F U N D A Ç Ã O
GETULIO VARGAS

EPGE

Escola de Pós-Graduação
em Economia

Ensaio Econômico

Escola de

Pós Graduação

em Economia

da Fundação

Getúlio Vargas

Nº 540

ISSN 0104-8910

On the statistical estimation of diffusion processes – A survey

Rubens Penha Cysne

Abril de 2004

On the Statistical Estimation of Diffusion Processes - A Survey

Rubens Penha Cysne*

April 1, 2004

Abstract

Data available on continuous-time diffusions are always sampled discretely in time. In most cases, the likelihood function of the observations is not directly computable. This survey covers a sample of the statistical methods that have been developed to solve this problem. We concentrate on some recent contributions to the literature based on three different approaches to the problem: an improvement of the Euler-Maruyama discretization scheme, the use of Martingale Estimating Functions and the application of Generalized Method of Moments (GMM).

1 Introduction

Rigorous definitions of a diffusion process can be found in Krylov ((1980) or in Karatzas and Shreve (1991). Loosely speaking, a diffusion process is a Markov processes with continuous sample paths which can be characterized by an infinitesimal generator (to be defined below). The simplest diffusion process is the Wiener process, the stochastic process that corresponds to the Brownian Motion.

Our main concern here will be with the statistical inference of diffusion processes using discretely-sampled data. This is always the case in economics and finance, since data on exchange rates, interest rates, stock prices etc., only change and are only recorded in discrete points of time.

*Getulio Vargas Foundation Graduate School of Economics.

As a general point of departure for the type of problem in which we shall be interested, consider the stochastic integral equation relative to a stochastic process X_t in $\mathbb{R}^d$:

$$X_t = X_0 + \int_0^t h(\theta, s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, \quad 0 \leq s \leq t$$

In this equation, X_0 is an $\mathcal{F}_0$ -measurable function independent of $\{W_u - W_v, u \geq v \geq 0\}$ and W is a standard Brownian Motion.

Let $\mathcal{F}_{0,t}^W$ be the completion of the σ -algebra generated by $\{W_u, t \geq u \geq 0\}$. Denote by $\mathcal{F}_{0,t}$ the σ -algebra generated by $\mathcal{F}_0$ and $\mathcal{F}_{0,t}^W$. To simplify notation, make $\mathcal{F}_{0,t} = \mathcal{F}_t$. Now suppose $h : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is $\mathcal{F}_t$ -measurable and σ is a $(d \times m)$ $\mathcal{F}_t$ -measurable matrix with $\sigma_{ij} : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$. Under a set of Lipschitz conditions (see, e.g., Prakasa Rao (1999) or Oksendal (2000, section 5.2)), the equation:

$$dX_t = h(\theta, t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x_0 \in \mathbb{R}^d, t \geq 0 \quad (1)$$

has a unique and continuous solution $X_t(t, \omega)$ and, for each $t \geq 0$, X_t is $(\mathcal{F}_t \mid \text{Borel})$ -measurable. $\{X_t\}$ is a continuous Markov process relative to $\mathcal{F}_t$.

The process is called homogenous if $h(\theta, s, X_s) = h(\theta, X_s)$ and $\sigma(s, X_s) = \sigma(X_s)$. In general, we shall be interested in homogenous processes in $\mathbb{R}$ ($d = 1$).

For the purpose of stochastic modelling, we can think of a diffusion process as a continuous version of a process:

$$X_{t+1} = f(X_t, s_t) + \epsilon_t$$

where X_t stands for the state at the t -th generation, s_t for a random or fixed parameter at the t -th generation and ϵ_t for a noise.

Example 1 (*Wiener Process with Drift μ and Diffusion σ^2*):

$$dX_t = \mu dt + \sigma dW_t, \quad X_0 = 0 \in \mathbb{R}, t \geq 0$$

In this case $X(t) - X(s)$, $t > s > 0$, is normal with independent increments, mean $E(X(t) - X(s)) = \mu |s - t|$ and $\text{Var}(X(t) - X(s)) = \sigma^2 |s - t|$.

Ideally, parametric inference for diffusion processes should be based on the likelihood function. Since such processes are Markovian, the likelihood function (given that the initial point is known) is the product of transition densities. However, the transition densities f_k on which the maximum

likelihood function has to rely can only be obtained in closed-form in very specific cases. Given this hindrance to the direct application of the likelihood method, different alternatives have been proposed in the literature.

