BANCA D'ITALIA

Temi di discussione

del Servizio Studi

On the estimation of stochastic differential equations: the continuous-time maximum-likelihood approach

by Riccardo Cesari

Numero 125 - Settembre 1989

BANCA D'ITALIA

Temi di discussione

del Servizio Studi

On the estimation of stochastic differential equations: the continuous-time maximum-likelihood approach

by Riccardo Cesari

Numero 125 - Settembre 1989

The purpose of the «Temi di discussione» series is to promote the circulation of working papers prepared within the Bank of Italy or presented in Bank seminars by outside economists with the aim of stimulating comments and suggestions.

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.

ABSTRACT

Thanks to modern information technology, economic and in particular financial phenomena tend to be systematically observed at increased time frequency (weekly, daily, hourly etc.). This paper investigates the limit case of a continuous record of observations, presenting a general maximum-likelihood method for estimating the parameters of stochastic differential equations (SDEs) of observable point variables. This method, whose foundations can be found in the probabilistic literature, has the advantage of avoiding the main difficulties met by the discrete-time approach in the case of nonlinear and/or multivariate SDEs. In this paper, its properties are presented and some examples, covering a large part of the recent financial literature, are worked out. The case of discrete observations is considered, suggesting the importance of Monte Carlo experiments to evaluate the alternative approaches particularly in presence of nonstationary time series.

CONTENTS

§ 1 INTRODUCTION (*)

Stochastic differential equations (SDEs) are currently used in many areas of economics, from macro analysis to micro theory.[1]

This concept, which is now part of a new and rapidly growing branch of the theory of probability called 'stochastic analysis', found its first analytical treatment in a problem of financial economics, when, at the beginning of this century, a model of the brownian motion was suggested and an application to option prices was given by L.Bachelier.[2] More recently, in the same area of financial economics, empirical applications of SDEs have been greatly stimulated by the seminal work of Black and Scholes (1973) who, assuming a simple SDE for the stock price, were able to find a valuation formula for stock options completely free from preference assumptions, dependent only on observable variables and their dynamics.

The recent increase in the opportunities of applications has augmented the importance to provide general estimation method for SDEs.

The literature concerning this inferential problem can be easily segmented into two different approaches. The 'discrete approach' is typically econometric and realistically assume discrete sampling: for simple linear-in-the-variables (v-linear) models this approach transforms the SDE into an equivalent

---

* This paper is a revised version of a chapter of my D.Phil thesis (Cesari(1987)) submitted to the University of Oxford. I would like to thank my examiners, J.Mirrlees and S.Schaefer, and a referee for their helpful comments. Financial support from the Bank of Italy is also gratefully acknowledged. Any responsibility is of course only my own.

stochastic difference equation for which a number of estimation and testing results are known.[3] When v-nonlinearities are involved, the transformation is, in general, no longer possible and linear approximations are called for.

The 'continuous approach', instead, is typical of the probabilistic literature and assumes, at the outset, the availability of a continuous record of observations over the sample interval. From this information, a maximum-likelihood (ML) procedure is developed by exploiting a number of powerful results of stochastic calculus.[4] Closed form solutions of the ML equations are easily obtained in the linear-in-the-parameters (p-linear) case. In particular, v-nonlinearities presents no additional difficulty and multidimensionality does not raise any aliasing problem (see Phillips (1973) and Hansen and Sargent (1983)) which is instead a result of the discretization of the SDE.

Clearly, the continuous approach concerns the abstract, limit hypothesis of the availability of a continuous record of observations. As a matter of fact, the recent technological improvement in data collection has allowed to systematically observe a number of economic (chiefly, financial) phenomena at increasing frequency, from monthly to weekly, daily, hourly, minute by minute, making the 'continuous record' assumption not irrealistic in many cases.

An additional restrictive condition is the complete observability of the model variables. If, for example, flow data are involved (a case of point unobservability: consumption, income etc., observable only on a time interval) the continuous approach is significantly complicated becoming a joint problem of parameter estimation and filtering. The discrete approach is, instead, the natural set up for models of this type.

Moreover, it must be noted that the typical situation in empirical applications is the availability of a set of discrete observations, so that the practical implementation of the continuous approach requires to approximate the (stochastic) integrals involved in the ML estimators by (stochastic) sums which, however, by the separability property of diffusion processes, converge in probability to the corresponding integrals over the set of refined partitions of the sample interval.

Notwithstanding these limitations, the continuous approach has a wide range of empirical applications and, as it will be apparent from the examples provided in the text, a number of popular theoretical models, particularly in finance, could be easily estimated by it.

Unfortunaletly, no comparative analysis concerning the two approaches is available in the literature: in particular, either approach does not have optimal properties in finite samples and it would be very useful to obtain simulation results in different conditions (linear/nonlinear, uni/multivariate models, high/low observation frequency etc.).

As a first step in this direction, in this paper we shall analyse the continuous approach in the case of observable point variables, presenting assumptions and results of the continuous-time ML method which, to our best knowledge, has been largely neglected by financial economists and econometricians. The main contributions in the literature are reviewed and presented in a general and unified framework. Marginal references to alternative approaches are provided, leaving to future work a proper comparative analysis through Monte Carlo experiments.

The paper is as follows: in § 2 definitions and assumptions are presented; the estimation of the diffusion matrix is considered in § 3; § 4 shows the martingale properties of the likelihood ratio and the score vector; § 5 is concerned with the optimal properties of the continuous-time ML estimator of the drift; § 6 presents some simple examples; § 7 is devoted to the problem of discrete observations and § 8 concludes.

§ 2 DEFINITIONS AND ASSUMPTIONS

Let $(\Omega, F, (F_t), P)$ $t \in [0,T]$ be a standard set-up[5] and let

$$dX^i_t = a_i(X_t)dt + g_t'(X_t)dW_t \qquad X_{io}=x_{io}, \quad i=1(1)m \tag{1}$$

be an m-dimensional stochastic differential equation (SDE) system s.t. $W_t'=(W^1_t,...,W^n_t)$ is an n-dimensional standard Wiener process[6], $n \geq m$, and $X_t=(X^1_t,...,X^m_t)'$ is its unique strong solution (a diffusion process). In matrix form we have[7]

$$dX_t = A(X_t)dt + G(X_t)dW_t \tag{2a}$$

$X_0 = x_0$ (2b)

We explicitly note that the solution X_t shares many properties with the Wiener process: it is Markov and continuous in probability, it has P-a.s. uniformly continuous sample paths and it is nowhere differentiable and therefore of unbounded variation.

In particular, X_t is not mean-square differentiable so that the generalized Langevin equation

$$dX_t/dt \equiv DX_t = A(X_t) + G(X_t)\epsilon_t \tag{3}$$

interpreted as an ordinary differential equation driven by a continuous-time zero-mean Gaussian white noise, ϵ_t, has no direct correspondence with the Itô equation (1) but with (1) as a Stratonovich equation.[8]

In (1) we assume that the coefficients a_i, g_i', $i=1(1)m$ are smooth functions of two unrelated sets of parameters [9]

$$a_i(X_t,\theta) \quad , \quad g_i'(X_t,\sigma) \qquad \theta \cap \sigma = \emptyset, \qquad \theta\in\Phi, \ \sigma\in H \tag{4}$$

where Φ and H are open subsets of R^k, R^h respectively, and that the parameters are identifiable in the sense that will be defined below. Moreover we assume that X_t is a process with observable components.[10]

The econometric problem in the 'continuous approach' is to find estimators $\theta\hat{}_T$, $\sigma\hat{}_T$ having optimal properties on the basis of the continuous record (X_t: $t\in[0,T]$).
Many different estimation methods can be found for this problem in the literature but the maximum-likelihood (ML) procedure is certainly the most important one on both theoretical and

practical grounds.

