NBER TECHNICAL PAPER SERIES

MAXIMUM LIKELIHOOD ESTIMATION OF GENERALIZED ITO PROCESSES WITH DISCRETELY SAMPLED DATA

Andrew W. Lo

Technical Working Paper No. 59

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1986

This paper has benefitted considerably from the comments of Lars Hansen, Jerry Hausman, A. Craig McKinlay, 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, the University of Wisconsin-Madison, and three anonymous referees. 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. The research reported here is part of the NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Technical Working Paper #59
September 1986

Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data

ABSTRACT

In this paper, we consider the parametric estimation problem for continuous time stochastic processes described by general first-order nonlinear stochastic differential equations of the Itô type. We characterize the likelihood function of a discretely-sampled set of observations as the solution to a functional partial differential equation. The consistency and asymptotic normality of the maximum likelihood estimators are explored, and several illustrative examples are provided.

Andrew W. Lo
Department of Finance
Wharton School
University of Pennsylvania
2254 Steinberg-Dietrich Hall
Philadelphia, PA 19104-6367

1. Introduction.

For purposes of modelling the evolution of various random systems, the theory of continuous-time stochastic processes has become indispensible. In particular, the use of continuous-time processes described by 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 [1951] 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 example, in the case of the well-known Black-Scholes [1973] 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 and range from labor economics to investment theory.[2] However, to date relatively little research in economics has been devoted to the econometric estimation problems associated with such continuous-time processes.[3] This is particularly surprising since the modelling of uncertainty via Itô processes usually yields a very specific statistical specification for purposes of estimation.

In this paper, we consider the parametric estimation problem for Itô processes using the method of maximum likelihood (ML).[4] The main result of the paper is a characterization of the likelihood function as a solution to a particular functional partial differential equation. We also discuss the asymptotic properties of the resulting estimators and provide some illustrative examples. Since the general theory of maximum likelihood estimation has been well-studied in several related areas of research, our discussion of the estimators' asymptotic behavior will be mainly expository in nature and somewhat abbreviated in collecting some of the more relevant results from the exant literature.

Because the theory of statistical inference for an alternative class of continuous-time processes is now well established and comprehensively developed in Bergstrom[5] [1976, 1983, 1984], a few remarks concerning the relation of that literature to inference for generalized Itô processes are in order before we begin our analysis. One important distinction between Itô processes and those studied by 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.[6] 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.[7] 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. In particular, Bergstrom [1983, 1984] 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 then well-specified. This, of course, does not obtain for functions of nth-order processes. In addition, 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.[8] For purposes of exposition, the loss in concreteness does not seem justified by such generality.

One further aspect of Itô processes which is distinct from nth-order processes is the Markovian nature of the former. It will be shown that 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 quite restrictive and may not be applicable to certain economic processes of interest. 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 intimately upon the underlying economic model at hand and must be considered on a case-by-case basis.[9]

The organization of the paper is as follows. In Section 2 the general parametric estimation problem is posed and the maximum likelihood procedure is discussed. Section 3 considers the asymptotic properties of the maximum likelihood estimators obtained in Section 2, and we conclude in Section 4.

2. Formulation of 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 $\{X(t): t \in T \subset \mathbf{R}^+, X(t) \in \mathbf{S} \subset \mathbf{R}\}$ be a stochastic process defined on a complete probability space $(\Omega, \mathbf{F}, \mu)$ and suppose X(t) satisfies the following stochastic integral equation:

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

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 f, g, and h are known functions which depend upon (X, t) and an unknown parameter vector $\bar{\theta} = (\alpha', \beta', \gamma')'$. Note that because the integrand g 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 [1983].[10] The integral may, however, be interpreted in the sense of Itô [1951] if the functions f, g, and h satisfy the following restrictions:

Assumption 1: Let $\{\mathbf{F}_t: t \in \mathbf{T}\}$ denote a right-continuous filtration σ-field defined on $(\Omega, \mathbf{F}, \mu)$ and let the pure Wiener process $\{W(t): t \in \mathbf{T}\}$ be adapted to this filtration.[11]

Assumption 2: If $\mathbf{B}$ is the σ-field of Borel sets on $\mathbf{R}^+$, then for all $\bar{\theta}$ in the parameter space $\bar{\mathbf{\Theta}} \equiv \mathbf{A} \times \mathbf{B} \times \mathbf{\Gamma}$ the functions f, g, and h are measurable in the product σ-field $\mathbf{B} \times \mathbf{F}$.

Assumption 3: For all $\bar{\theta} \in \bar{\boldsymbol{\theta}}$ and $t \in \mathbf{T}$ the functions f, g, h depend only upon X(t) and t, implying that the functions are trivially adapted to the filtration $\{\mathbf{F}_t : t \in \mathbf{T}\}$ and are therefore nonanticipating with respect to that filtration.