Pedersen (1995a, 1995b) derived the estimators departing from approximations of the continuous-time likelihood function using simulation methods. More recently, Ait-Sahalia (2002) has proposed the use of closed-form approximations of the (unknown) likelihood functions based on Hermite polynomials. The estimator so obtained is shown by the author to converge to the true maximum likelihood estimator and to share its asymptotic properties. The method proposed by Ait-Sahalia starts by making a transformation of the original process, from X to Z . Z is a process for which the Hermite expansion of the transition densities converges. It is the appropriate transformation of X whose expansion starts with a $N(0,1)$ term. Since Z is a known transformation of X , the expansion of the density of X can then be obtained by the use of the Jacobian formula, thereby leading (analytically) to closed-form approximations of the maximum likelihood function.

We will concentrate our analysis on three distinct alternatives to the problem. The first two of them, the Euler-Maruyama discretization and the use of Martingale Estimating Functions, are based on replacements of the true likelihood function, which is not known, by some approximation. Such procedures, sometimes classified as indirect inference (Gourieroux (2001)), are usually followed, in order to improve the quality of the estimators, by additional steps involving simulation and calibration. The third method object of our analysis will be the Generalized Method of Moments (GMM).

2 Some Applications in Finance and Economics

Diffusion processes provide an alternative to the discrete-time stochastic processes traditionally used in time series analysis. The need of modelling and estimating such processes has been particularly important in finance and economics, where they are fitted to time series of, for instance, stock prices, interest rates and exchange rates, in order to price derivative assets.

The applications shown below fall into a category (see section 4.2.1) in which the diffusion process is of a type such that the transition functions are known. In this case the parameters can be directly estimated by maximum likelihood. The reason we introduce such applications here is that they are a point of departure to more complicated models, using other underlying diffusion processes, which do not lead to transition functions that are known. For instance, there is considerable evidence that the increments of the logarithm of the price of the stock used to price the option in the Black and

Scholes model are neither independent nor Gaussian, as implied by equation (2) below. This leads to the necessity of more complex estimation processes, the analysis of which is the purpose of this survey.

- Black and Scholes (1973). We present here the version of Campbell et alii (1997). Suppose we want to find the price $G(P(t), \tau)$ at time t , of an (European) option with strike price X and expiration date $T > t$, with $\tau = T - t$. We assume that the relative changes of prices follow the equation:

$$\frac{dP(t)}{dt} = k(t)P(t), \quad P_0 \text{ given}$$

with $k(t) = \mu + \sigma Z(t)$, $Z(t)$ a white noise and μ and σ constants. In Itô's representation:

$$dP(t) = \mu P(t)dt + \sigma P(t)dW(t), \quad t \geq 0 \quad (2)$$

$W(t)$ standing for a standard Brownian motion. The hypothesis of the model is that $P(t)$ models the stock price upon which the option price is based. Now suppose (we omit the arguments of the function $P(\cdot)$) that an initial investment I is allocated in options and stocks according to

$$I = G(P, t) + \alpha P \quad (3)$$

Using Itô's Lemma:

$$\begin{aligned} dG(P, t) &= dt \left[\mu P G_P + G_t + \frac{1}{2} P^2 \sigma^2 G_{PP} \right] + P G_P \sigma dW \\ dI(t) &= dt \left[(\alpha + G_P) P \mu + G_t + \frac{1}{2} P^2 \sigma^2 G_{PP} \right] + (\alpha + G_P) P \sigma dW \end{aligned}$$

The risk is zero when $dI(t)$ does not depend on the stochastic component $(\alpha + G_P) P \sigma dW$, which implies $\alpha + G_P = 0$. In this case the expected income per unit of time is $G_t + \frac{1}{2} P^2 \sigma^2 G_{PP}$. Denoting by r the risk-free rate, the no-arbitrage condition demands:

$$G_t + \frac{1}{2} P^2 \sigma^2 G_{PP} = rI$$

Using (3) and the no-risk condition once more:

$$G_t + \frac{1}{2} P^2 \sigma^2 G_{PP} = r(G + \alpha P) = r(G - G_P P)$$

from which we get:

$$G_t + \frac{1}{2}P^2\sigma^2G_{PP} - r(G - G_P P) = 0 \quad (4)$$

Since the (European) option is only exercised if the price at time T is no less than the strike price X :

$$G(P(T), T) = \max(0, P(T) - X) \quad (5)$$

Solving (4) with condition (5) and the condition $G(0, t) = 0, \forall t > 0$, gives the price of the option as a function of time and of the parameter (to be estimated) σ .

