastic differential equation does not yield consistent estimators.

1. INTRODUCTION

For purposes of modelling the evolution of various random systems, the theory of continuous-time stochastic processes has become indispensable. The use of continuous-time processes described by Itô stochastic differential equations (SDE's) is now an integral part of such diverse fields as stochastic optimal control theory, financial economics, and statistical thermodynamics. This is due, in part, to the development of a fully operational "stochastic calculus" by Itô [36] which extends the standard tools of calculus to functions of a wide class of continuous-time random processes (now known as Itô processes).[1] Another important aspect of the class of Itô processes is its closure under quite general nonlinear transformations; that is, nonlinear functions of Itô processes are (under mild regularity conditions) also Itô processes. Moreover, given the SDE of the original process, the Itô calculus provides a method for explicitly calculating the SDE driving the transformed process dynamics. Due to this remarkable result, the stochastic properties of quite complex models driven by Itô processes may be readily deduced as, for

This paper has benefitted considerably from the comments of Peter Phillips and three referees. I also thank Gary Chamberlain, Lars Hansen, Jerry Hausman, A. Craig MacKinlay, Charles Manski, Rennie Mirollo, Whitney K. Newey, Krishna Ramaswamy, seminar participants at the University of Chicago, the University of Minnesota's IMA Conference on Stochastic Differential Equations, and the University of Wisconsin at Madison for comments on an earlier draft. I am grateful to Stephanie Hogue, Elizabeth Schmidt, and Madhavi Vinjamuri for preparing the manuscript and the National Science Foundation (Grant No. SES-8520054) for providing research support. Any errors are of course my own.

 0266-4666/88 $5.00 + .00

example, in the case of the well-known Black–Scholes [10] stock option-pricing model. Although Itô processes are used most often in economics as models of asset-price behavior, their applications are considerably more widespread. However, to date relatively little research in economics has been devoted to the econometric estimation problems associated with such continuous-time processes.[2] This is particularly surprising since the modelling of uncertainty via Itô processes often yields very precise statistical specifications for estimation.

In this paper, we consider the parametric estimation problem for Itô processes using the method of maximum likelihood (ML) with discretely sampled data.[3] Section 2.1 contains the main result of the paper, a characterization of the exact likelihood function of the discrete sample as the solution to a particular functional partial differential equation (p.d.e.). Although the usual existence theorems do not apply to this functional p.d.e., when the solution does exist it may often be obtained by solving this equation via standard methods to yield the likelihood function. Moreover, the functional p.d.e. may also be used as a conclusive check to an "educated guess" for the likelihood function. Two illustrative examples are provided. In Section 2.2, we show by a simple counterexample that the common approach of estimating parameters of an Itô process by employing maximum likelihood upon a discretization of the stochastic differential equation need not yield consistent estimators. We conclude in Section 3.

Because the theory of statistical inference for an alternative class of continuous-time processes is now well established and comprehensively developed by Phillips [54–56] and Bergstrom[4] [6–8], a few remarks concerning the relation of that literature to inference for generalized Itô processes are in order before we begin. One important distinction between Itô processes and those studied by Phillips and Bergstrom is that the latter (hereafter called "nth-order processes") are described by nth-order linear SDE's with constant coefficients, whereas the former satisfy first-order *nonlinear* SDE's.[5] Of course, an nth-order linear SDE with constant coefficients may in principle be considered a special case of a vector first-order nonlinear SDE in the usual fashion.[6] However, we assume throughout that all the state variables are observable by the econometrician. With unobservable state variables the parameters may no longer be identified, which may complicate the analysis severely and is beyond the scope of this paper.

Another important difference between nth-order and Itô SDE's is the type of randomness driving the processes. For example, Bergstrom [7,8] considers SDE's driven by white-noise disturbances but does not restrict them to be Gaussian. In contrast, the approach taken in this paper is to consider SDE's driven by the sum of Gaussian and Poisson white-noise components. Although almost all sample paths of Gaussian white noise (Brownian motion) are continuous, the introduction of a Poisson component allows for simple sample-path discontinuities. Also, the distributional assumptions on the disturbances facilitate the explicit calculation of statistical properties of Itô processes and the derivation of the likelihood function. Furthermore, given the closure of the class of Itô processes under smooth nonlinear transformations and the Itô calculus, the stochastic behavior of functions of such processes are well-specified. This does not obtain for functions of nth-order processes. Although nth-order processes may seem more general than Itô processes driven by Gaussian and Poisson white noise, the Itô calculus has been shown to extend to quite general martingale processes.[7] For purposes of exposition, the loss in concreteness is not justified by such generality.

One further aspect of Itô processes which is distinct from nth-order processes is the Markovian nature of the former. This leads to a considerable simplification in the calculation of the likelihood function of discretely sampled data which are not necessarily equally spaced in time. The Markov property is clearly restrictive and may not be applicable to certain economic processes. This may be partly remedied by the usual "expansion of the states" technique, although issues of tractability arise when the number of states is large. The appropriateness of the Markov property depends ultimately upon the underlying economic model at hand and must be considered on a case-by-case basis.[8]

2. THE ESTIMATION PROBLEM

For expositional clarity we consider the estimation problem only for univariate Itô processes with single jump and diffusion components. The extension to vector Itô processes with multiple jump and diffusion terms poses no conceptual difficulties but is notationally more cumbersome.