Assumption 4: For all $\bar{\theta} \in \bar{\boldsymbol{\theta}}$ the functions f, g, and h satisfy the following inequalities almost surely:

$$\int_{t_0}^{t} |f| d\tau < \infty \tag{2a}$$

$$\int_{t_0}^{t} |g|^2 d\tau < \infty \tag{2b}$$

$$\int_{t_0}^{t} |h|^2 d\tau < \infty . \tag{2c}$$

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

$$dX(t) = f(X, t; \alpha)dt + g(X, t; \beta)dW(t) + h(X, t; \gamma)dN_\lambda(t) . \tag{3}$$

Further assumptions are required in order to insure the existence and uniqueness of a solution to the stochastic integral (differential) equation given by equation (1) (equation (3)). They are:

Assumption 5: There exists some constant $K > 0$ such that the functions f, g, and h satisfy the following conditions for all $X, X' \in \mathbf{S}$ and $t, t' \in \mathbf{T}$:

(i) $$|f(X,t)-f(X',t)| + |g(X,t)-g(X',t)| + |h(X,t)-h(X',t)| \le K|X-X'| \tag{4a}$$

(ii) $$|f(X,t)-f(X,t')| + |g(x,t)-g(X,t')| + |h(X,t)-h(X,t')| \le K|t-t'| \tag{4b}$$

(iii) $$f^2(X, t) + g^2(X, t) + h^2(X, t) \le K^2(1 + X^2) . \tag{4c}$$

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

Assumption 6: The functions f, g, and h are twice continuously differentiable in (X, t) and three times continuously differentiable in $\bar{\theta}$; $g \neq 0$ a.e. in X for all $(t, \beta) \in \mathbf{T} \times \mathbf{B}$; the function $\tilde{h}(X, t; \gamma) \equiv X + h(X, t; \gamma)$ is bijective and $|\frac{\partial h}{\partial X} + 1| \neq 0$ for all $(t, \gamma) \in \mathbf{T} \times \mathbf{\Gamma}$ and $X \in \mathbf{S}$.

Assumption 7: 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 $\boldsymbol{\Theta} \equiv \bar{\boldsymbol{\Theta}} \times \boldsymbol{\Lambda}$.

Although the formal definition of the Itô process is somewhat abstract, its basic time series properties (e.g., conditional expectation, autocorrelation function, etc.) may often be deduced explicitly. In the Appendix, the general approach to calculating population moments is discussed and an illustrative example is provided.

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. 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), we now consider the maximum likelihood estimator $\hat{\theta}_{ML}$ of θ_0. Let $P(X_0, X_1, \ldots, X_n; \theta)$ denote the finite-dimensional distribution of the sample X associated with the process X(t) and let $\rho(X; \theta)$ denote the density representation of P.[12] When considered a function of θ, this joint density is obviously the desired likelihood function. Since X(t) is a Markov process (see Arnold [1974, chapter 9]), the joint density ρ may be re-written 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 1-7, the transition densities ρ_k solve the following functional partial differential equation:

$$\frac{\partial}{\partial t}[\rho_k] = -\frac{\partial}{\partial X}[f\rho_k] + \frac{1}{2}\frac{\partial^2}{\partial X^2}(g^2\rho_k) - \lambda\rho_k + \lambda\tilde{\rho}_k\left|\frac{\partial}{\partial X}[\tilde{h}^{-1}]\right| \quad (6)$$

subject to:

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

and any other relevant boundary conditions, where $\tilde{h}$ is defined in Assumption 6, $\tilde{\rho}_k \equiv \rho_k(\tilde{h}^{-1}, t)$, 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 [1984]) we have:

$$d\psi = \left[\psi_X f + \tfrac{1}{2}\psi_{XX}g^2\right]dt + \psi_X g\,dW + \left[\psi(X+h) - \psi(X)\right]dN_\lambda \quad (8a)$$

where

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

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:

$$D_{P,k}[\psi] = E_{t_k}\left[\psi_X f + \tfrac{1}{2}\psi_{XX}g^2\right] + \lambda E_{t_k}\left[\psi(X+h) - \psi(X)\right] . \quad (9)$$

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

$$D_{P,k}[\psi] = \int_S \{\psi_X f + \tfrac{1}{2}\psi_{XX}g^2 + \lambda[\psi(X+h) - \psi(X)]\}\rho_k(X, t)dX \quad (10a)$$

$$= \int_S \left[-\psi \frac{\partial}{\partial X}(f\rho_k) + \tfrac{1}{2}\psi \frac{\partial^2}{\partial X^2}(g^2\rho_k) - \psi\lambda\rho_k\right]dX$$
$$+ \lambda \int_S \psi(X+h)\rho_k dX . \quad (10b)$$

where the second equality is obtained by integrating by parts and collecting terms. By Assumption 6 the Inverse Function Theorem guarantees the existence of $\tilde{h}^{-1}$ such that $X = \tilde{h}^{-1}(\tilde{h}(X,t;\gamma), t; \gamma)$. Using the change of variables formula, we have:

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

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

We then conclude that

$$D_{P,k}[\psi] = \int_S \{-\frac{\partial}{\partial X}(f\rho_k) + \tfrac{1}{2}\frac{\partial^2}{\partial X^2}(g^2\rho_k) - \lambda\rho_k$$
$$- \lambda\tilde{\rho}_k\left|\frac{\partial}{\partial X}\tilde{h}^{-1}\right|\}\psi(X)\, dX . \quad (12)$$

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

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

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

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

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}.