We shal devote therefore our inquiry to the ML method giving marginal reference to alternative approaches.[11]

§ 3 ALMOST SURE ESTIMATION OF THE DIFFUSION MATRIX

We shall show the following

PROPOSITION 1: Given the system (1) and the continuous record $\{X_t \; t \in [0,T]\}$, the diffusion matrix GG' is a.s. known on $[0,T]$.

PROOF: The SDE system (1) in matrix and integral form is

$$X_t = X_0 + \int_0^t A(X_s,\theta)ds + \int_0^t G(X_s,\sigma)dW_s \qquad (5)$$

By hypothesis, X_t is a sample continuous semimartingale[12] and

$$M_t \quad X_0 + \int_0^t G(X_s,\sigma)dW_s \qquad (6)$$

is a sample continuous second order martingale. By Meyer (1962) decomposition theorem there exists a unique sample continuous increasing process of bounded variation (BV) null at zero, called the quadratic variation-covariation process[13] of M_t, $\langle M,M\rangle_t$ $[\langle M^i,M^j\rangle_t]$, and a sample continuous martingale N_t s.t.

$$M_t M_t' \quad N_t + \langle M,M\rangle_t \qquad (7)$$

and

$$\langle X,X\rangle_t \quad \langle M,M\rangle_t \qquad (8)$$

Using the Kunita-Watanabe (1967) characterization, the stochastic integral in (6) is a martingale uniquely determined by the following covariation property

$$\left\langle \int_0^t G(X_s;\sigma)dW_s,\ R_t \right\rangle = \int_0^t G(X_s;\sigma)d\langle W,R\rangle_s \quad \text{f.e. martingale } R_t \qquad (9)$$

This implies

$$\langle \int_0 G(X_s,\sigma)dW_s, W\rangle_t = \int_0^t G(X_s,\sigma)d\langle W,W\rangle_s = \int_0^t G(X_s,\sigma)I_n ds$$
$$= \int_0^t G(X_s,\sigma)ds \qquad (10)$$

$$\langle W, \int_0 G(X_s,\sigma)dW_s\rangle_t = \int_0^t G'(X_s,\sigma)ds \qquad (11)$$

$$\langle M,M\rangle_t = \langle \int_0 G(X_s,\sigma)dW_s, \int_0 G(X_s,\sigma)dW_s\rangle_t = \int_0^t G(X_s,\sigma)d\langle W, \int_0 G(X_u,\sigma)dW\rangle_s \qquad (12)$$

and using (8) and (12)

$$\langle X,X\rangle_t = \langle M,M\rangle_t = \int_0^t G(X_s,\sigma)G'(X_s,\sigma)ds \qquad (13)$$

From the constructive characterization of quadratic variations [14]

$$\langle X,X\rangle_T = \int_0^T dX_t dX_t' = \int_0^T G(X_t,\sigma)G'(X_t,\sigma)dt \qquad P\text{-a.s.} \qquad (14)$$

so that the covariance diffusion matrix $D(X_t,\sigma)=G(X_t,\sigma)G'(X_t,\sigma)$ can be assumed a.s. known. If it is nonsingular for every X_t, then $G(X_t,\sigma)$ is its unique positive definite square root.[15]

The parameter vector σ is defined to be identifiable if, from (14), it can be put in explicit form

$$\hat{\sigma}_T = f(X_t,\ t\in[0,T]) \qquad (15)$$

§ 4 MARTINGALE PROPERTIES OF THE LIKELOHOOD RATIO AND THE SCORE VECTOR

Let us consider the systems

$$dX_t = A(X_t,\theta)dt + G(X_t)dW_t \qquad \theta \in \Phi\backslash\theta_0 \qquad (16)$$

$$dY_t = A(Y_t,\theta_0)dt + G(Y_t)dW_t \qquad \theta_0 \in \Phi,\ \text{true value} \qquad (17)$$

$$X_0 = Y_0 \qquad (18)$$

having unique strong solutions X_t, Y_t which induce on the measurable space of continuous functions, (C^n_T, B^n_T), the probability measure P_θ, P_o respectively.[16]

We shall show the following

PROPOSITION 2: If

$\Gamma(Y_s,\theta) \equiv G'(Y_s)(G(Y_s)G'(Y_s))^{-1}(A(Y_s,\theta)-A(Y_s,\theta_o))$

is nonanticipative of class L^2 [17] satisfying the boundedness condition $E\exp(\mu|\Gamma(t)|^2) < \infty$, $\mu>0$, f.e. $t\in[0,T]$ and $\theta\in\Theta$ and $A'(Y_s,\theta)(G(Y_s)G'(Y_s))^{-1}A(Y_s,\theta)$, $\theta\in\Theta$ is of class L^1 then there exist a likelihood process $L_t(Y,\theta)$ on $[0,T]$ and it is a martingale.

PROOF: The hypothesis of Liptser and Shiryayev (1974 ch.7 p.279) are satisfied so that P_θ and P_o are equivalent (they have the same zero-measure sets) and in particular the Radon-Nikodym derivative is given by

$$\frac{dP_\theta}{dP_o}(Y,T) = L_T(Y,\theta) \tag{19}$$

where

$$L_T(Y,\theta) \equiv \exp\Big[\int_0^T (A(Y_s,\theta)-A(Y_s,\theta_o))'(G(Y_s)G'(Y_s))^{-1}dY_s - \tag{20}$$
$$-\tfrac{1}{2}\int_0^T (A(Y_s,\theta)-A(Y_s,\theta_o))'(G(Y_s)G'(Y_s))^{-1}(A(Y_s,\theta)+A(Y_s,\theta_o))ds\Big]$$

Equivalently, the likelihood ratio $L_T(\theta)$ can be written as

$$L_T(Y,\theta) = \exp\Big[\int_0^T \Gamma'(Y_s,\theta)dW_s - \tfrac{1}{2}\int_0^T |\Gamma(Y_s,\theta)|^2 ds\Big] \tag{21}$$

where $|.|$ is the Euclidean norm.

Clearly,

$$E_o(L_T(Y,\theta)) = 1, \tag{22}$$

the supermartingale $L_T(Y,\theta)$ is a P_o-martingale[18] and Girsanov (1960) theorem applies, so that

$$\tilde{W}_t = W_t - \int_0^t \Gamma(Y_s,\theta)ds \tag{23}$$

is a BM w.r.t. P_θ f.e. $\theta\in\Theta$.

Take the restriction of $L_T(Y,\theta)$ on $B_t \in B_T$

$$L_t(Y,\theta) = E_o(L_T(Y,\theta)|B_t) \tag{24}$$

Clearly it is a P_o-martingale as well and, using (21) we have

$$L_t(Y,T) = \exp[\int_0^t \Gamma'(Y_s,\theta)dW_s - \tfrac{1}{2}\int_0^t |\Gamma(Y_s,\theta)|^2 ds] \tag{25}$$

$$= \exp[\int_0^t \Gamma'(Y_s,\theta)d\tilde{W}_s + \tfrac{1}{2}\int_0^t |\Gamma(Y_s,\theta)|^2 ds] \tag{26}$$

By assumption on Γ,

$$\tilde{M}_t \equiv \int_0^t \Gamma'(Y_s,\theta)d\tilde{W}_s \tag{27}$$

is a P_θ-martingale and we can write

$$L_t(Y,\theta) = \exp(\tilde{M}_t + \tfrac{1}{2}\langle\tilde{M},\tilde{M}\rangle_t) = 1 - \int_0^t L_s(Y_s,\theta)d\tilde{M}_s \tag{28}$$

so that $L_t(Y,\theta)$ is also a P_θ-martingale f.e. $\theta\in\Theta$, with unit mean.[19]

We observe that, by Jensen inequality, the loglikelihood $\log L_t(Y,\theta)$ is a supermartingale.