Let $\{X(t)\colon t \in T \subset R^+,\ X(t) \in S \subset R\}$ be a stochastic process defined on a complete probability space (Ω, F, μ) and suppose $X(t)$ satisfies the following stochastic integral equation:

$$X(t) = X(t_0) + \int_{t_0}^{t} a(X,\tau;\alpha)\,d\tau + \int_{t_0}^{t} b(X,\tau;\beta)\,dW(\tau) + \int_{t_0}^{t} c(X,\tau;\gamma)\,dN_\lambda(\tau), \qquad (1)$$

where the last two (stochastic) integrals in (1) are defined with respect to the pure Wiener process $W(t)$ and a Poisson counter $N_\lambda(t)$, respectively, and a, b, and c are known functions which depend upon (X,t) and an unknown parameter vector $\bar{\theta} = [\alpha'\ \beta'\ \gamma']'$. Note that because the integrand b in (1) may be a function of X as well as of t, and because $W(t)$ is of unbounded variation, the corresponding stochastic integral cannot be interpreted in the wide or second-order sense as, for example, in Bergstrom [7].[9] The integral may, however, be interpreted in the sense of Itô [36] if the functions a, b, and c satisfy the following restrictions:

A1. Let $\{F_t\colon t \in T\}$ denote a right-continuous filtration σ-field defined on (Ω, F, μ) and let the pure Wiener process $\{W(t)\colon t \in T\}$ be adapted to this filtration.[10]

A2. If B is the σ-field of Borel sets on R^+, then for all $\bar{\theta}$ in the parameter space $\bar{\Theta} \equiv \mathrm{A} \times \mathrm{B} \times \Gamma$ the functions a, b, and c are measurable in the product σ-field $B \times F$.

A3. For all $\bar{\theta} \in \bar{\Theta}$ and $t \in T$ the functions a, b, c depend only upon $X(t)$ and t, implying that the functions are trivially adapted to the filtration $\{F_t\colon t \in T\}$ and are therefore *nonanticipating* with respect to that filtration.

A4. For all $\bar{\theta} \in \bar{\Theta}$ the functions a, b, and c satisfy the following inequalities almost surely:

$$\int_{t_0}^{t} |a|\,d\tau < \infty, \qquad \int_{t_0}^{t} |b|^2 d\tau < \infty, \qquad \int_{t_0}^{t} |c|^2 d\tau < \infty. \tag{2}$$

An equivalent and perhaps more familiar representation of $X(t)$ as a stochastic differential equation is given by

$$dX(t) = a(X,t;\alpha)dt + b(X,t;\beta)dW(t) + c(X,t;\gamma)dN_\lambda(t). \tag{3}$$

In order to ensure the existence and uniqueness of a solution to the stochastic integral [differential] equation given by (1) [(3)], we require:

A5. There exists some constant $K > 0$ such that the functions a, b, and c satisfy the following conditions for all $X, X' \in S$ and $t, t' \in T$:

$$|a(X,t) - a(X',t)| + |b(X,t) - b(X',t)| + |c(X,t) - c(X',t)| \le K|X - X'|, \tag{4a}$$

$$|a(X,t) - a(X,t')| + |b(x,t) - b(X,t')| + |c(X,t) - c(X,t')| \le K|t - t'|, \tag{4b}$$

$$a^2(X,t) + b^2(X,t) + c^2(X,t) \le K^2(1 + X^2). \tag{4c}$$

Finally, for purposes of estimation we make two additional assumptions:

A6. The functions a, b, and c are twice continuously differentiable in (X,t) and three times continuously differentiable in $\bar{\theta}$; the function $\tilde{c}(X,t;\gamma) \equiv X + c(X,t;\gamma)$ is bijective and $\left|\dfrac{\partial c}{\partial X} + 1\right| \neq 0$ for all $(t,\gamma) \in T \times \Gamma$ and $X \in S$.

A7. The true but unknown parameters $\theta_0 \equiv [\bar{\theta}_0' \ \lambda_0]'$ lie in the interior of a finite-dimensional closed and compact parameter space $\Theta \equiv \bar{\Theta} \times \Lambda$.

2.1 A Characterization of the Likelihood Function

Suppose the process $X(t)$ is sampled at $n + 1$ discrete points in time $t_0, t_1, \ldots, t_n$, *not* necessarily equally spaced apart and let $X \equiv (X_0, X_1, \ldots, X_n)$

denote this random sample where $X_k \equiv X(t_k)$. Given the discretely sampled data X and the stochastic specification of the process $X(t)$, denote by $P(X_0, X_1, \ldots, X_n; \theta)$ the finite-dimensional distribution of X and let $\rho(X; \theta)$ denote the density representation of P.[11] When considered a function of θ, this joint density is obviously the desired exact likelihood function of X. Since $X(t)$ is a Markov process (see Arnold [1, Chapter 9]), the joint density ρ may be rewritten as the following product of conditional densities:

$$\rho(X) = \rho_0(X_0) \prod_{k=1}^{n} \rho_k(X_k, t_k | X_{k-1}, t_{k-1}). \tag{5}$$

Deriving the likelihood function then reduces to calculating the transition density functions ρ_k. The main result of our paper is a characterization of these transition densities via the corresponding forward or Fokker–Planck equation which we derive in the following theorem:

THEOREM. *Under assumptions (A1)–(A7), the likelihood function ρ_k solves the following functional partial differential equation:*

$$\frac{\partial}{\partial t}[\rho_k] = -\frac{\partial}{\partial X}[a\rho_k] + \frac{1}{2}\frac{\partial^2}{\partial X^2}(b^2\rho_k) - \lambda\rho_k + \lambda\tilde{\rho}_k \left| \frac{\partial}{\partial X}[\tilde{c}^{-1}] \right| \tag{6}$$

subject to

$$\rho_k(X, t_{k-1}) = \delta(X - X_{k-1}), \tag{7}$$

and any other relevant boundary conditions, where $\tilde{c}$ is defined in (A6), $\tilde{\rho}_k \equiv \rho_k(\tilde{c}^{-1}, t)$, and $\delta(X - X_{k-1})$ is the Dirac-delta generalized function centered at X_{k-1}.