Q.E.D.

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 f, g, and h. However, when the existence of a density representation for a specific process has been assured by other means, equation (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 h = 0 (pure diffusion) and f and g satisfy the following reducibility condition:

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

it may be shown (see Schuss [1980, 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:[13]

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

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

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

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 differential rule then yields $dY = [\alpha - \frac{\beta^2}{2}]dt + \beta dW$ 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, equation (14) provides a conclusive check for such conjectured density representations as the following example illustrates:

Example 1. (Lognormal diffusion and jump process.)

Suppose we seek the likelihood function corresponding to the process X(t) which satisfies the following SDE:

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

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:[14]

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

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

In addition to the initial condition (14b), the solution of equation (14a) often depends critically upon particular auxiliary restrictions placed on the process $X(t)$ as a result of economic considerations. For example, when $X(t)$ represents an asset-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.)[15]

Let $X(t)$ satisfy the following SDE:

$$dX(t) = \mu dt + \sigma dW(t) \quad , \quad X(0) = X_0 > 0 \ , \qquad (21)$$

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 <u>between</u> 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:

$$\frac{\partial \rho_k}{\partial t} = \frac{1}{2}\sigma^2 \frac{\partial^2 \rho_k}{\partial X^2} - \mu \frac{\partial \rho_k}{\partial X} \tag{22a}$$

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

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" this may be solved to yield:

$$\rho_k(X_k, t_k|X_{k-1}, t_{k-1}) = [2\Pi\sigma^2\Delta t_k]^{-\frac{1}{2}}[\exp[- \frac{(X_k - X_{k-1} - \mu\Delta t_k)^2}{2\sigma^2\Delta t_k}] - \tag{23}$$

$$\exp[- \frac{2\mu X_{k-1}}{\sigma^2} - \frac{(X_k + X_{k-1} - \mu\Delta t_k)^2}{2\sigma^2\Delta t_k}]]$$

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:

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

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, . . ., n is given by:

$$\rho_k = [2\Pi\sigma^2\Delta t_k]^{-\frac{1}{2}}[\exp[-\frac{(X_k - X_{k-1} - \mu\Delta t_k)^2}{2\sigma^2\Delta t_k}] - \tag{26}$$

$$\exp[-\frac{2\mu X_{k-1}}{\sigma^2} - \frac{(X_k + X_{k-1} - \mu\Delta t_k)^2}{2\sigma^2\Delta t_k}]] +$$

$$P[\text{Absorption in } [t_{k-1}, t_k)]\delta(X) \; , \; X \geq 0 \; .$$

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. Note that although the conditional likelihood function in Example 2 is indeed a solution for equation (14), it contains a Dirac δ-function and is therefore not a function in the usual sense.[16] 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.[17] 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 following example:

Example 3. (Pure jump process.)

Let X(t) solve the SDE $dX(t) = dN_\lambda(t)$ and consider the transition density $\rho(X, t)$ of X conditional upon X(t) = 0 at time t = 0. Since in this case X(t) is just a Poisson counter with intensity λ, its transition density may be expressed as:

$$\rho(X, t) = \sum_{k=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!} \delta(X - k) . \tag{27}$$

For the Poisson counter the forward equation (14) reduces to:

$$\frac{\partial \rho}{\partial t} = \lambda[\rho(X - 1, t) - \rho(X, t)] . \tag{28}$$

This may be verified by explicitly calculating the derivative of ρ with respect to t and simplifying:

$$\frac{\partial \rho}{\partial t} = \sum_{k=0}^{\infty} \frac{-\lambda e^{-\lambda t}(\lambda t)^k}{k!} \delta(X - k) + \sum_{k=0}^{\infty} \frac{\lambda k e^{-\lambda t}(\lambda t)^{k-1}}{k!} \delta(X - k) \tag{29a}$$

$$= -\lambda \sum_{k=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!} \delta(X - k) + \lambda \sum_{k=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!} \delta(X - 1 - k) \tag{29b}$$

$$= \lambda[\rho(X - 1, t) - \rho(X, t)] . \tag{29c}$$

Even though the solution to (14) in this case is a generalized function which is associated with a measure not absolutely continuous with respect to Lebesgue measure, nevertheless maximum likelihood estimation is still feasible with respect to the proper carrier measure. In this example, the counting carrier measure ν yields a likelihood function which is simply the probability:[18]

$$\frac{dP[X(t)|X(0) = 0]}{d\nu} = \frac{e^{-\lambda t}(\lambda t)^{X(t)}}{X(t)!} . \tag{30}$$

Having characterized the likelihood function as the solution to equation (14), we now suppose its existence and define the maximum likelihood estimator in the usual manner:

$$\hat{\theta}_{ML} \equiv \arg \underset{\theta \in \boldsymbol{\Theta}}{\text{Max}}\, G(\theta;\, X) \qquad \text{where} \tag{31a}$$

$$G(\theta;\, X) \equiv \ln \rho_0(X_0,\, t_0) + \sum_{k=1}^{n} \ln \rho k(X_k,\, t_k | X_{k-1},\, t_{k-1};\, \theta) \tag{31b}$$

$$\equiv \ell_0(X_0,\, t_0) + \sum_{k=1}^{n} \ell_k(X_k,\, t_k | X_{k-1},\, t_{k-1};\, \theta)\ . \tag{31c}$$

In the next section, the asymptotic properties of $\hat{\theta}_{ML}$ are discussed.