PROPOSITION 3: If the order of differentiation and intergration can be interchanged[20], it can be shown that the score vector is a zero-mean martingale null at zero.

PROOF: In fact, symplifying notation,

$$q_t(Y,\theta) \equiv \frac{\delta}{\delta\theta}\log L_t(Y,\theta) = \int_0^t \frac{\delta}{\delta\theta}[(A(\theta)-A(\theta_o))'](GG')^{-1}dY_s -$$

$$-\tfrac{1}{2}\int_0^t \frac{\delta}{\delta\theta}[(A(\theta)-A(\theta_o))'(GG')^{-1}(A(\theta)+A(\theta_o))]ds$$

$$= \int_0^t \frac{\delta}{\delta\theta} A'(\theta)(GG')^{-1}dY_s - \int_0^t \frac{\delta}{\delta\theta} A'(\theta)(GG')^{-1}A(\theta)ds \qquad (29)$$

given that

$$\frac{\delta}{\delta\theta} [A'(\theta)(GG')^{-1}A(\theta)] = \frac{\delta}{\delta\theta} A'(\theta) \frac{\delta}{\delta\theta} [A'(\theta)(GG')^{-1}A(\theta)]$$

$$= 2 \frac{\delta}{\delta\theta} A'(\theta)(GG')^{-1}A(\theta)$$

Using (17) and (24) we can write

$$q_t(Y,\theta) = \int_0^t \frac{\delta}{\delta\theta} A'(\theta)(GG')^{-1}GdW_t - \int_0^t \frac{\delta}{\delta\theta} A'(\theta)(GG')^{-1}(A(\theta)-A(\theta_o))ds$$

$$= \int_0^t \frac{\delta}{\delta\theta} A'(\theta)(GG')^{-1}GdW^{\sim}_t(\theta) \qquad (30)$$

which is a zero-mean square-integrable P_θ-martingale.

A maximum likelihood (ML) estimator $\theta\hat{}_t$ is an adapted process satisfying

$$L_t(Y,\theta\hat{}_t) = \sup_\theta (L_t(Y,\theta),\ \theta\epsilon\Theta) \qquad (31)$$

so that, L_t being exponential, $\theta\hat{}_t$ solves the likelihooh equation

$$q_t(Y,\theta\hat{}_t) = 0 \qquad (32)$$

Note that in this case of Wiener disturbances we obtain, in continuous time, the well known result that the MLE coincides with the minumum-distance (weighted least-square) estimator

$$\min_\theta \int_0^t (dY_s - A(\theta)ds)'(GG')^{-1}(dY_s - A(\theta)ds)$$

PROPOSITION 4: In the hypotheses of Proposition 3 the Fisher information matrix is given by [21]

$$I_t(\theta) = \int_0^t \frac{\delta}{\delta\theta} A'(\theta)(GG')^{-1} \frac{\delta}{\delta\theta} A(\theta)\ ds$$

PROOF: From the Itô integration by parts formula and Meyer (1962) theorem we obtain

$$q_t(\theta)q_t'(\theta) = \int_0^t dq_s(\theta)\, q_s'(\theta) + \int_0^t q_s(\theta)\, dq_s'(\theta) + \langle q(\theta),q(0)\rangle_t \qquad (33)$$

Writing I_t for the quadratic covariation process of q_t

$$I_t(\theta) \quad \langle q(\theta),q(\theta)\rangle_t \qquad (34)$$

it can be considered a random Fisher conditional information matrix for dependent observations. In fact, if the order of differentiation and integration can be changed we have (using E for E_θ)

$$E[I_t(\theta)] \quad E[q_t(\theta)q_t'(\theta)] \equiv E[\frac{\delta \log L_t}{\delta\theta}\frac{\delta \log L_t}{\delta\theta'}] \quad -E[\frac{\delta^2 \log L_t}{\delta\theta\delta\theta'}]$$

given that

$$\frac{\delta^2 \log L_t}{\delta\theta\delta\theta'} = \frac{\delta}{\delta\theta}(\frac{\delta \log L_t}{\delta\theta'}) = \frac{\delta}{\delta\theta}(\frac{\delta L_t}{\delta\theta'}\frac{1}{L_t}) = \frac{\delta^2 L_t}{\delta\theta\delta\theta'}\frac{1}{L_t} - \frac{\delta \log L_t}{\delta\theta}\frac{\delta \log L_t}{\delta\theta'}$$

and, using (23)

$$E[\frac{\delta^2 L_t}{\delta\theta\delta\theta'}\frac{1}{L_t}] = E_0[\frac{\delta^2 L_t}{\delta\theta\delta\theta'}] = \frac{\delta^2}{\delta\theta\delta\theta'} E_0[L_t] = 0 \qquad (35)$$

The following property holds

$$E(I_t(\theta)-I_s(\theta)|B_s) = E(q_t(\theta)q_t'(\theta)-q_s(\theta)q_s'(\theta)|B_s) \qquad (36)$$

$$= E[(q_t(\theta)-q_s(\theta))(q_t(\theta)-q_s(\theta))'|B_s]$$

so that

$$E(dI_t(\theta)|B_t) = E(dq_t(\theta)dq_t'(\theta)|B_t) = dI_t(\theta) \qquad (37)$$

the last equality being a result of the predictability property of the quadratic variation process[22].

Consequently the Fisher information matrix is

$$I_t(\theta) \quad \int_0^t dq_s(\theta)dq_s'(\theta) = \int_0^t \frac{\delta}{\delta\theta} A'(\theta)(GG')^{-1} \frac{\delta}{\delta\theta} A(\theta)\, ds \qquad (38)$$

Equations (29) and (32) are the basic results to calculate the ML estimator $\theta\hat{}_t$; equations (30) and (38) allow to investigate its distributional properties. The identifiability of θ means that it can be made explicit from (32).

§ 5 PROPERTIES OF THE ML ESTIMATOR

We shall consider the asymptotic properties of the ML estimator in some special cases.

P-LINEAR V-NONLINEAR DRIFT. Let us assume that the drift vector $A(X,\theta)$ is linear in the parameters (p-linear) but quite general w.r.t. the variables.

We have

$$dX_t = (a(X_t) + B(X_t)\theta)dt + G(X_t)dW_t \qquad (39a)$$

$$X_0 = x_0 \qquad (39b)$$

and, omitting arguments,

$$q_t(Y,\theta) = \int_0^t B'(GG')^{-1}dY_s - \int_0^t B'(GG')^{-1}(a+B\theta)ds \qquad (40)$$

$$I_t(Y,\theta) = \int_0^t B'(GG')^{-1}Bds = I_t(Y) \qquad (41)$$

giving (for I_t non singular)

$$\theta\hat{}_t = I_t^{-1} \left[\int_0^t B'(GG')^{-1}dY_s - \int_0^t B'(GG')^{-1}ads\right] \qquad (42)$$

so that

$$\theta\hat{}_t = \theta_0 + I_t^{-1} \int_0^t B'(GG')^{-1}GdW_s \qquad (43)$$

and using (40) q_t has the representation

$$q_t(Y,\theta) = I_t(Y,\theta)(\theta\hat{}_t-\theta) \qquad (44)$$

The following theorem gives sufficient conditions for asymptotic normality.