Proof. Let $\psi(X)$ be an arbitrary infinitely differentiable function with compact support, i.e., $\psi \in C_c^\infty(R)$. By Itô's Lemma (see Brockett [12]) we have

$$d\psi = \left[\psi_X a + \frac{1}{2}\psi_{XX} b^2\right] dt + \psi_X b dW + [\psi(X + c) - \psi(X)] dN_\lambda, \tag{8a}$$

where

$$\psi_X \equiv \frac{d\psi}{dX}, \qquad \psi_{XX} \equiv \frac{d^2\psi}{dX^2}. \tag{8b}$$

Define $D_{P,k}$ to be the Dynkin operator at time t_k, i.e., $D_{P,k} \equiv \frac{d}{dt} E_{t_k}[\cdot]$. Applying it to ψ yields

$$D_{P,k}[\psi] = E_{t_k}\left[\psi_X a + \frac{1}{2}\psi_{XX} b^2\right] + \lambda E_{t_k}[\psi(X + c) - \psi(X)]. \tag{9}$$

We may express $D_{P,k}[\psi]$ as the following integral:

$$D_{P,k}[\psi] = \int_S \left\{ \psi_X a + \frac{1}{2} \psi_{XX} b^2 + \lambda[\psi(X+c) - \psi(X)] \right\} \rho_k(X,t)dX, \tag{10a}$$

$$= \int_S \left[-\psi \frac{\partial}{\partial X}(a\rho_k) + \frac{1}{2}\psi \frac{\partial^2}{\partial X^2}(b^2\rho_k) - \psi\lambda\rho_k \right] dX + \lambda \int_S \psi(X+c)\rho_k dX, \tag{10b}$$

where the second equality is obtained by integrating by parts and collecting terms. By (A6), the Inverse Function Theorem guarantees the existence of $\tilde{c}^{-1}$ such that $X = \tilde{c}^{-1}(\tilde{c}(X,t;\gamma),t;\gamma)$. Using the change of variables formula, we have

$$\int_S \psi(X+c)\rho_k(X,t)dX = \int_S \psi(Y)\rho_k(\tilde{c}^{-1}(Y,t;\gamma)) \cdot \left| \frac{\partial}{\partial Y}(\tilde{c}^{-1}(Y,t;\gamma)) \right| dY, \tag{11a}$$

$$= \int_S \psi(X)\rho_k(\tilde{c}^{-1}(X,t;\gamma)) \left| \frac{\partial}{\partial X}(\tilde{c}^{-1}(X,t;\gamma)) \right| dX. \tag{11b}$$

We then conclude that

$$D_{P,k}[\psi] = \int_S \left\{ -\frac{\partial}{\partial X}(a\rho_k) + \frac{1}{2}\frac{\partial^2}{\partial X^2}(b^2\rho_k) - \lambda\rho_k - \lambda\tilde{\rho}_k \left| \frac{\partial}{\partial X}\tilde{c}^{-1} \right| \right\} \psi(X)dX. \tag{12}$$

But $D_{P,k}[\psi]$ may be calculated alternatively as

$$D_{P,k}[\psi] = \frac{d}{dt} E_{t_k}[\psi] = \int_S \psi(X) \frac{\partial}{\partial t}[\rho_k(X,t)]dX. \tag{13}$$

Equating (12) and (13) and noting that the equality obtains for arbitrary $\psi \in C_c^\infty(R)$ allows us to conclude that

$$\frac{\partial}{\partial t}[\rho_k] = -\frac{\partial}{\partial X}[a\rho_k] + \frac{1}{2}\frac{\partial^2}{\partial X^2}(b^2\rho_k) - \lambda\rho_k + \lambda\tilde{\rho}_k \left| \frac{\partial}{\partial X}[\tilde{c}^{-1}] \right|, \tag{14a}$$

with the initial condition

$$\rho_k(X, t_{k-1} | X_{k-1}, t_{k-1}) = \delta(X - X_{k-1}), \tag{14b}$$

where $\delta(X - X_k)$ is the Dirac-delta generalized function centered at X_{k-1}. ∎

Because the differential equation in (14) is a *functional* partial differential equation, the usual existence and uniqueness theorems for p.d.e.'s do not apply and a solution is unfortunately not guaranteed for general coefficient functions a, b, and c. However, when the existence of a density representation for a specific process has been assured by other means, (14) may often be solved by standard methods (Fourier transforms, etc.) to yield the likelihood function. Also, additional restrictions upon the coefficient functions may simplify these calculations. As an example, if $c \equiv 0$ (pure diffusion) and a and b satisfy the following reducibility condition

$$\frac{\partial}{\partial X}\left[\frac{1}{b^2}\frac{\partial b}{\partial t} - \frac{\partial}{\partial X}\left[\frac{a}{b}\right] + \frac{1}{2}\frac{\partial^2 b}{\partial X^2}\right] = 0, \tag{15}$$

it may be shown (see Schuss [65, Chapter 4]) that there exists a transformed process $Z(t)$ of $X(t)$ for which the coefficient functions are independent of $Z(t)$. That is, for some suitable change of variables $F[X(t)] \equiv Z(t)$, an application of Itô's lemma will yield:[12]

$$dZ = p(t;\theta)dt + q(t;\theta)dW. \tag{16}$$

In this case the transition density function for the transformed data is readily derived as

$$\rho_k(Z,t) = \left[2\pi\int_{t_{k-1}}^{t} q^2 d\tau\right]^{-1/2} \exp\left[-\frac{\left(Z - Z_{k-1} - \int_{t_{k-1}}^{t} p d\tau\right)^2}{2\int_{t_{k-1}}^{t} q^2 d\tau}\right] \tag{17}$$

For example, it is easily established that the lognormal diffusion process $dX = \alpha X dt + \beta X dW$ satisfies the reducibility condition and the transformation $Y = F(X)$ is readily derived as ln X. Applying this to X and using Itô's differentiation rule then yields $dY = \left[\alpha - \frac{\beta^2}{2}\right]dt + \beta dW$ which has a simple Gaussian likelihood function.

Because the usual methods for solving partial differential equations are in some cases quite cumbersome, solutions are often obtained by "educated guesses." In these cases, (14) provides a conclusive check for such conjectured density representations as the following example illustrates:

Example 1. (Lognormal diffusion and jump process.) We seek the likelihood function corresponding to the process $X(t)$ which satisfies the following SDE

$$dX = \alpha X dt + \beta X dW + \gamma X dN_\lambda, \qquad \gamma \geq 0. \tag{18}$$

Using the log-transformation $Y = \ln X$ and Itô's lemma yields

$$dY = \left(\alpha - \frac{\beta^2}{2}\right)dt + \beta dW + \ln(1+\gamma)dN_\lambda. \tag{19}$$

Since dW and dN_λ are assumed to be independent and the coefficient functions in (19) do not depend upon Y, a reasonable guess for the conditional likelihood function of Y_t given $Y_{t-\tau}$ is the convolution $\rho_v * \rho_z$ of a Poisson density ρ_v with intensity λ and a Gaussian density ρ_z with mean $\mu\tau \equiv \left(\alpha - \frac{\beta^2}{2}\right)\tau$ and variance $\beta^2\tau$, and is given by:[13]

$$\rho_Y(Y,t) = \sum_{k=0}^{\infty} \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!} [2\pi\beta^2\tau]^{-1/2} \times \exp\left[-\frac{[Y_t - Y_{t-\tau} - k\ln(1+\gamma) - \mu\tau]^2}{2\beta^2\tau}\right]. \tag{20}$$

This guess is readily vindicated by performing the required differentiation and checking that (14) is satisfied.