3. Asymptotic Properties of the Maximum Likelihood Estimator.

Because the estimator $\hat{\theta}_{ML}$ is based on the sample X which is neither independently nor identically distributed, the standard proofs of consistency and asymptotic normality (as in Huber [1967] for example) are not directly applicable. However, we may appeal to the results of several authors who have investigated the asymptotic properties of maximum likelihood estimators with dependent heterogeneous osbervations. In doing so, it is useful to distinguish between two cases: asymptotics based upon sampling intervals which are positive in the limit, and more frequent sampling of X(t) in a fixed span of calendar time [0, T]. More precisely, define the limiting sampling interval Δ as the following:

$$\Delta \equiv \lim_{n\to\infty} (\inf_{k\leq n} |\Delta t_k|)\ . \tag{32}$$

If Δ is strictly positive and two further regularity conditions are satisfied:

(R1) ρ_k is time-homogeneous for all $\theta \in \boldsymbol{\Theta}$.

(R2) X(t) admits a unique stationary or "steady state" distribution for all $\theta \in \boldsymbol{\Theta}$.

then it has been shown by Billingsley [1961], Roussas [1965], and Prakasa Rao [1972] that $\hat{\theta}_{ML}$ is consistent and asymptotically normal, i.e.,

$$\operatorname*{plim}_{n\to\infty} \hat{\theta}_{ML} = \theta_0 \tag{33a}$$

$$\sqrt{n}\,(\hat{\theta}_{ML} - \theta_0) \overset{a}{\sim} N\left(0,\ I^{-1}(\theta_0)\right) \quad \text{where} \tag{33b}$$

$$I(\theta_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E\left[- \frac{\partial^2 \ell(X_k | X_{k-1};\ \theta_0)}{\partial\theta\partial\theta'}\right] . \tag{33c}$$

Note that conditions (R1) and (R2) are necessary and sufficient conditions for the strict stationarity of $X(t)$.[19] Assuming strict stationary and a strictly positive limiting sampling interval essentially reduces the sample X to observations of a strictly stationary discrete-time Markov process, for which the references cited above have demonstrated the consistency and asymptotic normality of $\hat{\theta}_{ML}$.

The assumption of the strict stationarity of $X(t)$ is quite restrictive and excludes many processes of interest such as the simple Wiener process with drift or the Ornstein-Uhlenbeck process. In addition, it is of some interest to consider asymptotic properties of processes when the limiting sampling interval is zero.[20] A somewhat less restrictive set of conditions for consistency and asymptotic normality may be obtained by appealing to Crowder's [1978] results. The following two rather contrived examples demonstrate the importance of checking the Crowder conditions, and we refer the reader to Crowder's original paper for the precise statement of those conditions.

Example 4. (Wiener process with $\Delta > 0$.)

Let X be a sample of equally spaced observations of the process $X(t)$ where:

$$dX = \mu dt + \sigma dW\ ,\ X(t_0 = 0) = X_0\ .$$

Since $X(t)$ is clearly not stationary in the "wide" or "second-order" sense (observe that $E[X(t)] = \mu t + E[X_0]$), it cannot be strictly stationary hence the results of Billingsley, Roussas, and Prakasa Rao are not strictly applicable. However, since the Crowder conditions may be shown to obtain in this case, the consistency and asymptotic normality of $\hat{\theta}_{ML}$ are insured. Indeed, the likelihood function G and corresponding estimator $\hat{\theta}_{ML} \equiv (\hat{\mu}_{ML} \quad \hat{\sigma}^2_{ML})'$ may be readily derived as:

$$G(\theta;\, X) = -\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{k=1}^{n} \frac{(X_k - X_{k-1} - \mu\Delta t)^2}{2\sigma^2\Delta t} \tag{33}$$

$$\hat{\mu}_{ML} = \frac{1}{n\Delta t}\sum_{k=1}^{n}(X_k - X_{k-1}) \tag{34a}$$

$$\hat{\sigma}^2_{ML} = \frac{1}{n\Delta t}\sum_{k=1}^{n}[X_k - X_{k-1} - \hat{\mu}_{ML}]^2 \tag{34b}$$

where $\Delta t \equiv t_k - t_{k-1}$, $k = 1, \ldots, n$. The ML estimators (33) are then the usual ones for the mean and variance of a normally distributed random variable.

Example 5. (Wiener process with $\Delta = 0$.)