PROPOSITION 5: If as $t \to \infty$

$$\frac{1}{t}\int_0^t B'(X_t)(G(X_t)G'(X_t))^{-1}B(X_t)ds \xrightarrow{P_\theta} C(\theta) \quad (45a)$$

$$\frac{1}{t}\int_0^t E\,[B'(X_t)(G(X_t)G'(X_t))^{-1}B(X_t)]ds \to C(\theta) \quad (45b)$$

where $C(\theta)$ is a nonrandom nonsingular matrix, then a.s. the likelihood equation has a unique solution which is consistent and normally distributed in the sense that

$$\sqrt{t}\,(\hat{\theta}_t-\theta_o) \xrightarrow{D} N(0,C^{-1}(\theta_o)) \quad (46)$$

If a,B,G are time-independent (homogeneous diffusion) and the process is recurrent with a (steady state) ergodic distribution $\bar{P}_\theta$ then, if B,G have bounded norms,

$$C(\theta) \equiv \bar{E}_\theta[B'(X)(G(X)G'(X))^{-1}B(X)] \quad (47)$$

where $\bar{E}_\theta$ is the expectation operator w.r.t. the steady state distribution.[23]

PROOF: See Taraskin (1970) p.218.

<u>P- V- NONLINEAR DRIFT</u> If the drift vector is nonlinear in the parameters as well as in the variables (p- v- nonlinear) then Taraskin's (1970) theorem can be extended provided that the process is ergodic f.e. $\theta\epsilon\Theta$ and the function

$$\frac{\delta}{\delta\theta} A'(X_o,\theta)(G(X_o)G'(X_o))^{-1}G(X_o)$$

satisfies a boundedness condition (see Kutoyants (1978) p.401). Under these assumptions $\hat{\theta}_t$ is consistent and asymptotically normal:

$$\sqrt{t}\,(\hat{\theta}_t-\theta_o) \xrightarrow{D} N(0,D^{-1}(\theta_o)) \quad (48)$$

$$D(\theta) \equiv \bar{E}_\theta\,[\frac{\delta}{\delta\theta} A'(X_o,\theta)\,(G(X_o)G'(X_o))^{-1}\,\frac{\delta}{\delta\theta'} A(X_o,\theta)\,] \quad (49)$$

<u>ASYMPTOTIC EFFICIENCY</u> The strong law of large numbers (SLLN) for brownian motions, saying that

$$\lim_{t\uparrow\infty} \frac{W_t}{t} = 0 \text{ a.s.} \quad (50)$$

can be extended to any square-integrable sample-continuous martingale M_t s.t.

$$\lim_{t\uparrow\infty} \langle M,M\rangle_t = \infty$$

in which case it is stated as

$$\lim_{t\uparrow\infty} \langle M,M\rangle_t^{-1} M_t = 0 \text{ a.s.} \quad (51)$$

As a result we have

$$\lim_{t\uparrow\infty} I_t^{-1}(\theta)q_t(\theta) = 0 \quad P_\theta\text{-a.s.} \quad (52)$$

provided that

$$\lim_{t\uparrow\infty} I_t(\theta) = \infty \quad (53)$$

In the univariate case, using a random-time substitution it is possible to write the martingale q_t as a BM with a random time change

$$q_t(\theta) = W_{I_t(\theta)} \quad (54)$$

so that a central limit theorem (CLT) holds[24]

$$I_t^{-1/2}(\theta)W_{I_t(\theta)} \xrightarrow{D} N(0,1)$$

If the score can be factorized as

$$q_t(\theta) = I_t(\theta)(\hat{\theta}_t-\theta) \quad (55)$$

then, using SLLN and CLT, we obtain strong consistency and asymptotic normality[25] in the sense that

$$I_t^{½}(\theta)(\theta\hat{}_t-\theta) \overset{D}{\rightarrow} N(0,1) \qquad (56)$$

Following Feigin(1976,1978) and Heyde and Feigin (1975) define the estimator $\theta\tilde{}_t$ asymptotically efficient if as $t \rightarrow \infty$

$$[I_t^{½}(\theta)(\theta\tilde{}_t-\theta) - Q(\theta)I_t^{-½}(\theta)q_t(\theta)] \overset{P_\theta}{\rightarrow} 0 \qquad (57)$$

for some constant nonsingular matrix $Q(\theta)$.

Using the SLLN and CLT you can see that $\theta\tilde{}_t$ is consistent for θ and asymptotically normal in the sense that, as $t \rightarrow \infty$

$$I_t^{½}(\theta)(\theta\tilde{}_t-\theta) \rightarrow N(0,\Sigma(\theta)) \qquad (58)$$

where $\Sigma^{½}(\theta) = Q(\theta)$

If the factorization (55) holds, then the MLE is asymptotically efficient and therefore consistent and asymptotically normal with $Q(\theta)=I_k$.

§ 6 SOME EXAMPLES[24]

Example 1: Ornstein-Uhlenbeck (O-U) process

$$dX_t = \theta X_t dt + \sigma dW_t \qquad \theta \in R\backslash\theta_0 \qquad (59a)$$

$$dY_t = \theta_0 Y_t dt + \sigma dW_t \qquad (59b)$$

$$X_0 = Y_0 \qquad (59c)$$

The diffusion coefficient is obtained as

$$\sigma^2 = \frac{1}{t}\int_0^t (dX_t)^2 = \frac{1}{t}\langle X,X\rangle_t$$

where, as above, $(dX_t)^2 = d\langle X,X\rangle_t$. From (29)

$$q_t(\theta) = \frac{1}{\sigma^2}\left[\int_0^t Y_s dY_s - \int_0^t \theta Y_s^2 ds\right] = \frac{1}{\sigma}\int_0^t Y_s d\tilde{W}_t$$

$$\theta\hat{}_t = \left(\int_0^t Y_s^2 ds\right)^{-1}\left(\int_0^t Y_s dY_s\right) = \left(\int_0^t Y_s^2 ds\right)^{-1} ½\left[Y_T^2 - Y_0^2 - \int_0^t (dY_s)^2\right]$$

$$I_t(\theta) = \frac{1}{\sigma^2}\int_0^t Y_s^2 ds = I_t$$

so that

$$q_t(\theta) = I_t(\theta\hat{}_t - \theta)$$

It can be shown[27] that I_t is an integrable process increasing to ∞ so that $\theta\hat{}_t$ is strongly consistent f.e. θ. Moreover, for θ_o n.e. 0,

$$I_t^{½} (\theta\hat{}_t - \theta_o) \xrightarrow{D} N(0,1)$$

For $\theta_o = 0$, however, asymptotic normality does not hold (in particular, the sufficient conditions (45) in Proposition 5 are non satisfied). In this case, in fact, $Y_t - Y_o = \sigma W_t$ and, assuming for simplicity $Y_o = 0$, we have

$$I_t^{½}\theta\hat{}_t = \left(\int_0^t W_s dW_s\right)\left(\int_0^t W_s^2 ds\right)^{-½} = ½(W_t^2 - t)\left(\int_0^t W_s^2 ds\right)^{-½}$$

implying that the asymptotic distribution is non normal, being

$$\lim_{t\uparrow\infty} P(I_t^{½}\theta\hat{}_t \le 0) = \lim_{t\uparrow\infty} P\left(\frac{W_t^2}{t} \le 1\right) = X^2_1(1) > ½$$

where X^2_1 is a chi-square distribution function with 1 d.f. Therefore $\theta_o = 0$ represents a point of singularity in the asymptotic distribution results.[28]

Moreover, optimal properties are not valid in finite sample intervals: as shown by Liptser and Shiryayev (1974 theorem 17.2), given the SDE

$$dX_t = \theta B(X_t, t)dt + dW_t \qquad \theta \in R$$

$$dY_t = \theta_o B(X_t, t)dt + dW_t$$

$$X_o = Y_o = 0$$

the MLE is biased for any finite record length [0,t], being

$$\hat{\theta}_t = \left(\int_0^t B^2(Y_s,s)ds\right)^{-1} \int_0^t B(Y_s,s)dY_s$$

and if B satisfies regularity conditions (e.g. if it is bounded) then[29]

$$E_0(\hat{\theta}_t) = \theta_0 + \frac{\delta}{\delta\theta_0} E_0\left(\int_0^t B^2(Y_s,s)ds\right)^{-1}$$