- Krugman (1991) is an example of a model used in the determination of exchange rates. Krugman develops a model to analyze the exchange rate under a target-zone regime. All variables are expressed in logarithms. The exchange rate (s) is supposed to be determined by the money supply (m), the velocity of money (v) and the expectation of exchange rate devaluation $E(ds/dt)$:

$$s = m + v + \gamma E(ds/dt) \quad (6)$$

m is an exogenous variable determined by the Central Bank in order to maintain the price of the foreign currency of reference in the target zone $[s_l, s_h]$. v is supposed to follow a continuous-time random walk given by $dv = \sigma dz$. γ is a parameter to be determined. (6) and the monetary rule leads to $s = g(m, v, s_l, s_h)$. For a fixed m , Itô's rule implies:

$$E\left(\frac{ds}{dt}\right) = \frac{1}{2}\sigma^2 g_{vv}(m, v, s_l, s_h)$$

Substituting $E(\frac{ds}{dt})$ in (6):

$$s = g(m, v, s_l, s_h) = m + v + \frac{1}{2}\gamma\sigma^2 g_{vv}(m, v, s_l, s_h)$$

a second order non-homogeneous differential equation with constant coefficients. The characteristic function associated with this differential equation is $f(r) = r^2 k - 1$, where $k = \gamma\sigma^2/2$. The general solution of the equation is then:

$$g(m, v, s_l, s_h) = A_1 \exp(-v/\sqrt{k}) + A_2 \exp(-v/\sqrt{k}) + v + m$$

A_1 and A_2 being constants to be determined. By making, without loss of generality, $m = 0$, by symmetry $A_1 = -A_2$ and:

$$g(m, v, s_l, s_h) = A_1 \left[\exp(-v/\sqrt{k}) - \exp(-v/\sqrt{k}) \right] + v + m$$

Note that this model has as parameters to be estimated A_1, γ and σ .

- Cox, Ingersoll and Ross (1985): In this model the state variable follows a diffusion process given by:

$$dX_t = (\alpha + \beta X_t)dt + \sigma \sqrt{X_t}dW_t$$

in which case the parameters to be estimated are $\theta = (\alpha, \beta, \sigma)$. This diffusion process can be easily solved by means of a transformation of variables. The solution leads to non-central chi-square transition densities, there being no need of approximations.

3 The Generator of a Diffusion Process

Let $f(\cdot)$ be a bounded twice continuously differentiable function, with bounded derivatives and X_t a generic time-homogeneous diffusion process defined on the probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{Q}$ be the probability measure induced by X_t on R^n (for any t) and $L^2(\mathcal{Q})$ be the space of Borel measurable functions $f(X_t) : R^n \rightarrow R$, $\mathcal{Q}$ -square integrable. In this space (not distinguishing between the space itself and the equivalent-classes space) we define, for $t \geq 0$, the family of operators:

$$\Gamma_t f(x_0) = E[f(X_t) | X_0 = x_0] \quad (\equiv E^0(f(X_t))) \quad (7)$$

It can be shown that these operators ($L^2(\mathcal{Q}) \rightarrow L^2(\mathcal{Q})$) are well defined ($f = f^*$ $\mathcal{Q}$ -ae $\rightarrow \Gamma_t(f) = \Gamma_t(f^*)$ $\mathcal{Q}$ -ae and $\Gamma_t(f^*) \in L^2(\mathcal{Q})$), a weak contraction ($\|\Gamma_t(f)\| \leq \|f\|$) and a semi-group (by the law of iterated expectations, $E^0(X_{t+s}) = E^0(E^t(X_{t+s}))$, implying $\Gamma_{t+s} = \Gamma_t \Gamma_s$).

In the remaining of this text, we will some times refer to the infinitesimal generator of a diffusion process $f(X_t)$. The infinitesimal generator gives a measure of the infinitesimal drift of a diffusion. For some functions $f \in L^2(\mathcal{Q})$ for which the limit below exists (call it Ψ , a proper subset of $L^2(\mathcal{Q})$), this is defined as:

$$\Lambda f(x_0) = \lim_{t \downarrow 0} \frac{\Gamma_t f(x_0) - f(x_0)}{t}, \quad t \geq 0 \quad (8)$$

Γ and Λ commute on Ψ and Ψ is dense in $L^2(\mathcal{Q})$. We need a Proposition about the way how this generator materializes in the case of a particular diffusion process. The initial part of the proof is done in Scheinkman and Hansen (1995).

Proposition 1 *Consider the one-dimensional diffusion process defined as solution to the stochastic differential equation:*

$$dX_t = b(X_t; \theta)dt + \sigma(X_t; \theta)dW_t \quad (9)$$

where W_t is a Wiener process. Let L_θ be a (differential) operator defined by:

$$L_\theta = b(x; \theta) \frac{d}{dx} + \frac{1}{2} \sigma^2(x; \theta) \frac{d^2}{dx^2} \quad (10)$$

Then $\Lambda f(x_0) = L_\theta f(x_0)$.