In addition to the initial condition (14b), the solution of equation (14a) often depends critically on particular auxiliary restrictions placed on the process $X(t)$ as a result of economic considerations. For example, when $X(t)$ represents an asset's price a non-negativity condition is required. Such restrictions usually take the form of boundary conditions for (14) as in the following example:

Example 2. (Diffusion with absorbing barrier.)[14] Let $X(t)$ satisfy the following SDE

$$dX(t) = \alpha dt + \beta dW(t), \qquad X(0) = X_0 > 0, \tag{21}$$

with the added restriction that $X = 0$ is an absorbing state, i.e., once the process reaches 0 it remains at that state thereafter. In addition, suppose that we have the observations $X_1 > 0, \ldots, X_{n-1} > 0$, $X_n = 0$ so that absorption is realized in this sample some time *between* t_{n-1} and t_n. Consider the transition density for $X(t_k)$ conditional upon $X(t_{k-1})$ where $k < n$. It may be shown that in this case the forward equation (14) reduces to

$$\frac{\partial \rho_k}{\partial t} = \frac{1}{2}\beta^2 \frac{\partial^2 \rho_k}{\partial X^2} - \alpha \frac{\partial \rho_k}{\partial X}, \tag{22a}$$

$$\rho_k(X, t_k | X_{k-1}, t_{k-1}) = \delta(X - X_{k-1}), \tag{22b}$$

with the added boundary condition that

$$\rho_k(0, t_k | X_{k-1}, t_{k-1}) = 0. \tag{22c}$$

Using the "method of images" (see [19] or [38]) this may be solved to yield

$$\rho_k(X_k,t_k|X_{k-1},t_{k-1}) = [2\pi\beta^2\Delta t_k]^{-1/2}\left\{\exp\left[-\frac{(X_k - X_{k-1} - \alpha\Delta t_k)^2}{2\beta^2\Delta t_k}\right] - \exp\left[-\frac{2\alpha X_{k-1}}{\beta^2} - \frac{(X_k + X_{k-1} - \alpha\Delta t_k)^2}{2\beta^2\Delta t_k}\right]\right\}, \tag{23}$$

where $\Delta t_k \equiv t_k - t_{k-1}$. Now the transition density of $X(t_n)$ conditional upon $X(t_{n-1})$ will not be defined in the usual sense since X has been absorbed by time t_n. However, the *probability* that absorption has occurred by time t_n conditional upon $X(t_{n-1})$ may be derived as

$$P[\text{Absorption in } [t_{n-1},t_n)] = \Phi\left[\frac{-X_{n-1} - \alpha\Delta t_n}{\beta\sqrt{\Delta t_n}}\right] + \exp\left[-\frac{2\alpha X_{n-1}}{\beta^2}\right]\Phi\left[\frac{-X_{n-1} + \alpha\Delta t_n}{\beta\sqrt{\Delta t_n}}\right], \tag{24}$$

thus the transition density may be defined as the following generalized function

$$\rho_n(X,t_n|X_{n-1},t_{n-1}) = P[\text{Absorption in } [t_{n-1},t_n)]\,\delta(X). \tag{25}$$

More generally the transition density for any observation $k, k = 1,\ldots,n$ is given by

$$\rho_k = [2\pi\beta^2\Delta t_k]^{-1/2}\left[\exp\left[-\frac{(X_k - X_{k-1} - \alpha\Delta t_k)^2}{2\beta^2\Delta t_k}\right] - \exp\left[-\frac{2\alpha X_{k-1}}{\beta^2} - \frac{(X_k + X_{k-1} - \alpha\Delta t_k)^2}{2\beta^2\Delta t_k}\right]\right] + P[\text{Absorption in } [t_{k-1},t_k)]\,\delta(X), \qquad X \geq 0. \tag{26}$$

This conditional likelihood function is quite similar to the likelihood of the well-known censored linear regression model which is composed of a discrete and continuous part. Although the conditional likelihood function in Example 2 is indeed a solution for (14), it contains a Dirac δ-function and is therefore not a function in the usual sense.[15] However, this poses no problems for maximum likelihood estimation but merely requires some care in choosing an appropriate carrier measure. Specifically, although the joint distribution function of the sample X in Example 2 is not absolutely continuous with respect to Lebesgue measure, it is absolutely continuous with respect to the *sum* of Lebesgue and counting measures. The proper likelihood function may then be derived by taking the Radon–Nikodym derivative of the joint probability measure with respect to the alternative carrier measure. In Example 2, this results in a joint likelihood function which is simply the product of the densities and probabilities, as in the censored regression model.[16]

2.2 Maximum Likelihood Estimation Via Discretized Itô Processes

Having characterized the likelihood function as the solution to (14), we assume its existence and define the maximum likelihood estimator in the usual manner.

$$\hat{\theta}_{\mathrm{ML}} \equiv \arg \operatorname*{Max}_{\theta} G(\theta; X), \tag{27a}$$

where

$$G(\theta; X) \equiv \ln \rho_0(X_0, t_0) + \sum_{k=1}^{n} \ln \rho_k(X_k, t_k | X_{k-1}, t_{k-1}; \theta). \tag{27b}$$

Since $\hat{\theta}_{\mathrm{ML}}$ is a true maximum likelihood estimator, it possesses the standard properties of consistency and asymptotic normality under appropriate regularity conditions.[17]