<u>Example 2: Feller diffusion (diffusion branching process)</u>

$$dX_t = \theta X_t dt + \sigma\sqrt{X_t}dW_t \qquad \theta<0,\ X_t\geq 0 \tag{60a}$$

$$dY_t = \theta_0 Y_t dt + \sigma\sqrt{Y_t}dW_t \tag{60b}$$

$$X_0 = Y_0 > 0 \tag{60c}$$

We have

$$\sigma^2 = \left(\int_0^t Y_s ds\right)^{-1}\left(\int_0^t (dY_s)^2\right)$$

$$q_t(\theta) = \frac{1}{\sigma^2}\int_0^t dY_s - \theta\int_0^t Y_s ds = \frac{1}{\sigma}\int_0^t \sqrt{Y_s}d\tilde{W}_s$$

$$\hat{\theta}_t = \left(\int_0^t Y_s ds\right)^{-1}(Y_t - Y_0)$$

$$I_t(\theta) = \frac{1}{\sigma^2}\int_0^t Y_s ds = I_t$$

$$q_t(\theta) = I_t(\hat{\theta}_t - \theta)$$

It can be shown (Feigin (1976)) that $\hat{\theta}_t$ is strongly consistent and asymptotically normal. Note that the change of variable $R_t=\sqrt{Y_t}$ would give

$$dR_t = \left(\frac{\theta_0}{2} R_t - \frac{\sigma^2}{2R_t}\right)dt + \frac{\sigma}{2} dW_t$$

so that, given

$$\sigma^2 = \frac{4}{t}\int_0^t (dR_s)^2$$

we would obtain the same ML estimator of θ

$$\theta\hat{}_t = \left(\int_0^t R_s^2 ds\right)^{-1}(R_t^2 - R_0^2)$$

In other words, the MLE is invariant under regular (C^2) transformations of the state space.[30]

Example 3: Multivariate diffusion

$$dX_{1t} = \theta_1 X_{1t} dt + g_1' \sqrt{X_{2t}} dW_t \tag{61a}$$

$$dX_{2t} = \theta_2 X_{2t} dt + g_2' \sqrt{X_{2t}} dW_t \tag{61b}$$

$$(X_{1o}, X_{2o}) = (x_{1o}, x_{2o}) \tag{61c}$$

We outline the steps to obtain the ML estimator. Write in vector form $\theta' = (\theta_1, \theta_2)$, $X_t' = (X_{1t}, X_{2t})$ and $G = (g_1, g_2)'$ a 2xn matrix s.t.

$$GG' \equiv \Sigma = \begin{bmatrix} s_1^2 & \alpha s_1 s_2 \\ \alpha s_1 s_2 & s_2^2 \end{bmatrix} = \left(\int_0^t X_{2s} ds\right)^{-1}\left(\int_0^t dX_s dX_s'\right)$$

where $s_1^2 = g_1'g_1$, $s_2^2 = g_2'g_2$, $\alpha s_1 s_2 = g_1'g_2 = g_2'g_1$.

$$q_t(\theta) = \int_0^t \begin{bmatrix} X_{1s} & 0 \\ 0 & X_{2s} \end{bmatrix} \frac{\Sigma^{-1}}{X_{2s}} dX_s - \int_0^t \begin{bmatrix} X_{1s} & 0 \\ 0 & X_{2s} \end{bmatrix} \frac{\Sigma^{-1}}{X_{2s}} \begin{bmatrix} \theta_1 & X_{1s} \\ \theta_2 & X_{2s} \end{bmatrix} ds$$

Calculate Σ^{-1} and solve the likelihood equation for $\theta\hat{}_t$.

The Fisher conditional matrix is

$$I_t = \int_0^t \begin{bmatrix} X_{1s} & 0 \\ 0 & X_{2s} \end{bmatrix} \frac{\Sigma^{-1}}{X_{2s}} \begin{bmatrix} X_{1s} & 0 \\ 0 & X_{2s} \end{bmatrix} ds$$

and therefore

$$q_t(\theta) = I_t(\theta\hat{}_t - \theta).$$

§ 7 DISCRETE OBSERVATIONS

In many instances, whenever a continuous monitoring is not possible, a discrete sample of observations, usually taken at equally spaced time points, is available. In this case, if the

SDE is sufficiently simple such as a nonzero-mean homogeneous O-U SDE

$$dX_t = (\theta X_t + b)dt + \sigma dW_t \tag{62}$$

corresponding to the integral equation

$$X_t - X_s = \int_s^t (\theta X_u + b)du + \sigma \int_s^t dW_u \tag{63}$$

it is possible to obtain the closed form solution for the process given, in this example, by

$$X_t = X_0 e^{\theta t} + b(e^{\theta t} - 1)/\theta + \sigma \int_0^t e^{\theta(t-s)} dW_s \tag{64}$$

so that, as in Sargan (1976), writing X_τ for $X_{\tau\delta}$, $\tau = 0,1,\ldots,\infty$ δ being the fixed time interval between successsve observations, we obtain

$$X_\tau = e^{\theta\delta} X_{\tau-1} + b(e^{\theta\delta} - 1)/\theta + \sigma \int_{(\tau-1)\delta}^{\tau\delta} e^{\theta(\tau\delta - s)} dW_s \tag{65a}$$

or

$$X_\tau = c_0 + c_1 X_{\tau-1} + u_\tau \tag{65b}$$

where $c_0 = b(e^{\theta\delta} - 1)/\theta$, $c_1 = e^{\theta\delta}$ and u_τ is a zero mean Gaussian white noise with variance

$s^2 = \sigma^2(e^{2\theta\delta} - 1)/(2\theta)$.

From the OLS estimates of c_0, c_1 and s^2 (and the assumption $\delta \equiv 1$) the ML estimates for θ, b and σ^2 can be easily obtained.

This discrete approach, which involves no approximation, has, however, two major limitations: it needs an explicit solution of the SDE, i.e. a result quite difficult to obtain in nonlinear cases, and it involves identification problems in the linear but multivariate case (see Phillips (1973) and Hansen and Sargent (1983)). In general some sort of linearization must be introduced.

Alternatively, assuming for simplicity $b=0$, the continuous-time ML approach would give, as in Example 1,

$$\hat{\theta}_T = \left(\int_0^T X_s^2 ds\right)^{-1}\left(\int_0^T X_s dX_s\right) \qquad (66)$$

which, using the Cauchy approximation to integrals[31], become

$$\hat{\theta}_{n,T} = (\Sigma_1^{N_n} X_{\tau-1}^2\delta_n)^{-1}(\Sigma_1^{N_n} X_{\tau-1}\Delta X_\tau) \qquad N_n \equiv T/\delta_n \qquad (67)$$

i.e. the OLS estimator of

$$\Delta X_\tau = \theta X_{\tau-1}\delta_n + u_\tau \qquad \tau=1,\dots,N_n \qquad (68)$$

The basis for this approximation comes from the separability property of continuous processes (see Wong and Hajek (1985) p.45) by which operations involving a continuoum of realizations of the process can be consistently approximated using a discrete number of observations, with the convergence in probability defined in terms of infinitesimal asymptotics i.e. as the maximum distance, δ_n, between successive times of observation goes to zero.

More generally, following Prakasa-Rao and Rubin (1981) it can be shown that for a stationary ergodic process satisfying the SDE with nonlinear drift

$$dX_t = A(X_t,\theta)dt + dW_t \qquad (69a)$$

$$X_0 = x_0 \qquad (69b)$$

the least square criterion applied to

$$\Delta X_\tau = A(X_{\tau-1},\theta)\delta_n + u_\tau \qquad (70)$$

giving the LS estimator

$$\hat{\theta}_{n,T} = \min_\theta \Sigma_1^{N_n} (\Delta X_\tau - A(X_{\tau-1},\theta)\delta_n)^2 \qquad (71)$$

is equivalent to the ML criterion on the basis of the continuous record over [0,T] provided that the maximum distance δ_n between successive observations vanishes.