Proof. We divide the Proof in six parts.

I- Write $Y_t = f(X_t)$, use (9) and apply Itô's formula to get:

$$dY_t = f'(X_t)(b(X_t; \theta)dt + \sigma(X_t; \theta)dW_t) + \frac{1}{2}f''(X_t)\sigma^2(X_t; \theta)(dW_t)^2$$

II- Substitute dt for $(dW_t)^2$ and integrate to get:

$$\begin{aligned} Y_t &= f(X_t) = f(x_0) + \int_0^t \left[b(X_s; \theta)f'(X_s; \theta) + \frac{1}{2}f''(X_s)\sigma^2(X_s; \theta) \right] ds + \\ &\quad + \int_0^t f'(X_s)\sigma(X_s; \theta)dW_s \end{aligned}$$

III- By the construction of the Itô's Integral, for $u < t$,

$$\Gamma_u \int_0^t f'(X_s)\sigma(X_s; \theta)dW_s = \int_0^u f'(X_s)\sigma(X_s; \theta)dW_s$$

Using the definition of L_θ established by (10),

$$Y_t - f(x_0) - \int_0^t L_\theta f(X_s)ds$$

is a continuous martingale.

IV- Taking expectations conditional on x_0 :

$$E[f(X_t) | x_0] - f(x_0) = E\left[\int_0^t L_\theta f(X_s)ds | x_0\right]$$

V- Using Fubini's theorem (by assumption, the integrand is quasi-integrable w.r.t the product measure) and definition (7):

$$\frac{\Gamma_t f(x_0) - f(x_0)}{t} = (1/t) \int_0^t \Gamma_s L_\theta f(x_0)ds$$

VI- Now take limits with $t \downarrow 0$ on both sides of the above equation. The left side, by definition, is equal to Λf . Therefore, the demonstration will be

finished once we show that the limit of the right side equals $L_\theta f$. We need to show:

$$\int_{R^n} \left\{ \left[(1/t) \int_0^t (\Gamma_s L_\theta f(x_0) - L_\theta f(x_0)) ds \right]^2 \right\} d\mathcal{Q} \rightarrow 0$$

Using the Cauchy-Schwarz (Hölder) inequality:

$$\begin{aligned} & \int_{R^n} \left\{ \left[(1/t) \int_0^t (\Gamma_s L_\theta f(x_0) - L_\theta f(x_0)) ds \right]^2 \right\} d\mathcal{Q} \\ & \leq (1/t)^2 \int_{R^n} \left\{ \int_0^t [\Gamma_s L_\theta f(x_0) - L_\theta f(x_0)]^2 ds \int_0^t 1^2 ds \right\} d\mathcal{Q} \\ & = (1/t) \int_{R^n} \left\{ \int_0^t [\Gamma_s L_\theta f(x_0) - L_\theta f(x_0)]^2 ds \right\} d\mathcal{Q} \end{aligned}$$

Using Fubini again,

$$\begin{aligned} & (1/t) \int_{R^n} \left\{ \int_0^t [\Gamma_s L_\theta f(x_0) - L_\theta f(x_0)]^2 ds \right\} d\mathcal{Q} \\ & = (1/t) \int_0^t \left\{ \int_{R^n} [\Gamma_s L_\theta f(x_0) - L_\theta f(x_0)]^2 d\mathcal{Q} \right\} ds \\ & = (1/t) \int_0^t \|\Gamma_s L_\theta f(x_0) - L_\theta f(x_0)\|^2 ds \end{aligned}$$

which goes to zero by the assumption that X_t is (Borel) measurable with respect to the product sigma-algebra (the sigma algebra generated by the measurable rectangles AxB , $A \in \mathcal{R}$, $B \in \mathcal{F}$ (this implies ⁽¹⁾ that, for each ϕ , $\{\Gamma_t \phi, t \geq 0\}$ converges to ϕ as $t \downarrow 0$.) ■

4 Maximum Likelihood Estimation (MLE)

4.1 Continuously Observed Data

Likelihood methods for continuously observed diffusions are standard in the

literature. We base our exposition here on Prakasa Rao (1999) and Basawa and Prakasa Rao (1980). The important point to note is that likelihood function in this case can be obtained by a classical result on change of measure.

Consider the diffusion process (1). Note that the function $\sigma(s, X_t)$ does

¹See footnote 4 in Hansen and Scheinkman (1995).

not depend on the parameter θ . We assume that $\sigma(s, X_t)$ is known or, alternatively, that it is a constant, in which case it can be estimated from a standard quadratic-variation property of the Wiener process. Stokey (2000, chapter 2) presents this as an exercise.