In the event that (14) cannot be solved explicitly to obtain the likelihood function of the sample X, several authors have estimated θ by applying maximum likelihood to a suitably discretized SDE.[18] Specifically, equally spaced discretely sampled data are assumed to be generated by the following difference equation:

$$X_{k+1} = X_k + a(X_k, t_k; \alpha)h + b(X_k, t_k; \beta) \cdot \Delta W(t_{k+1}) + c(X_k, t_k; \gamma) \cdot \Delta N_\lambda(t_{k+1}), \tag{28}$$

where

$$\Delta W(t_{k+1}) \equiv W(t_{k+1}) - W(t_k), \qquad \Delta N_\lambda(t_{k+1}) \equiv N_\lambda(t_{k+1}) - N_\lambda(t_k), \qquad t_k \equiv kh$$

for $k = 0, 1, \ldots, n$ and $h \equiv T/n$. The parameters $\theta \equiv [\alpha' \; \beta' \; \gamma' \; \lambda']$ are then estimated via maximum likelihood using (28). Because it is well-known that the sample paths of the discretization (28) converge to those of the continuous-time Itô process $X(t)$ as h approaches 0 (see, for example, Gihman and Skorohod [24]), such an estimation procedure may seem sensible.[19] However, we show by a simple counterexample that the resulting "discretized maximum likelihood" estimator $\hat{\theta}_D$ need not be consistent.

Example 3. (Inconsistency of $\hat{\theta}_D$.) Let $X(t)$ denote the lognormal diffusion on the interval $[0, T]$:

$$dX(t) = \alpha X(t)dt + \beta X(t)dW(t), \tag{29}$$

and consider its discretization according to (28):

$$X_{k+1} = \alpha X_k h + \beta X_k \Delta W_{k+1} \equiv \alpha X_k h + X_k \epsilon_{k+1}, \tag{30}$$

where ϵ_{k+1} is an i.i.d. $N(0,\beta^2 h)$ random variate. From (30), it is apparent that the discretized maximum likelihood estimators of α and β^2 are given by:

$$\hat{\alpha}_D = \frac{1}{T}\sum_{k=1}^{n}\left[\frac{X_k}{X_{k-1}} - 1\right], \qquad \hat{\beta}_D^2 = \frac{1}{T}\sum_{k=1}^{n}\left[\frac{X_k}{X_{k-1}} - 1 - \hat{\alpha}_D h\right]^2. \tag{31}$$

But for fixed observation intervals h, as the number of observations increases without bound the discretized estimators do not converge to the population parameters of interest:

$$\operatorname*{plim}_{n\to\infty} \hat{\alpha}_D = \frac{1}{h}\,[e^{\alpha h} - 1] \neq \alpha, \qquad \operatorname*{plim}_{n\to\infty} \hat{\beta}_D^2 = \frac{1}{h}\,e^{2\alpha h}[e^{\beta^2 h} - 1] \neq \beta^2. \tag{32}$$

From (32), it is evident that for small h the asymptotic bias may be negligible. Of course, whether or not the bias is economically meaningful is an empirical question which is process-specific and must be resolved for each application individually. However, it should be clear from (28) that for arbitrary coefficient functions a, b, and c the discretized ML estimator is generally inconsistent.

Since (32) indicates that the asymptotic bias is decreasing in the observation interval h, it might be conjectured that consistency may be restored if we draw observations more frequently within the fixed time span $[0,T]$, thereby letting h approach zero as n increases without bound so as to keep $T \equiv nh$ fixed. In fact, it may be shown that such a limiting operation, which Phillips [55] terms "continuous data recording",[20] does guarantee that $\hat{\beta}_D$ converges to β in probability. However, the same cannot be said for the drift estimator $\hat{\alpha}_D$. In particular, using functional central limit theory techniques developed in Phillips [55], we conclude that:

$$\hat{\alpha}_D \Rightarrow \alpha + \frac{\beta}{\sqrt{T}}\,[W(1) - W(0)], \tag{33}$$

where the symbol "$\Rightarrow$" denotes weak convergence. Thus, with more frequent sampling within a fixed time span, $\hat{\alpha}_D$ converges weakly to a Gaussian random variate with mean α and variance β^2/T.[21]

Of course, the inconsistency of the discretized maximum likelihood estimators does not imply that there exists no consistent estimators of the parameters. Indeed, there are at least two consistent estimators for the parameter α in (29).[22] Example 3 merely illustrates the inappropriateness of applying maximum likelihood estimation to the discretized process when it is the parameters of the *continuous-time* process that are of interest. That this is so should not be surprising since the discretization may be viewed as a misspecification of the true likelihood function. Although the discretization in Example 3 incorrectly assumes conditionally Gaussian observations, the exact conditional likelihood of each observation is the lognormal density.

In contrast to the discretized estimators, the exact maximum likelihood estimators $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$ converge in probability to the true parameters for fixed observation intervals h. However, under continuous data recording it may be demonstrated that $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$ behave identically to their discretized counterparts in the limit, i.e., $\hat{\alpha}_{ML}$ converges weakly to (33) and $\hat{\beta}_{ML}$ converges in probability to β.

3. CONCLUSION

In this paper, we have derived a functional partial differential equation which characterizes the likelihood function of a discretely sampled continuous-time generalized Itô process. Because the asymptotic properties of maximum likelihood estimators are well established, statistical inference for many continuous-time models of economic behavior may readily be performed. One such example is the estimation and testing of contingent claims asset-pricing models, as in Lo [42]. Important future applications include the empirical estimation of general equilibrium asset-pricing models such as those in Chamberlain [14], and Cox, Ingersoll, and Ross [20,21].

We have also shown by a simple example that the maximum likelihood estimator derived from a discretized stochastic differential equation need not be a consistent estimator of the parameters of the continuous-time process. Although this does not preclude the existence of consistent estimators, the example suggests that the naive approach of estimating parameters of a generalized Itô process using a discrete approximation is not appropriate. Consistency may sometimes be restored for a subset of parameters by sampling more frequently within a fixed time span. Of course, since the results in this paper are exclusively asymptotic in nature, the finite-sample properties of the estimators must be studied through Monte Carlo experiments on a case-by-case basis. This is especially important for applications in which the total time span of the data set is small, since it is precisely in such cases that inferences must depend critically upon the continuous data recording concept.[23] A potentially fruitful direction for further investigation may be to ascertain the empirical relevance of this new form of asymptotics.