This problem has also been studied by Le Breton (1976) for the O-U process. He found, in particular, that the discrete-time Cauchy-analog, $\theta\hat{}_{n,T}$, of the ML estimator as well as the exact discrete-time MLE converge in probability to the continuous-time MLE $\theta\hat{}_T$ as $\delta_n \downarrow 0$. Moreover, they differ from $\theta\hat{}_T$ by an error which is of order $\sqrt{\delta_n}$ in probability.

The difficulty, however, is that the discrete-time analog (Cauchy sums instead of integrals) of the ML estimator converges (for $\delta_n \downarrow 0$) to an estimator which is only asymptotically (for $T\uparrow\infty$) optimal: as shown by Bergstrom (1984 theorem 4) for the stationary O-U case the asymptotic ($T\uparrow\infty$) bias of $\theta\hat{}_{n,T}$ is of order δ_n in probability[32].

In the nonstationary case $\theta_0=0$ Phillips (1987) has shown that, as $\delta_n \downarrow 0$,

$$\theta\hat{}_{n,T} \equiv (\alpha\hat{}_{n,T} - 1)/\delta_n \xrightarrow{D} \theta\hat{}_T$$

as expected, but also that the asymptotic ($\delta_n \downarrow 0$) distribution of $\alpha\hat{}_{n,T}$ is degenerate, the estimator being consistent for 1 as $\delta_n \downarrow 0$, so that in this nonstationary case $\alpha\hat{}_{n,T}$ has optimal properties not only in the infinite ($T\uparrow\infty$) but also in the infinitesimal ($\delta_n \downarrow 0$) convergence: a suitable discrete sampling scheme turns out to be preferable and more informative than a single continuous record over a finite time interval.

§ 8 CONCLUSIONS

We have presented a general ML method for estimating SDEs based on the assumption of the availability of a continuous record of observations, an assumption not irrealistic whenever the modern information technology is used in collecting the data. From basic results of stochastic analysis we have shown the almost sure estimability of the diffusion matrix, the martingale properties of the likelihood ratio and the score vector as well as the equivalence of a Fisher conditional information matrix to the quadratic variation process of the score. Conditions for consistency, asymptotic normality and the Heyde-Feigin (1975) asymptotic efficiency are given in the general case of nonlinear drifts both in the parameters and in the variables. The usual difficulties related to the discrete-time estimator of nonlinear in the variables and/or multivariate SDEs are overcome.

With discrete observations, however, the continuous-time MLE must be, at best, consistently approximated in the sense that its discrete-time approximation tends to the continuous-time MLE as the discrete observations become more and more dense in the sample interval.

In actual practice, both the continuous-time and the discrete-time estimation methods involve, in general, some degree of approximation and Monte Carlo studies, along with more theoretical analysis, are an important point in the agenda on the subject. In particular, the availability of a continuous-time estimation method as the one presented in this paper raises a number of questions: does it pay to transform a continuous model into a discrete one? At which stage of the analysis? In which setting linearizations and other approximations have weaker effects? Which are the relative properties, in small and large samples, of discrete and continuous estimates of a given parameter vector when the data have the typical nonstationary dynamics of economic time series? As suggested by Novikov (1972) and Shiryayev (1974) or by Feigin (1979) and Phillips (1987), alternative sampling schemes seem to be preferable in the case of (near-) nonstationary processes but general results have not yet been obtained.

FOOTNOTES

*. This paper is a revised version of a chapter of my D.Phil. thesis (Cesari(1987)) submitted to the University of Oxford. I would like to thank my examiners, J.Mirrlees and S.Schaefer, and a referee for their helpful comments. Financial support from the Bank of Italy is also gratefully acknowledged. Any responsibility is of course only my own.

1. See Malliaris and Brock (1982) for an extended survey. Arguments in favour of continuous-time economic modelling are given, for example, in Merton (1975) and Gandolfo (1981) ch.1.

2. Bachelier's pioneering work is reproduced in Cootner (ed.) (1964). Note that a careful distinction should be kept between the brownian motion, which is a 'matter of fact', result of experimental observations by R.Brown[1828] and others and the brownian motion process (Bachelier[1900]-Einstein[1905]-Wiener[1923]) process) which is one of possible models of the former and, in general, of erratic movements over space and time.

3. See Bergstrom (1984) for a recent survey of the discrete approach. Pioneering works by J.D.Sargan, C.R.Wymer and P.C.B.Phillips can be found in Bergstrom (ed.) (1976).

4. The foundations of the continuous-time ML procedure are in Liptser and Shiryayev (1974). A recent extended framework is given in Hutton and Nelson (1986). A survey is provided by Basawa and Prakasa-Rao (1980).

5. (F_t) $t\in[0,T]$ is a standard filtration if it is a complete, right-continuous increasing family of sub σ-algebra of F. See, for example, Wong and Hajek (1985) p.210.

6. A (one-dimensional) standard Wiener process (or standard brownian motion process, BM) is defined by the properties
1) $W_t \sim N(0,t)$; 2) $E(W_t W_s)=\min(t,s)$; 3) $W_0=0$ P-a.s.
In the multidimensional case, each component is a standard BM.

7. The drift vector $A(X_t)$ and the diffusion matrix $G(X_t)\equiv[g_i{}'(X_t)]$ are assumed to satisfy by components the conditions of bounded growth and Lipschitz continuity of Itô[1951] theorem which guarantee existence and pathwise uniqueness of the solution process. The coefficients of the SDE (1) do not depend directly on time so that X_t has stationary transition densities (time-homogeneous diffusion). This is not necessary for the estimation method but it is required for the

existence of a steady-state random process associated whith X_t and therefore for the asymptotic properties of the estimators.

8. See Jazwinsky (1970) ch.4, Wong and Hajek (1985) ch.4, McShane (1972), Sethi and Lehoczky (1981). Itô and Stratonovich stochastic integrals have different definitions and properties. Even if equivalence theorems exist in order to identify the same process with apparently different Itô and Stratonovich SDEs, a particular interpretation must be adopted when a SDE is used to model some random phenomenon. Different interpretations will give different solutions.

9. The fact that θ,σ are assumed to be disjoint (no parameters in common) in not restrictive. Using the result of the next paragraph, if θ,σ have σ_o in common (i.e. $\theta_o=\sigma_o$) and σ is identifiable then σ_o can be assumed a.s. known in the drift vector and θ substituted by $\theta\backslash\sigma_o\equiv\theta_1$.

10. In general, if X_t contains unobservable components we face a joint problem of estimation and filtering about which very little, if any, is known in the general case of nonlinear coefficients or linear but interactive systems. (See also the following note). On the problem of instantaneous flow variables (consumption, income, etc.), observable only in integral form over a time interval, see Phillips (1978).

11. Alternative estimation methods are presented in Lanska (1979) (minimum contrast estimation), Banon (1978) (nonparametric estimation of the drift using the steady-state estimated density and the Fokker-Plank equation), Aase (1982) whose recursive estimator (for linear-in-the-parameter drifts) generalizes the Kalman-Bucy method and encompasses the ML estimator.
In the case of random varying "parameters", i.e. in the case of multivariate diffusions with unobservable components, filtering methods can be found in Ershov (1970), Liptser and Shiryayev (1974), Kallianpur (1980), Jazwinsky (1970).

12. A process X_t is a semimartingale if it can be decomposed as $X_t = A_t + M_t$ where A_t is an F_t-measurable (i.e. adapted) process null at zero ($A_0=0$) with a.s. sample paths of finite (total) variation and M_t is an F_t-martingale. Such a decomposition, if it exists, must be pathwise unique. Note that diffusion processes are only a subset of the very general class of semimartingales (e.g. Shiryayev (1981)) which includes non-continuous as well as non-Markov processes. Estimation methods for semimartingales are developed in Hutton and Nelson (1986) and Christopeit (1986).