Proposition 2 *If X is a (μ, σ^2) Brownian motion, then over any finite interval $[S, S + T]$:*

$$QV_n \equiv \sum_{j=1}^{2^n} \left[X\left(\frac{jT}{2^n}, \omega\right) - X\left(\frac{(j-1)T}{2^n}, \omega\right) \right]^2 \rightarrow \sigma^2 T, \quad P - a.e. \text{ as } n \rightarrow \infty.$$

(QV stands for quadratic variation).

Proof. We omit the ω . Make

$$\Delta_{j,n} = X\left(\frac{jT}{2^n}\right) - X\left(\frac{(j-1)T}{2^n}\right)$$

Then $E(\Delta_{j,n}) = 0$ and

$$E(\Delta_{j,n})^2 = E\left(\frac{X\left(\frac{jT}{2^n}\right) - X\left(\frac{(j-1)T}{2^n}\right)}{\sqrt{\frac{T}{2^n}\sigma^2}}\right)^2 \frac{\sigma^2 T}{2^n} = \frac{\sigma^2 T}{2^n}$$

because $\frac{X\left(\frac{jT}{2^n}\right) - X\left(\frac{(j-1)T}{2^n}\right)}{\sqrt{\frac{T}{2^n}\sigma^2}}$ is a $N(0, 1)$ r.v. Define $x_n = \sum_{j=1}^{2^n} (\Delta_{j,n})^2$. Then

$$Ex_n \equiv E\left(\sum_{j=1}^{2^n} (\Delta_{j,n})^2\right) = \sigma^2 T$$

and

$$E(x_n - \sigma^2 T)^2 = \text{Var}(x_n) = \text{Var}\left(\sum_{j=1}^{2^n} (\Delta_{j,n})^2\right)$$

By the independence of increments:

$$E(x_n - \sigma^2 T)^2 = \sum_{j=1}^{2^n} \text{Var}(\Delta_{j,n})^2 = \sum_{j=1}^{2^n} \frac{\sigma^4 T^2}{2^{2n}} \text{Var}\left(\left(\frac{X\left(\frac{jT}{2^n}\right) - X\left(\frac{(j-1)T}{2^n}\right)}{\sqrt{\sigma^2 \frac{T}{2^n}}}\right)^2\right) = \frac{2\sigma^4 T^2}{2^n} \quad (11)$$

because $\left(\frac{X\left(\frac{jT}{2^n}\right) - X\left(\frac{(j-1)T}{2^n}\right)}{\sqrt{\sigma^2 \frac{T}{2^n}}}\right)^2$ has a chi-square distribution with one degree of freedom. (11) implies convergence of x_n to $\sigma^2 T$ in $L_2(P)$. To prove, as required, that $QV \equiv \lim_n x_n = \sigma^2 T$, $P - a.e.$, note that:

$$\sum_{n=1}^{\infty} E(x_n - \sigma^2 T)^2 = 2\sigma^4 T^2 < \infty$$

Since $(x_n - T)^2 \geq 0$, $\sum_{n=1}^{\infty} E(x_n - \sigma^2 T)^2 = E(\sum_{n=1}^{\infty} (x_n - \sigma^2 T)^2) < \infty$. This implies

$$\sum_{n=1}^{\infty} (x_n - \sigma^2 T)^2 < \infty, \text{ P-a.e.}$$

$$\lim_n (x_n - \sigma^2 T)^2 = 0, \text{ P-a.e.}$$

and

$$QV \equiv \lim_n x_n = \sigma^2 T \text{ P-a.e.}$$

■

This proposition allows us to consider (1) with $\sigma = 1$.

Formally, let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t, t \geq 0\}$ a filtration in $(\Omega, \mathcal{F})$. Suppose $\{X_t\}$ is adapted to this filtration and satisfies (1). Let P_{θ}^T be the probability measure generated by $\{X_t, 0 \leq t \leq T\}$ on the space $(C[0, T], \mathcal{B}_T)$, $\mathcal{B}$ corresponding to the Borel sigma-algebra defined in $C[0, T]$. By this we mean:

$$P_{\theta}^T(B) = P\{w \in \Omega : X_t \in B, \quad B \in \mathcal{B}_T\}$$

$P_{\theta}^T(B)$ is the measure induced by the process $X_t(\theta)$ on $C[0, T]$.