NOTES

1. Generalized Itô processes satisfy quite general SDE's driven by both standard white noise (Brownian motion) and Poisson counters. The term "generalized" emphasizes the presence of discontinuities and serves to distinguish such processes from the more common Itô diffusions. The inclusion of the Poisson term is one of the principal advantages of the Itô process over

higher-order processes driven purely by sample-path continuous white noise since discrete changes in the state variables cannot be modelled by diffusion alone (whereas Brockett [12] has shown that *any* finite-state continuous-time jump process may be expressed as a generalized Itô process).

2. Most of the empirical applications seem to be within asset-pricing studies. See, for example, Rosenfeld [61]; Marsh and Rosenfeld [44]; Grossman, Melino, and Shiller [26]; Ball and Torous [4]; and Lo [42].

3. Note that in this context, the term "estimation" is used in the classical statistical sense. This is in contrast to its usage in the engineering and stochastic control literature, in which estimation is associated with the filtering, smoothing, and prediction problems. Of course, the parametric statistical estimation problem may be posed as a very special case of the filtering problem. However, because the focuses of the two approaches are quite different, the distinction between the two forms of estimation is significant.

The maximum likelihood approach to inference for continuous-time processes is, of course, not new. The recent volume by Basawa and Prakasa Rao [5, Chapter 9] contains an excellent survey and extensive bibliography of this vast and still growing literature. However, their discussion of ML estimation of Itô processes focuses exclusively on the special case of linear Itô diffusions. This is also true of other studies which consider maximum likelihood estimation of diffusion processes such as those by Liptser and Shiryayev [40, Chapter 17], Bagchi [2,3], Le Breton [39], Ljung [41], Brown and Hewitt [13], Tugnait [70–72], Borkar and Bagchi [11], and Loges [43]. In contrast, we consider maximum likelihood estimation for the general nonlinear Itô process which includes both diffusion and jump components. Of course, our goal here is more modest than those in some of the aforementioned papers: we only derive a characterization of the likelihood function, but do not prove its general existence.

4. Other examples include A.W. Phillips [53]; Sims [68]; Hansen and Sargent [28–30]; Christiano [15–17]; Harvey and Stock [32]; and Grossman, Melino, and Shiller [26].

5. One implication of this is that nth-order processes are "smoother" than Itô processes in the mean-square sense. More precisely, an nth-order process possesses mean-square derivatives up to order $n - 1$; Itô processes are not mean-square differentiable. This non-differentiability is an important property especially for purposes of modelling asset-prices since, as Harrison, Pitbladdo, and Schaefer [31] have shown, continuous-time equilibrium price processes generated by frictionless markets must be of unbounded variation.

6. For practical purposes, this quickly becomes intractable when systems of nth-order SDE's are considered, as in Bergstrom's approach.

7. See, for example, Skorohod [69, Chapters 2 and 3].

8. In particular, in this paper we do not deal with the important issues of time-aggregation and stock/flow distinctions which would render the Markov property inappropriate. These issues, however, are explicitly investigated in Sims [68]; Bergstrom [7,8]; Christiano [15,16]; and Grossman, Melino, and Shiller [26].

9. Since $N_\lambda(t)$ is of bounded variation, the stochastic integral with respect to the Poisson counter may be defined as a Lebesgue–Stieltjes integral.

10. That is, let $\{F_t\colon t \in T\}$ be a sequence of sub-σ-fields of the σ-field F such that:

(i) $F_t \subset F_s$ for $t \leq s$,

(ii) $F_t = \bigcap_{\tau > t} F_\tau$,

and let $W(t)$ be F_t-measurable for all $t \in T$.

11. More formally, let the measure corresponding to P be absolutely continuous with respect to some σ-finite carrier measure ν. Then ρ is simply the Radon–Nikodym derivative of the P measure with respect to ν. Note that ν need not be Lebesgue measure.

12. Furthermore, this transformation F may be explicitly derived by solving a simple ordinary differential equation given in Schuss [65, Chapter 4.1].

13. More formally, we define:

$$dV = \ln(1+\gamma)dN_\lambda, \qquad dZ = \left(\alpha - \frac{\beta^2}{2}\right)dt + \beta dW,$$

$$\rho_v(V_t, t) = c\sum_{k=0}^{\infty} \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!}\,\delta(cV_t - k), \qquad \text{where } c \equiv [\ln(1+\gamma)]^{-1},$$

$$\rho_z(Z_t, t) = (2\pi\beta^2 t)^{-1/2} \exp\left[-\frac{\left[Z_t - Z_{t-\tau} - \left(\alpha - \frac{\beta^2}{2}\right)\tau\right]^2}{2\beta^2\tau}\right].$$

14. A more satisfactory economic model might allow the absorbing barrier to change over time; however, the first-passage probability for this general case is analytically much more complicated. See, for example, Park and Paranjape [49]; Park and Schuurmann [50]; Park and Beekman [48]; and Siegmund [67].

15. More formally, the δ-function is an example of a generalized function or "distribution" (no relation to probability distributions), which is defined to be a real-valued continuous linear functional on $C_c^\infty(R)$. One important property of generalized functions is that their "derivatives" always exist. Moreover, all the standard formal rules of calculus obtain for these objects (differentiation, chain rule, integration by parts, etc.), a fact which is implicitly used in our derivation of the forward equation. See Gel'fand and Shilov [23, Chapter 1] or Rudin [63, Chapters 6 and 8] for a formal development of this theory.

16. Observe that $\frac{dP_\delta}{d\nu} = 1$ where P_δ is the probability measure associated with the δ-function and ν is counting measure. See Hoadley [35] for a more detailed discussion. That maximum likelihood estimation may still be performed when the solution of (14) is not a function in the classical sense is best illustrated by the example of the pure jump process $dX(t) = dN_\lambda(t)$ which yields the well-defined likelihood function $(\lambda t)^{X(t)}\exp[-\lambda t]/X(t)!$ for $X(t)$ conditional upon $X(0) = 0$, even though the solution to (14) in this case is a generalized function. Note that this likelihood function is simply the Radon–Nikodym derivation of the probability with respect to the counting carrier measure.