13. Given two onedimensional continuous martingale, M,N, the quadratic covariation process $\langle M,N\rangle_t$ is defined by
$\langle M,N\rangle_t = \frac{1}{4} \langle M+N,M+N\rangle_t - \frac{1}{4} \langle M-N,M-N\rangle_t$
and the quadratic variation process $\langle M,M\rangle_t$ is defined, by Meyer

(1962, 1963) theorem, as the unique continuous increasing process s.t. $M_t^2 - \langle M,M\rangle_t$ is a martingale.

14. Note that in eq.(14) the second integral is defined as the limit -over nested partitions of [0,T] refining to zero- of the cross products of the increments of the elements of X_t. By the properties of continuous martingales it can be shown that this integral is a constructive characterization of $\langle M,M\rangle_T$.

15. See Le Breton (1977) for the Ornstein-Uhlenbeck case and McKeague (1984) on the effects of misspecifying the matrix G.

16. Let C^n_T be the space of continuous functions $f(t)$: $t \rightarrow R^n$. Let B^n_T be the σ-algebra generated by the sets $\{t: f(t)\in A\}$ where $t\in[0,T]$ and A is any Borel set in R^n. Any sample continuous process X_t in (Ω,F,P) induce a probability, say P_x, on (C^n_T, B^n_T) through the definition
$P_x(B) = P(w\in\Omega;\ X.(w)\in B)$ f.e. $B\in B^n_T$
where $X.(w)$ is the sample path $t \rightarrow X_t(w)$.

17. A separable and measurable process X_t F_t-adapted is called nonanticipative (w.r.t. (F_t)). It is of class L^2 if

$$P\left(\int_0^\infty |X_t|^2 dt < \infty\right) = 1$$

The class L^1 is defined analogously.

18. See for example Liptser and Shiryayev (1974) ch.6 p.216 and Friedman (1975) p.156.

19. Analogous results can be obtained for $\frac{dP_o}{dP_\theta}(X,t) \equiv V_t(X,\theta)$ and for $L_t(X,\theta)$, $V_t(Y,\theta)$ which equal, from the equivalence of measures, $V_t^{-1}(X,\theta)$ and $L_t^{-1}(Y,\theta)$ respectively.

20. See Hutton and Nelson (1984) for sufficient conditions.

21. See Feigin (1976) for the univariate case.

22. Predictability essentially means that I_{t+dt} is B_t-measurable. The quadratic variation process is also called predictable quadratic variation or predictable compensator: see Wong and Hajek (1985 p.224).

23. See also Brown and Hewitt (1975a). Kulinich (1975) gives an example of non recurrent diffusion with random limit $C(\theta)$ such that the consistency of $\hat{\theta}_t$ is preserved but its asymptotic distribution (with a different normalization) is no longer Gaussian.

24. See Feigin (1976). Taraskin's (1973) theorem 1 and 5 deal with the multivariate case.

25. McKeague (1984) proved that if the diffusion is stationary ergodic, then strong consistency and asymptotic normality of the MLE continue to hold even if the diffusion coefficient function $G(X_t)$ has been misspecified (loss in efficiency). If also the drift function is misspecified, consistency and asymptotic normality are valid w.r.t. the pont of minimum θ^* of the MSE between the true drift $A(X_t,\theta_0)$ and the misspecified one.

26. The recent literature on asset valuation provides a wide range of theoretical models in continuous time the ML procedure could be applied to. For example, Black and Scholes (1973) used a lognormal model, Vasicek (1977) an O-U model, Cox, Ingersoll and Ross (1985) a branching process; Langetieg (1980), Merton (1973), Richard (1978) and others have suggested multivariate theoretical models. Cesari (1987) part 2 contains an application of the continuous-time approah to the estimation of a multivariate model of the real and nominal term structures of interest rates in Italy.

27. See Feigin (1976) and Brown and Hewitt (1975a).

28. See Feigin (1979) and Phillips (1987) theorem 6.3 for Y_0 n.e. 0. Note that for every $\theta \in R$ the O-U process does not explode in the sense that it cannot reach $\pm\infty$ at a finite time (Aase (1982)). For $\theta > 0$, however, it is nonstationary while for $\theta < 0$ it is stationary ergodic.

29. Shiryayev (1974) and Novikov (1972) suggests the sequential or stopped ML estimator: take $H>0$ and define the stopping rule

$$\tau(H) = \inf(t: \int_0^t B^2(X_s,s)ds \geq H)$$

If f.e. $\theta \in \Phi$,

$$P_\theta(\int_0^\infty B^2(X_s,s)ds = \infty) = 1$$

then $P_\theta(\tau(H)<\infty)=1$ and the sequential estimator (for B bounded)

$$\theta^{\sim}_{\tau(H)} = \frac{1}{H} \int_0^{\tau(H)} B(Y_s,s)dY_s$$

is unbiased, has constant variance abd is normally distributed i.e.

$$\theta^{\sim}_{\tau(H)} \sim N(\theta,1/H)$$

For $B(X_t,t)=X_t$ Novikov (1972) studies the mean record length $E_\theta(\tau(H))$ and the mean square error of $\theta^{\wedge}_t$ and $\theta^{\sim}_{\tau(H)}$ showing that the gain of using the sequential ML estimator is especially large if $\theta \geq 0$ (nonstationary case). On sequential ML estimators see also Brown and Hewitt (1975b). Swensen (1985) shows that both estimators are non admissible for the quadratic loss but that the former is minimax (as well as Bayes for uniform prior) for a suitable weighted quadratic loss function.

30. A general proof can be easily obtained via Itô lemma and (29). See also Aase (1982).

31. On discrete approximations to integrals see McShane (1983) p.61.

32. Along the same line, it can be shown that the trapezoidal approximation to (66)

$\theta^{\sim}_{n,T} = (2/\delta_n)(\Sigma_1^{Nn} X_T^2 + \Sigma_1^{Nn} X_{T-1}^2)^{-1}(\Sigma_1^{Nn} X_{T-1}\underline{\triangle} X_T)$

has an asymptotic bias of order δ_n in probability, while the 2SLS estimator

$\theta^{\circ}_{n,T} = (2/\delta_n)(\Sigma_1^{Nn} X_{T-1}(X_T+X_{T-1}))^{-1}(\Sigma_1^{Nn} X_{T-1}\underline{\triangle} X_T)$

derived from the trapezoidal approximation to (63) using X_{T-1} as instrument han an asymptotic bias of order δ_n^2 in probability.

REFERENCES

Aase, K.K. (1982), Stochastic continuous-time model reference adaptive system with decreasing gain, "Advances in Applied Probability", n.14, pp.763-788

Banon, G. (1978), Nonparametric identification for diffusion processes, "SIAM Journal on Control and Optimization", n.16, pp.380-395

Basawa, I.V. and Prakasa-Rao, B.L.S. (1980), Asymptotic Inference for Stochastic Processes, "Stochastic Processes and their Applications", n.10, pp.221-254

Bergstrom, A.R. (ed.) (1976), Statistical Inference in Continuous Time Economic Models, Amsterdam, North Holland.