In the same way, let P_W^T be the probability measure induced by the Wiener process in $C[0, T]$:

$$P_W^T(B) = P\{w \in \Omega : W_t \in B, \quad B \in \mathcal{B}_T\}$$

Then, under regularity conditions ensuring (for all $\theta \in \Theta$) the absolute continuity of P_{θ}^T with respect to P_W^T , the Radon-Nikodym derivative $\frac{dP_{\theta}^T}{dP_W^T}$ is given by (see Oksendall (2000), Girsanov's theorem):

$$\frac{dP_{\theta}^T}{dP_W^T} = \exp \left\{ \int_0^T h(\theta, s, X_t) dX_t - \frac{1}{2} \int_0^T h^2(\theta, s, X_t) dt \right\}, [a.s.(P)]$$

By definition, the MLE $\hat{\theta}_T(X_T)$ of θ is defined by the measurable map $\hat{\theta}_T : ((C[0, T], \mathcal{B}_T) \rightarrow (\Theta, \tau)$, such that:

$$\frac{dP_{\hat{\theta}}^T}{dP_W^T} = \sup_{\theta \in \Theta} \left(\frac{dP_{\theta}^T}{dP_W^T} \right)$$

where τ is the σ -algebra of Borel subsets of Θ .

Example 2 Making $h(\theta, s, X_t) = \theta$, the above equation leads to the maximization of $f(\theta) = \theta X_T - \theta \frac{T^2}{2}$, with solution $\hat{\theta} = X_T/T$.

4.2 Discretely Observed Data

4.2.1 The Case When the Transition Densities are known

There are three well known cases when the stochastic differential equation (1) is easily solvable, and the corresponding transition functions known²: i) $h(\theta, t, X_t) = \mu X_t$, $\sigma(t, X_t) = \sigma X_t$, called geometric brownian motion, used, for instance, in the Black and Scholes model; ii) $h(\theta, t, X_t) = \alpha(\beta - X_t)$, $\sigma(t, X_t) = \sigma$, the Orstein-Uhlenbeck process, used, for instance, by Vasicek (1977) to analyze the dynamics of the short-term interest rate and; iii) $h(\theta, t, X_t) = \mu(\beta - X_t)$, $\sigma(t, X_t) = \sigma\sqrt{X_t}$, which is the diffusion used in the Cox-Ingersoll-Ross model of the term structure of interest rates. The first of these processes leads to log-normal, the second to normal, and the third to non-central chi-square transition densities.

To exemplify the use of maximum likelihood in this case, let us consider the equation that describes the evolution of the price of the underlying stock in the Black and Scholes model³, which falls into the first case considered above:

$$dP(t) = \mu P(t)dt + \sigma P(t)dW(t), \quad t \geq 0 \quad (12)$$

Define $Y_t = \log(P_t)$. Using Itô's rule:

$$dY_t = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW(t)$$

The above equation implies a normal distribution for the transition densities of Y_t . Integrating,

$$\log P_t = \log P_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W(t) \quad (13)$$

$$P_t = P_0 e^{(\mu - \frac{1}{2}\sigma^2)t} e^{\sigma W(t)}$$

The conditional distribution of P_t given P_0 is a log normal with mean $\log P_0 + (\mu - \frac{1}{2}\sigma^2)t$ and variance $\sigma^2 t$. The conditional mean of $P(t)$ given P_0 can be obtained by using the formula for the moment generating function of a normal random variable of mean $(\mu - \frac{1}{2}\sigma^2)t$ and variance $\sigma^2 t$:

$$E(P_t | P_0) = P_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \frac{\sigma^2}{2}t} = P_0 e^{\mu t}$$

²Wong (1964) investigates some other particular cases.

³Note that this same diffusion equation (usually called geometric Brownian motion) could be used to model different phenomena, in particular populational growth.

Since the transition densities of Y_t are known, the application of maximum likelihood in this case follows in a straightforward way.

One nice feature of continuous modelling is that we can analyze what happens when the time between observations tends to zero. In the present case, this can be done with the help of equation (13). If the time between observations is h , $\log P_t - \log P_{t-h}$ has a Gaussian distribution with mean $(\mu - \frac{\sigma^2}{2})h$ and variance $\sigma^2 h$. The estimators of both μ and σ^2 (trivially obtained by maximum likelihood), as well as their asymptotic variances, are functions of h and T , the number of observations. One can show that the variance of the volatility parameter depends only on the number of observations T . It does not depend upon the sampling frequency h . The variance of the drift parameter μ , though, depends on both T and h . The drift parameter μ cannot be consistently estimated when the whole time span of the observations is fixed, even if $h \rightarrow 0$ (with $hT \rightarrow k \in \mathbb{R}$)⁴.

4.2.2 The Case when the Transition Densities are not Known

As mentioned before, the diffusion for the prices of stocks described in the Black and Scholes model is usually not supported by the data. Here we analyze estimation in more general cases.