17. Because $\hat{\theta}_{ML}$ is based upon observations X_t which are neither independently nor identically distributed and possibly nonstationary, the standard proofs of consistency and asymptotic normality are not directly applicable. However, several authors have extended these proofs to much more general conditions using martingale convergence theorems. See, for example, Billingsley [9], Roussas [62], Gordin [25], Prakasa Rao [59], McLeish [45], Heyde [34], Crowder [22], and Herrndorf [33]. See Hall and Heyde [27] for details and more complete references. For further results and econometric applications, see Phillips [57,58] and Park and Phillips [51].

18. See, for example, Christie [18] and Ogden [46].

19. Although (28) is perhaps the most popular method of discretizing the Itô stochastic differential equation, there are several other methods. See, for example, Rao, Borwankar, and Ramakrishna [60], Rumelin [64], Janssen [37], and Pardoux and Talay [47].

20. "Continuous data recording" asymptotic inference is distinct from inference with a continuous data record (which involves no limit operations). For properties of maximum likelihood estimators of linear diffusions when sampling is *continuous*, see Liptser and Shiryayev [40, Chapter 17], Brown and Hewitt [13], Le Breton [39], and Basawa and Prakasa Rao [5, Chapter 9.5].

21. See Perron [52] for the testing analogue of consistency under continuous data recording.

22. Specifically, $\hat{\alpha}_1 \equiv \frac{1}{h}\ln[1 + \hat{\alpha}_D]$ and $\hat{\alpha}_2 \equiv \frac{1}{T}\ln(X_T/X_0) + \frac{1}{2}\hat{\beta}^2_{ML}$ are both consistent estimators of α.

23. In addition to the problem of inconsistency, the nominal size and power of standard statistical tests may differ substantially from their actual values, as Shiller and Perron's [66] example demonstrates.

REFERENCES

1. Arnold, L. *Stochastic Differential Equations: Theory and Applications.* New York: John Wiley and Sons, 1974.
2. Bagchi, A. Continuous-time systems identification with unknown noise covariance. *Automatica* 11 (1975): 533–536.
3. Bagchi, A. Consistent estimates of parameters in continuous-time systems. In O.L.R. Jacobs et al. *Analysis and Optimization of Stochastic Systems.* New York: Academic Press, 1980.
4. Ball, C.A. & W. Torous. On jumps in common stock prices and their impact on call option pricing. *Journal of Finance* 40 (1985): 155–173.
5. Basawa, I.V. & B.L.S. Prakasa Rao. *Statistical Inference for Stochastic Processes.* New York: Academic Press, 1980.
6. Bergstrom, A.R. *Statistical Inference in Continuous Time Economic Models.* Amsterdam: North-Holland, 1976.
7. Bergstrom, A.R. Gaussian estimation of structural parameters in higher-order continuous-time dynamic models. *Econometrica* 51 (1983): 117–152.
8. Bergstrom, A.R. Continuous time stochastic models and issues of aggregation over time. In Z. Griliches and M.D. Intriligator (ed.), *Handbook of Econometrics* Vol. II. Amsterdam: North-Holland Publishing Company, 1984.
9. Billingsley, P. *Statistical Inference for Markov Processes.* Chicago: University of Chicago Press, 1961.
10. Black, F. & M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy* 81 (1973): 637–654.
11. Borkar, B. & A. Bagchi. Parameter estimation in continuous-time stochastic processes. *Stochastics* 8 (1982): 193–212.
12. Brockett, R.W. Lecture Notes on Nonlinear Stochastic Control. Unpublished course notes, Harvard University, Spring 1984.
13. Brown, B.M. & J.I. Hewitt. Asymptotic likelihood theory for diffusion processes. *Journal of Applied Probability* 12 (1978): 228–238.
14. Chamberlain, G. Asset pricing in multiperiod securities markets. University of Wisconsin-Madison S.S.R.I. Working Paper No. 8510, 1985.
15. Christiano, L. The effects of aggregation over time on tests of the representative agent model of consumption. Mimeo, December 1984.
16. Christiano, L. A critique of conventional treatments of the model timing interval in applied econometrics. Mimeo, January 1985.
17. Christiano, L. Estimating continuous-time rational expectations models in frequency domain: A case study. Carnegie-Mellon University Working Paper No. 34-84-85, April 1985.
18. Christie, A. The stochastic behavior of common stock variances: Value, leverage, and interest rate effects. *Journal of Financial Economics* 10 (1982): 407–432.
19. Cox, D.R. & H.D. Miller. *The Theory of Stochastic Processes.* New York: Chapman and Hall, 1965.
20. Cox, J., J. Ingersoll, & S. Ross. An intertemporal general equilibrium model of asset prices. *Econometrica* 53 (1985): 363–384.
21. Cox, J., J. Ingersoll, & S. Ross. A theory of the term structure of interest rates. *Econometrica* 53 (1985): 385–408.
22. Crowder, M.J. Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society* Series B, 38 (1976): 45–53.