_______, (1984), Continuous Time Stochastic Models and Issues of Aggregation over Time, in Griliches and Intriligator (eds.) (1984), vol.2, ch.20

Bergstrom, A.R. et al. (eds.) (1978), Stability and Inflation Chichester, Wiley

Black, F. and Scholes, M., (1973), The Pricing of Options and Corporate Liabilities, "Journal of Political Economy", n.81, pp.637-657

Brown, B.M. and Hewitt, J.I. (1975a), Asymptotic Likelihood Theory for diffusion processes, "Journal of Applied Probability", n.12, pp.228-238

_______, (1975b), Inference for the diffusion branching process, "Journal of Applied Probability", n.12, pp.588-594

Cesari, R. (1987) Risks and Equilibrium Prices of Contingent Claims with Application to Italian Securities, unpublished D.Phil. thesis, Oxford University

Christopeit, N. (1986), Quasi-Least-Squares Estimation in Semimartingales Regression Models, "Stochastics", n.16, pp.255-278

Cootner, P.H. (ed.) (1964), The Random Character of Stock Market Prices, Cambridge (MA), MIT Press

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985), A Theory of the Term Structure of Interest Rates, "Econometrica", n.53, pp.385-407

Ershov, M.P. (1970), Sequential Estimation of Diffusion Processes, "Theory of Probability and its Applications", n.15, pp.688-700

Feigin, P.D. (1976), Maximum Likelihood Estimation for Continuous-Time Stochastic Processes, "Advances in Appllied Probability", n.8, pp.712-736

________, (1978), The efficiency criteria problem for stochastic processes, "Stochastic Processes and their Applications", n.6, pp.115-127

________, (1979), Some comments concerning a curious singularity, "Journal of Applied Probability", n.16, pp.440-444

Friedman, A. (1975), Stochastic Differential Equations and Applications, vol.1, New York, Academic Press

Gandolfo, G. (1981), Qualitative Analysis and Econometric Estimation of Continuous Time Dynamic Models, Amsterdam, North Holland

Gani, J., Sarkadi, K. and Vincze, I. (eds.) (1974), Progress in Statistics, Amsterdam, North-Holland, 2 voll.

Gihman, I.I. (1973), Theory of Stochastic Processes n.1, trans.by D.Louvish (1974), New York, Wiley

Girsanov, I.V. (1960), On trasforming a certain class of stochastic processes by absolutely continuous substitution of measures, "Theory of Probability and its Applications", n.5, pp.285-301

Griliches, Z. and Intriligator, M.D. (eds.) (1984), Handbook of Econometrics, Amsterdam, North-Holland, 3 voll.

Hansen, L.P. and Sargent, T.J. (1983), The dimensionality of the aliasing problem in model with rational spectral densities, "Econometrica", n.51, pp.377-387

Heyde, C.C. and Feigin P.D. (1975), On efficiency and exponantial families in stochastic process estimation, in Patil et al. (eds.) (1975), vol.1, 227-240

Hutton, J.E. and Nelson, P.I. (1984), Interchanging the order of differentiation and stochastic integration, "Stochastic Processes and their Applications", n.18, pp.371-377

________, (1986), Quasi-Likelihood Estimation for Semimartingales, "Stochastic Processes and their Applications", n.22, pp.245-257

Jazwinski, A.H. (1970), Stochastic Processes and Filtering Theory New York, Academic Press

Kallianpur, G. (1980), Stochastic Filtering Theory, Berlin, Springer

Kulinich, G.L. (1975), On the estimation of the drift parameter of a stochastic diffusion equation, transl. by K.Wickwire, "Theory of Probability and its Applications", n.20, pp.384-387

Kunita, H. and Watanabe, S. (1967), On square integrable martingales, "Nagoya Mathematical Journal", n.30, pp.209-245

Kutoyants, Yu. A. (1978), Estimation of the trend parameter of a diffusion process in the smooth case, trans. by Durri-Hamdani, "Theory of Probability and its Applications", n.23, pp.399-405

Langetieg, T.C. (1980), A Multivariate Model of the Term Structure, "Journal of Finance", n.35, pp.71-97

Lanska, V. (1979), Minimum contrast estimation in diffusion processes, "Journal of Applied Probability", n.16, pp.65-75

Le Breton, A. (1976), On continuous and discrete sampling for parameter estimation in diffusion type processes, "Mathematical Programming Study", n.5, pp.124-144

_______, (1977), Parameter estimation in a linear stochastic differential equation, "Transactions of the 7th Prague Conference and 1974 EMS", Prague, Academia, pp.353-366

Liptser, R.S. and Shiryayev, A.N. (1974), Statistics of Random Processes, 2 voll., transl. edition (1978), Berlin, Springer

Malliaris, A.G. and Brock, W.A. (1982), Stochastic Methods in Economics and Finance, Amsterdam, North-Holland

McKeague, I.W. (1984), Estimation for diffusion processes under misspecified models, "Journal of Applied Probability", n.21, pp.511-520

McShane, E.J. (1972), Stochastic Differential Equations and Models of Random Processes, "VI Berkeley Symposium on Mathematical Statistics and Probability", pp.263-294

_______, (1983), Unified Integration, New York, Academic Press

Merton, R.C. (1973), An Intertemporal Capital Asset Pricing Model, "Econometrica", n.41, pp.867-887

_______, (1975), Theory of Finance from the Perspective of Continuous Time, "Journal of Financial and Quantitative Analysis", n.10, pp.659-674

Meyer, P.A. (1962), A decomposition theorem for supermartingales, "Illinois Journal of Mathematics", n.6, pp.193-205

_______, (1963), Decomposition of supermartingales: the uniqueness theorem, "Illinois Journal of Mathematics", n.7, pp.1-17

Novikov, A.A. (1972), Sequential estimation of the parameters of diffusion processes, "Mathematical Notes", n.12, pp.812-818

Patil, G.P. et al. (eds.) (1975), Statistical Distributions in Scientific Work, vol.1, Dordrecht, Reidel

Phillips, P.C.B. (1973), The problem of identification in finite parameter continuous time models, "Journal of Econometrics", n.1, pp.351-362

_______, (1978), The treatment of flow data in the estimation of continuous time systems, in Bergstrom et al. (eds.) (1978), ch.15

_______, (1987), Time series regression with a unit root, "Econometrica", n.55, pp.277-301

Prakasa-Rao, B.L.S. and Rubin, H. (1981), Asymptotic Theory of Estimation in Nonlinear Stochastic Differential Equations, "Sankhyá", n.43, Series A, Pt.2, pp.170-189

Richard, S.F. (1978), An Arbitrage Model of the Term Structure of Interest Rates, "Journal of Financial Economics", n.2, pp.187-203

Sargan, J.D. (1976), Some discrete approximation to continuous time stochastic models, in Bergstrom (ed.) (1976), pp.27-80

Sethi, S.P. and Lehoczky, J.P. (1981), A comparison of the Itô and Stratonovich formulations of problems in finance, "Journal of Economic Dynamics and Control", n.3, pp.343-356

Shiryayev, A.N. (1974), Statistics of diffusion processes, in Gani, Sarkadi and Vincze (eds.) (1974), pp.737-751

_______, (1981), Martingales: recent developments results and applications, "International Statistical Review", n.49, pp.199-233

Swensen, A.R. (1985), A note on statistical inference for a class of diffusions and approximate diffusions, "Stochastic Processes and their Applications", n.19, pp.111-123

Taraskin, A.F. (1970), On the asymptotic normality of vector-valued stochastic integrals and estimates of drift parameters of a multidimensional diffusion process, trans.by J.Malek, "Theory of Probability and Mathematical Statistics", n.2, 1974, pp.209-224

________, (1973), Some limit theorems for stochastic integrals, in Gihman (ed.) (1973), pp.136-151

Vasicek, O. (1977), An equilibrium characterization of the term structure, "Journal of Financial Economics", n.5, pp.177-188

Wong, E. and Hajek, B. (1985), Stochastic Processes in Engineering Systems, New York, Springer

RECENTLY PUBLISHED «TEMI» (*)

(*) Requests for copies should be sent to: Banca d'Italia - Servizio Studi, Biblioteca, via Nazionale 91, 00184 Rome.

BANCA D'ITALIA - CENTRO STAMPA