Let $X(t)$ be the same process as in Example 4 but suppose that equally-spaced observations are taken in a <u>fixed</u> interval of calendar time $[0, T]$ so that $\Delta t \equiv T/n$ hence $\Delta = 0$. Consider the ML estimator of μ:

$$\hat{\mu}_{ML} = \frac{1}{T}\sum_{k=1}^{n}(X_k - X_{k-1}) = \frac{X(T) - X(0)}{T}\,. \tag{35}$$

It is obvious that as n approaches infinity, $\hat{\mu}_{ML}$ does not converge in probability to the true parameter μ. In this case, Crowder's [1978] condition

(2.3) is violated. More specifically, the upper left diagonal element of the information matrix of θ:

$$I(\theta, n) \equiv E[-\frac{\partial^2 G}{\partial\theta\partial\theta'}] = \begin{vmatrix} \frac{T}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{vmatrix} \tag{36}$$

does not tend to infinity as n increases without bound. Loosely speaking, this implies that information about μ does not accrue as more observations of X(t) are taken. Crowder terms such parameters as <u>transient</u> and specifies conditions under which consistent estimation of the remaining parameters is still possible in the presence of transient parameters. In this example, it may be shown that such conditions are satisfied for the variance parameter, therefore $\hat{\sigma}^2_{ML}$ is consistent and $\sqrt{n}(\hat{\sigma}^2_{ML} - \sigma^2)$ has the limiting distribution $N(0, 2\sigma^4)$. When the limiting sampling interval is zero, the Crowder conditions must be checked on a case-by-case basis.

4. Conclusion.

In this paper, we have considered the parametric estimation problem for continuous-time stochastic processes which satisfy general first-order nonlinear stochastic differential equations of the Itô type. By exploiting the Markov nature of such processes, under suitable regularity conditions consistent and asymptotically normal estimators of the unknown parameters may be obtained even if the process is discretely-sampled and sampled at unequally spaced intervals. Because the asymptotic properties of maximum likelihood estimators are well established, statistical inference for many continuous time asset-pricing models may readily be performed. One example given is the testing of contingent claims model, which is pursued in more detail in Lo [1986]. Other possible applications include the empirical estimation and testing of general equilibrium asset-pricing models such as those in Chamberlain [1985] and Cox, Ingersoll, and Ross [1985a, 1985b] for example.

Of course, since the results in this paper are exclusively asymptotic in nature, the finite-sampling properties must be studied through Monte Carlo simulations for each application separately. This is especially important for asymptotic inference when the limiting sampling interval is zero since, in the case of "continuous sampling" over a fixed length of calendar time, consistency does not always obtain and the nominal size and power of the usual statistical tests may differ substantially from their actual values (see, for example, Shiller and Perron [1985]).[21] Also, in cases where the density representation of the process is analytically intractable, numerically simulating the behavior of a discrete-time approximation to the process may be fruitful.[22]

APPENDIX

In this appendix, the calculation of population moments of Itô processes is considered. Let $X(t)$ satisfy the following stochastic differential equation (the dependence of X upon parameters θ has been suppressed for notational convenience):

$$dX = f(X, t)dt + g(X, t)dW + h(X, t)dN_\lambda \ , \ X(t_0) = X_0 \ . \tag{A1}$$

To compute the conditional expectation of $X(t)$ (conditional upon the initial value $X(t_0) = X_0$) we perform the following (heuristic) calculation:

$$\begin{aligned} E_0[dX] = dE_0[X] &= E_0[f]dt + E_0[gdW] + E_0[hdN_\lambda] \\ &= E_0[f]dt + E_0[g]E_0[dW] + E_0[h]E_0[dN_\lambda] \\ &= E_0[f]dt + \lambda E_0[h]dt \end{aligned}$$

therefore

$$\frac{d}{dt} E_0[X] = E_0[f] + \lambda E_0[h] \tag{A2}$$

where $E_0[\cdot] \equiv E[\cdot \mid X_0]$, $E_0[dW] = 0$, and $E_0[dN_\lambda] = \lambda dt$.

Note that the conditional expectations of the products gdW and hdN_λ are equal to the product of their expectations respectively because g and h are assumed to be nonanticipating hence they are independent of the stochastic differentials dW and dN_λ. The result is a (functional) ordinary differential equation in $E_0[X(t)]$, $E_0[f(X(t), t)]$, and $E_0[h(X(t), t)]$ subject to the initial condition $E_0[X(t_0)] = X_0$. If f and h are linear in X then (A2) reduces to an ordinary differential equation in the conditional expectation of X which may be solved under weak regularity conditions. The heuristic computation of taking expectations and "dividing by dt" is more rigorously defined by the Dynkin operator $D_t \equiv \frac{d}{dt} E_0[\cdot]$ (see, for example, Karlin and Taylor [1981]).