23. Gel'fand, I.M. & G.E. Shilov. *Generalized Functions Volume I: Properties and Operations.* New York: Academic Press, 1964.
24. Gihman, I.I. & A.V. Skorohod. *Controlled Stochastic Processes.* Springer, Berlin, 1979.
25. Gordin, M.I. The central limit theorem for stationary processes. *Soviet Mathematical Doklady* 10 (1969): 1174–1176.
26. Grossman, S., A. Melino, & R. Shiller. Estimating the continuous-time consumption-based asset pricing model. *Journal of Business and Economic Statistics* 5 (1987), 315–327.
27. Hall, P. & C.C. Heyde. *Martingale Limit Theory and Its Application.* New York: Academic Press, 1980.
28. Hansen, L. & T. Sargent. Formulating and estimating continuous-time rational expectations models. Unpublished manuscript, August 1981.
29. Hansen, L. & T. Sargent. The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* 51 (1983): 377–388.
30. Hansen, L. & T. Sargent. Identification of continuous-time rational expectations models from discrete-time data. Federal Reserve Bank of Minneapolis Staff Report 73, March 1983.
31. Harrison, J.M., R. Pitbladdo, & S.M. Schaefer. Continuous-price processes in frictionless markets have infinite variation. *Journal of Business* 57 (1984): 353–365.
32. Harvey, A. & J. Stock. The estimation of higher-order continuous-time autoregressive models. *Econometric Theory* 1 (1985): 97–112.
33. Herrndorf, N. A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability* 12 (1984): 141–153.
34. Heyde, C.C. On the central limit theorem and iterated logarithm law for stationary processes. *Bulletin of the Australian Mathematical Society* 12 (1975): 1–8.
35. Hoadley, B. Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Annals of Mathematical Statistics* 42 (1971): 1977–1991.
36. Itô, K. On stochastic differential equations. *Memoirs of the American Mathematical Society* 4 (1951): 1–51.
37. Janssen, R. Discretization of the Wiener process in difference methods for stochastic differential equations. *Stochastic Processes and their Applications* 18 (1984): 361–369.
38. Karlin, S. & H.M. Taylor. *A Second Course in Stochastic Processes.* New York: Academic Press, 1981.
39. Le Breton, A. On continuous and discrete sampling for parameter estimation in diffusion type processes. In R.J.-B. Wets (ed.), *Stochastic Systems: Modeling, Identification and Optimization, I.* Amsterdam: North-Holland Publishing Company, 1976.
40. Liptser, R.S. & A.N. Shiryayev. *Statistics of Random Processes II: Applications.* New York: Springer-Verlag, 1978.
41. Ljung, L. Convergence analysis of parametric identification methods. *IEEE Transactions on Automatic Control* AC-23 (1978): 770–783.
42. Lo, A. Statistical tests of contingent claims asset-pricing models: A new methodology. *Journal of Financial Economics* 17 (1986): 143–174.
43. Loges, W. Girsanov's theorem in Hilbert space and an application to the statistics of Hilbert space-valued stochastic differential equations. *Stochastic Processes and their Applications* 17 (1984): 243–263.
44. Marsh, T. & E. Rosenfeld. Stochastic processes for interest rates and equilibrium bond prices. *Journal of Finance* 38 (1983): 635–645.
45. McLeish, D.L. Dependent central limit theorems and invariance principles. *Annals of Probability* 2 (1974): 620–628.
46. Ogden, J. An analysis of yield curve notes. *Journal of Finance* 42 (1987): 99–110.
47. Pardoux, E. & D. Talay. Discretization and simulation of stochastic differential equations. *Acta Applicandae Mathematicae* 3 (1985): 23–47.
48. Park, C. & J. Beekman. Stochastic barriers for the Wiener process. *Journal of Applied Probability* 20 (1983): 338–348.

49. Park, C., J. Beekman, & S. Paranjape. Probabilities of Wiener paths crossing differentiable curves. *Pacific Journal of Mathematics* 53 (1974): 579–583.
50. Park, C., J. Beekman, & F. Schuurmann. Evaluations of barrier-crossing probabilities and Wiener paths. *Journal of applied Probability* 13 (1976): 267–275.
51. Park J. & P.C.B. Phillips. Statistical inference in regressions with integrated processes: Part 1. Cowles Foundation Discussion Paper No. 811, November 1986.
52. Perron, P. Testing consistency with varying sampling frequency. Working Paper, University of Montreal, August 1987.
53. Phillips, A.W. The estimation of parameters in systems of stochastic differential equations. *Biometrika* 46 (1959): 67–76.
54. Phillips, P.C.B. The structural estimation of a stochastic differential equation system. *Econometrica* 40 (1972): 1021–1041.
55. Phillips, P.C.B. The problem of identification in finite parameter continuous-time processes. *Journal of Econometrics* 1 (1973): 351–362.
56. Phillips, P.C.B. The estimation of some continuous-time models. *Econometrica* 42 (1974): 803–823.
57. Phillips, P.C.B. Time series regression with unit roots. *Econometrica* 55 (1987): 277–302.
58. Phillips, P.C.B. Regression theory for near integrated time series. Cowles Foundation Discussion Paper No. 781-R, January 1987.
59. Prakasa Rao, B.L.S. Maximum likelihood estimation for Markov processes. *Annals of the Institute of Statistics and Mathematics* 24 (1972): 333–345.
60. Rao, N., J. Borwankar, & D. Ramakrishna. Numerical solution of Itô integral equations. *SIAM Journal of Control* 12 (1974): 123–139.
61. Rosenfeld, E. *Stochastic Processes of Common Stock Returns: An Empirical Examination.* Unpublished Ph.D. thesis, Sloan School of Management, M.I.T., February 1980.
62. Roussas, G.G. Asymptotic inference in Markov processes. *Annals of Mathematical Statistics* 36 (1965): 978–992.
63. Rudin, W. *Functional Analysis.* New York: McGraw-Hill Book Company, 1973.
64. Rumelin, W. Numerical treatment of stochastic differential equations. *SIAM Journal of Numerical Analysis* 19 (1982): 604–613.
65. Schuss, Z. *Theory and Applications of Stochastic Differential Equations.* New York: John Wiley and Sons, 1980.
66. Shiller, R. & P. Perron. Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters* 18 (1985), 381–386.
67. Siegmund, D. Boundary crossing probabilities and statistical applications. Stanford University Department of Statistics Technical Report No. E&S/NSF230, June 1985.
68. Sims, C.A. Discrete approximations to continuous-time distributed lags in econometrics. *Econometrica* 39 (1971): 545–563.
69. Skorohod, A.V. *Studies in the Theory of Random Processes.* New York: Dover Publications, 1965.
70. Tugnait, J.K. Identification and model approximation for continuous-time systems on finite parameter sets. *IEEE Transactions on Automatic Control* 25 (1980): 1202–1206.
71. Tugnait, J.K. Global identification of continuous-time systems with unknown noise covariance. *IEEE Transactions on Information Theory* 28 (1982): 531–536.
72. Tugnait, J.K. Continuous-time system identification on compact parameter sets. *IEEE Transactions on Information Theory* 31 (1985): 652–659.