Given $n + 1$ observations of a diffusion process like (1) consider the data $X(t)$ sampled at non-stochastic dates $t_0 = 0 < t_1 < \dots < t_n$ (equally spaced or not). The joint density of the sample is given by:

$$p(X_0, X_1, \dots, X_n) = p_0(X_0, \theta) \prod_{j=1}^n p_k(X_{t_j}, t_j \mid X_{t_{j-1}}, t_{j-1}; \theta) \quad (14)$$

where $p_0(X_0)$ is the marginal density function of X_0 and $p(X_{t_j}, t_j \mid X_{t_{j-1}}, t_{j-1}; \theta)$

represents the transition density functions. Such functions are usually not known. In this Section we examine how this problem can be dealt with by using Gaussian distributions to approximate the densities. When the distance between observations is sufficiently small, such approximations lead to reasonable estimators, although biased. Sorensen (1995) is the main source of our analysis in this Section and the next.

To see how it works (Prakasa Rao (1999), Florens-Zmirou (1989)), let us start with the diffusion:

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t, \quad X_0 = x_0 \quad (15)$$

Following Sorensen (1995), we assume that the functions $b(X_t, \theta)$ and $\sigma(X_t, \theta)$ are known, apart from the parameter θ , which varies in a subset Θ

⁴The asymptotic variance of the MLE estimator of σ^2 is equal to $2\sigma^4/T$.

of $\mathbb{R}^d$. We discretize this process by assuming that the drift and the diffusion are constant in the time interval⁵ $\Delta_i = t_i - t_{i-1}$:

$$X_{t_i} - X_{t_{i-1}} = b(X_{t_{i-1}}, \theta)\Delta_i + \sigma(X_{t_{i-1}}, \theta)(W_{t_i} - W_{t_{i-1}})$$

Since $W_{t_i} - W_{t_{i-1}} \mid W_{t_{i-1}} \sim N(0, \sigma^2(X_{t_{i-1}}, \theta)\Delta_i)$, we are actually assuming:

$$E_\theta(X_{t_i} \mid X_{t_{i-1}}) = b(X_{t_{i-1}}, \theta)\Delta_i + X_{t_{i-1}} \quad (16)$$

$$E_\theta [(X_{t_i} - E_\theta(X_{t_i} \mid X_{t_{i-1}}))^2 \mid X_{t_{i-1}}] = \sigma^2(X_{t_{i-1}}, \theta)\Delta_i \quad (17)$$

Note that there are two types of approximation here, one regarding the moments and the other regarding the distribution of the transition densities, which are assumed to be Gaussian. The former usually introduces biases, whereas the latter leads to inefficiency (Sorensen, 2002).

The transition density of the discretized process then reads:

$$p(X_{t_i} \mid X_{t_{i-1}}) = \frac{1}{\sqrt{2\pi\sigma^2(X_{t_{i-1}}, \theta)\Delta_i}} \exp\left(-\frac{1}{2} \frac{(X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}}, \theta)\Delta_i)^2}{\sigma^2(X_{t_{i-1}}, \theta)\Delta_i}\right)$$

The joint density of $X_{t_0}, \dots, X_{t_n}$ is then given by:

$$L_n(\theta) = \prod_{i=1}^n p(X_{t_i} \mid X_{t_{i-1}}) p(X_{t_0}) \quad (18)$$

Using the last two results and taking logs, the parameters of the problem can be found by the maximization of:

$$l_N(\theta) = -\frac{1}{2} \sum_{i=1}^N \left\{ \frac{(X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}}, \theta)\Delta_i)^2}{\sigma^2(X_{t_{i-1}}, \theta)\Delta_i} + \log(2\pi\sigma^2(X_{t_{i-1}}, \theta)\Delta_i) \right\} + \log p(X_{t_0}) \quad (19)$$

Taking the derivative with respect to θ leads to the score function:

$$H_N(\theta) = \sum_{i=1}^N \left\{ \frac{\frac{b_\theta(X_{t_{i-1}}, \theta)}{\sigma^2(X_{t_{i-1}}, \theta)} [(X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}}, \theta)\Delta_i)] + \frac{\sigma_\theta^2(X_{t_{i-1}}, \theta)}{2(\sigma^2(X_{t_{i-1}}, \theta))^2\Delta_i} [(X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}}, \theta)\Delta_i)^2 - \sigma^2(X_{t_{i-1}}, \theta)\Delta_i] \right\} \quad (20)$$

⁵For equidistant intervals, such approximation, usually called an Euler-Maruyama approximation, can be written

$$X_{t_i} - X_{t_{i-1}} = b(X_{t_{i-1}}, \theta) + \epsilon_{t_i}, \quad \epsilon_{t_i} \mid X_{t_{i-1}} \sim N(0, \sigma^2(X_{t_{i-1}}, \theta))$$

with the subindex $(.)_\theta$ standing for the vector of partial derivatives with respect to θ .