In order to obtain the k-th (noncentral) conditional moment, a similar procedure is adopted. First, using Itô's differentiation rule, the stochastic differential equation of $F(X) \equiv X^k$ is derived, yielding:

$$dX^k = (kX^{k-1}f + \frac{1}{2}k(k-1)X^{k-2}g^2)dt + kX^{k-1}gdW + ([X+h]^k - X^k)dN_\lambda \ . \quad (A3)$$

Applying the Dynkin operator to (A3) results in the expression:

$$\frac{d}{dt} E_0[X^k] = kE_0[X^{k-1}f] + \frac{1}{2}k(k-1)E_0[X^{k-2}g^2] + \lambda E_0[(X+h)^k - X^k] \ . \quad (A4)$$

This relation also does not necessarily yield an ordinary differential equation in the k-th moment but, depending upon the coefficient functions f, g, and h, may involve expectations of nonlinear functions of X. This would suggest that a reasonable restriction on the coefficient functions is that they be polynomials in X. Of course, even in this case the k-th moment equation may contain moments of order higher than k in which case the system of k differential equations would not be closed without further restrictions. In a similar manner, conditional autocovariances, cross-variances, and correlations may be computed. Unconditional moments may also be obtained by specifying a particular marginal distribution for the initial value X_0 and then taking iterated expectations in the usual way. Central moments may be computed in the obvious way. The following example presents some illustrative calculations:

Example 6. (Random telegraph wave.)

Let X(t) satisfy the SDE:

$$dX(t) = -2X(t)dN_\lambda(t) \quad , \ X(t_0 = 0) = X_0 = 1 \ .$$

This process is known as the random telegraph wave since it jumps randomly between only 2 states: 1 and -1. Its conditional expectation, variance, autocovariance and autocorrelation functions may be calculated as:

$$\frac{d}{dt} E_0[X] = -2\lambda E_0[X]$$

$$E_0[X] = X_0 e^{-2\lambda t} = e^{-2\lambda t}$$

$$dX^2 = \left[(X - 2X)^2 - X^2\right]dN_\lambda = 0 \qquad \text{(Itô's Lemma)}$$

$$\frac{d}{dt} E_0[X^2] = 0$$

$$E_0[X^2] = k_0 = 1 \quad \text{since } E_0\left[X^2(0)\right] = 1$$

$$Var_0\left[X(t)\right] = 1 - e^{-4\lambda t}$$

$$F\left(X(\tau)\right) \equiv X(t)X(\tau) \qquad 0 \le t \le \tau$$

$$dF\left(X(\tau)\right) = X(t)dX(\tau) = -2X(t)X(\tau)dN_\lambda(\tau)$$

$$\frac{d}{d\tau} E_0\left[X(t)X(\tau)\right] = -2\lambda E_0\left[X(t)X(\tau)\right]$$

$$E_0\left[X(t)X(\tau)\right] = k_1 e^{-2\lambda(\tau-t)} = e^{-2\lambda(\tau - t)} \quad \text{since} \quad E_0\left[X^2(t)\right] = 1$$

$$E_0\left[X(t)X(t + \alpha)\right] = e^{-2\lambda\alpha}$$

Thus we have:

$$E_0\left[X(t)\right] = e^{-2\lambda t}$$

$$Var_0\left[X(t)\right] = 1 - e^{-4\lambda t}$$

$$Cov_0\left[X(t),\ X(t + \alpha)\right] = e^{-2\lambda\alpha}[1 - e^{-4\lambda t}]$$

$$Corr_0\left[X(t),\ X(t + \alpha)\right] = e^{-2\lambda\alpha}\left[\frac{1 - e^{-4\lambda t}}{1 - e^{-4\lambda(t+\alpha)}}\right]^{\frac{1}{2}} .$$

FOOTNOTES

[1]More specifically, 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 [1984] has shown that any finite-state continuous-time jump process may be expressed as a generalized Itô process).

[2]See, for example, Hausman and Wise [1983] and Abel [1983].

[3]Most of the empirical applications seem to be within asset-pricing studies. See, for example, Rosenfeld [1980]; Marsh and Rosenfeld [1983]; Grossman, Melino, and Shiller [1985]; Ball and Torous [1985]; and Lo [1986].

[4]Note that in this context, the term "estimation" is used in the classical parametric statistical sense. This is in contrast to its usage in the engineering and stochastic control literature, in which estimation is identified 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.

[5]Other examples include A. W. Phillips [1959]; P. C. B. Phillips [1972, 1973, 1974]; Hansen and Sargent [1981, 1983a, 1983b]; Borkar and Bagchi [1982]; Christiano [1984, 1985a, 1985b]; Harvey and Stock [1985]; and Grossman, Melino, and Shiller [1985].

[6]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 in contrast 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 [1984] have shown, continuous-time equilibrium price processes generated by frictionless markets must be of unbounded variation.

[7]For practical purposes, this quickly becomes intractable when <u>systems</u> of nth-order SDE's are considered, as in Bergstrom's approach.

[8]See, for example, Skorokhod [1965, chapters 2 and 3].

[9]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 [1971]; Bergstrom [1983, 1984]; Christiano [1984, 1985a]; and Grossman, Melino, and Shiller [1985].

[10]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.

[11]That is, let $\{\mathbf{F}_t : t \in \mathbf{T}\}$ be a sequence of sub-σ-fields of the σ-field $\mathbf{F}$ such that:

(i) $\mathbf{F}_t \subset \mathbf{F}_s$ for $t \leq s$

(ii) $\mathbf{F}_t = \bigcap_{\tau > t} \mathbf{F}_\tau$

and let W(t) be $\mathbf{F}_t$-measurable for all $t \in \mathbf{T}$.

[12]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.

[13]Furthermore, this transformation F may be explicitly derived by solving a simple ordinary differential equation given in Schuss [1980, chapter 4.1].

[14]More formally, we define:

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

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

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

[15]For perhaps its first econometric implementation, see Hausman and Wise [1983]. 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 [1974]; Park and Schuurmann [1976]; Park and Beekman [1983]; and Siegmund [1985].

[16]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 $C_c^\infty(\mathbf{R})$. 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 [1964, chapter 1] or Rudin [1973, chapters 6 and 8] for a formal development of this theory.

[17] 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 [1971] for a more detailed discussion.

[18] Note that as long as the diffusion component of an Itô process does not vanish (as is assumed in Assumption 6), the transition probability measures will be absolutely continuous with respect to Lebesgue measure. Loosely speaking, this is due to the fact that the convolution of two densities is as smooth as the "smoother" of the two (see, for example, Chung [1974, chapter 6.1, problem 6]).

[19] See Arnold [1974, chapter 2.2].

[20] Indeed, several authors have examined the properties of maximum likelihood estimators of certain pure diffusion processes when sampling is continuous. See, for example, Brown and Hewitt [1975], Le Breton [1976], Liptser and Shiryayev [1978, chapter 17], and Basawa and Prakasa Rao [1980, chapter 9.5].

[21] Phillips' [1985] theoretical results concerning asymptotic theory for the continuous data-recording case is especially relevant in explaining the source of such differences. See also Le Breton [1976] and Basawa and Prakasa Rao [1980, chapter 9.4.2].

[22] For the numerical simulation of Itô processes, see Rao, Borwankar, and Ramakrishna [1974]; Rumelin [1982]; and Pardoux and Talay [1985].

REFERENCES

Abel, A.: "Optimal Investment Under Uncertainty," American Economic Review, 73(1983), 228-233.

Arnold, L.: Stochastic Differential Equations: Theory and Applications. New York: John Wiley and Sons, 1974.

Ball, C. A. and W. Torous: "On Jumps in Common Stock Prices and Their Impact on Call Option Pricing," Journal of Finance, 40(1985), 155-173.

Basawa, I. V. and B. L. S. Prakasa Rao: Statistical Inference for Stochastic Processes. New York: Academic Press, 1980.

Bergstrom, A. R.: Statistical Inference in Continuous Time Economic Models. Amsterdam: North-Holland, 1976.

_________: "Gaussian Estimation of Structural Parameters in Higher Order Continuous Time Dynamic Models," Econometrica, 51(1983), 117-152.

_________: Continuous Time Stochastic Models and Issues of Aggregation Over Time," in Handbook of Econometrics, Vol. II, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland Publishing Company, 1984.

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

Black, F. and M. Scholes: "The Pricing of Options and Corporate Liabilities," Journal of Political Economy, 81(1973), 637-654.

Borkar, B. and A. Bagchi: "Parameter Estimation in Continuous Time Stochastic Processes," Stochastics, 8(1982), 193-212.

Brockett: "Lecture Notes on Nonlinear Stochastic Control," Unpublished course notes, Harvard University, Spring 1984.

Brown, B. M. and J. I. Hewitt: "Asymptotic Likelihood Theory for Diffusion Processes," Journal of Applied Probability, 12(1978), 228-238.

Chamberlain, G.: "Asset Pricing in Multiperiod Securities Markets," University of Wisconsin-Madison S.S.R.I. Working Paper No. 8510, 1985.

Christiano, L.: "The Effects of Aggregation Over Time on Tests of the Representative Agent Model of Consumption," mimeo, December 1984.

_________: "A Critique of Conventional Treatments of the Model Timing Interval in Applied Econometrics," mimeo, January 1985a.

_________: "Estimating Continuous Time Rational Expectations Models in Frequency Domain: A Case Study," Carnegie-Mellon University Working Paper No. 34-84-85, April 1985b.

Chung, K. L.: A Course in Probability Theory. New York: Academic Press, 1974.

Cox, D. R. and H. D. Miller: The Theory of Stochastic Processes. New York: Chapman and Hall, 1965.

Cox, J., Ingersoll, J. and S. Ross: "An Intertemporal General Equilibrium Model of Asset Prices," Econometrica, 53(1985a), pp. 363-384.

__________: "A Theory of the Term Structure of Interest Rates," Econometrica, 53(1985b), pp. 385-408.

Crowder, M. J.: "Maximum Likelihood Estimation for Dependent Observations," Journal of the Royal Statistical Society, Series B, 38(1976), 45-53.

Gel'fand, I. M. and G. E. Shilov: Generalized Functions Volume I: Properties and Operations. New York: Academic Press, 1964.

Grossman, S., Melino, A., and R. Shiller: "Estimating the Continuous Time Consumption Based Asset Pricing Model," N.B.E.R. Working Paper No. 1643, June 1985.

Hansen, L. and T. Sargent: "Formulating and Estimating Continuous Time Rational Expectations Models," unpublished manuscript, August 1981.

__________: "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities," Econometrica, 51(1983a), 377-388.

__________: "Identification of Continuous Time Rational Expectations Models from Discrete Time Data," Federal Reserve Bank of Minneapolis Staff Report 73, March 1983b.

Harrison, J. M., Pitbladdo, R. and S. M. Schaefer: "Continuous Price Processes in Frictionless Markets Have Infinite Variation," Journal of Business, 57(1984), 353-365.

Harvey, A. and J. Stock: "The Estimation of Higher Order Continuous Time Autoregressive Models," Econometric Theory, 1(1985), 97-117.

Hausman, J. and D. Wise: "Social Security, Health Status, and Retirement," NBER Conference on Pensions, Labor, and Individual Choice, March 1983.

Hoadley, B.: "Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case," Annals of Mathematical Statistics, 42(1971), 1977-1991.

Huber, P. J.: "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics, and Probability, 1(1967), 221-233.

Itô, K.: "On Stochastic Differential Equations," Memoirs of the American Mathematical Society, 4(1951), 51.

Karlin, S. and H. M. Taylor: A Second Course in Stochastic Processes. New York: Academic Press, 1981.

Krishnan, V.: Nonlinear Filtering and Smoothing: An Introduction to Martingales, Stochastic Integrals and Estimation. New York: John Wiley and Sons, 1984.

Le Breton, A.: "On Continuous and Discrete Sampling for Parameter Estimation in Diffusion Type Processes," in Stochastic Systems: Modeling, Identification and Optimization, I, ed. by R. J.-B. Wets. Amsterdam: North-Holland Publishing Company, 1976.

Liptser, R. S. and A. N. Shiryayev: Statistics of Random Processes II: Applications. New York: Springer-Verlag, 1978.

Lo, A.: "Statistical Tests of Contingent Claims Asset-Pricing Models: A New Methodology," Journal of Financial Economics, 17(1986).

Marsh, T. and E. Rosenfeld: "Stochastic Processes for Interest Rates and Equilibrium Bond Prices," Journal of Finance, 38(1983), 635-645.

Pardoux, E. and D. Talay: "Discretization and Simulation of Stochastic Differential Equations," Acta Applicandae Mathematicae, 3(1985), 23-47.

Park, C. and J. Beekman: "Stochastic Barriers for the Wiener Process," Journal of Applied Probability, 20(1983), 338-348.

__________ and S. Paranjape: "Probabilities of Wiener Paths Crossing Differentiable Curves," Pacific Journal of Mathematics, 53(1974), 579-583.

__________ and F. Schuurmann: "Evaluations of Barrier-Crossing Probabilities and Wiener Paths," Journal of Applied Probability, 13(1976), 267-275.

Phillips, A. W.: "The Estimation of Parameters in Systems of Stochastic Differential Equations," Biometrika, 46(1959), 67-76.

Phillips, P. C. B.: "The Structural Estimation of a Stochastic Differential Equation System," Econometrica, 40(1972), 1021-1041.

__________: "The Problem of Identification in Finite Parameter Continuous Time Processes," Journal of Econometrics, 1(1973), 351-362.

__________: "The Estimation of Some Continuous Time Models," Econometrica, 42(1974), 803-823.

__________: "Time Series Regression With Unit Roots," Cowles Foundation Discussion Paper No. 740, 1985.

Prakasa Rao, B. L. S.: "Maximum Likelihood Estimation for Markov Processes, Annals of the Institute of Statistics and Mathematics, 24(1972), 333-345.

Rao, N., Borwankar, J., and D. Ramakrishna: "Numerical Solution of Itô Integral Equations," SIAM Journal of Control, 12(1974), 123-139.

Rosenfeld, E.: Stochastic Processes of Common Stock Returns: An Empirical Examination, unpublished Ph.D. thesis, Sloan School of Management, M.I.T., February 1980.

Roussas, G. G.: "Asymptotic Inference in Markov Processes," Annals of Mathematical Statistics, 36(1965), pp. 978-992.

Rudin, W.: Functional Analysis. New York: McGraw-Hill Book Company, 1973.

Rumelin, W.: "Numerical Treatment of Stochastic Differential Equations," SIAM Journal of Numerical Analysis, 19(1982), 604-613.

Schuss, Z.: Theory and Applications of Stochastic Differential Equations. New York: John Wiley and Sons, 1980.

Shiller, R. and P. Perron, "Testing the Random Walk Hypothesis: Power Versus Frequency of Observation," N.B.E.R. Technical Working Paper No. 45, 1985.

Siegmund, D.: "Boundary Crossing Probabilities and Statistical Applications," Stanford University Department of Statistics Technical Report No. E&S/NSF230, June 1985.

Sims, C. A.: "Discrete Approximations to Continuous Time Distributed Lags in Econometrics," Econometrica, 39(1971), 545-563.

Skorokhod, A. V.: Studies in the Theory of Random Processes. New York: Dover Publications, 1965.