This technique is known as the Euler-Maruyama approximation of the likelihood function (Kloeden and Platen, 1992). By making, in (20), $\sigma^2(X_{t_{i-1}}, \theta) = v(X_{t_{i-1}}, \theta)$ we obtain equation [(2.3)] derived in Sorensen (1995). By deleting the quadratic term (which would be the case when σ^2 is known), we get:

$$\tilde{H}_N(\theta) = \sum_{i=1}^N \left\{ \frac{b_\theta(X_{t_{i-1}}; \theta)}{v(X_{t_{i-1}}; \theta)} [(X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}}; \theta)\Delta_i)] \right\} \quad (21)$$

This estimating function has been studied by Dorogovtsev (1976), Prakasa Rao (1983, 1988), Florens-Zmirou (1989) and Yoshida (1992) in the case when the diffusion coefficient is constant and the parameter θ is unidimensional.

Basically, these authors have shown that expecting these estimators to be consistent and asymptotically normal requires assuming that the length of the observation interval ($n\Delta_n$) goes to infinity and that the time between consecutive observations (Δ_n) goes to zero. Yoshida (1992) and Florens-Zmirou (1989) worked out the case when $\sigma^2(x, \theta) = \theta\sigma^2(x)$ (multiplicative condition). Under such assumption, σ was estimated on the basis of quadratic variation and the estimator $\hat{\theta}$ was obtained from the maximization of (19). Florens-Zmirou imposed $n\Delta_n^2 \rightarrow 0$, and Yoshida the less restrictive assumption $n\Delta_n^3 \rightarrow 0$.

Summing up, the estimation by discretization of the transition function works reasonably well when the time between observations, Δ , is sufficiently small. Kloeden et al. (1992) confirmed this fact through simulation, whereas Pedersen (1995a) and Bibby and Sorensen (1995a) have shown that if Δ is not small the bias can be severe⁶.

Improving the Approximations for the Moments Lemma 1 in Florens-Zmirou⁷ (1989) provides an expansion of $E_\theta(X_\Delta | X_0 = x)$ which can be used to improve (16) and (17) to second or higher order. This Lemma will allow us to get better approximations of the average and of the variance of the Gaussian approximations to the transition functions. It reads:

⁶Estimators (20) and (21) have sometimes been referred to in the econometrics literature as belonging to the GMM class. Sorensen (1995) calls this denomination “odd”, arguing that the method is not one of moments, except approximately.

⁷Florens-Zmirou refers to Dacunha-Castelle and Duflo (1982) as the original reference for the Lemma.

Lemma 3 (Lemma 1 in Florens-Zmirou, 1989): Let $f \in C^{(2s+2)}$ and denote by E^k the conditional expectation w.r.t. $\sigma(X_u, u \leq k\Delta)$ (the σ -algebra generated by $(X_u, u \leq k\Delta)$). Then, with E^{k-1} denoting the conditional expectation w.r.t the information available at date $k - 1$:

$$E^{k-1}f(X_{k\Delta}) = \sum_{l=0}^s \frac{\Delta^l}{l!} L^l f(X_{(k-1)\Delta}) + \int_0^\Delta \int_0^{u_1} \dots \int_0^{u_s} E^{k-1}(L^{s+1}f)(X_{(k-1)\Delta+u_s+1}) du_1 \dots du_{s+1} \quad (22)$$

Notice in the expression above the presence of the operator L derived in Section 3. This expression, among other things, can be used to determine the bias of the estimator $\tilde{\theta}$ derived from (21).

The new equations for the conditional average and variance in (16) and (17) (writing b for $b(x; \theta)$ and v for $v(x; \theta)$) read:

$$E_\theta(X_\Delta \mid X_0 = x) = x + \Delta b + \frac{1}{2} \Delta^2 \left\{ bb_x + \frac{1}{2} vb_{xx} \right\} + O(\Delta^3) \quad (23)$$

and :

$$Var_\theta(X_\Delta \mid X_0 = x) = v\Delta + \Delta^2 \left[\frac{1}{2} bv_x + v \left\{ b_x + \frac{1}{4} v_{xx} \right\} \right] + O(\Delta^3) \quad (24)$$

Note that (16) and (17) are a particular case of these expressions, for the cases when $l = 1$ in (22).

In order to derive these expressions, note that, from Lemma 1 in Florens-Zmirou, making $f(x) = x$, we have::

$$\begin{aligned} E_\theta(X_\Delta \mid X_0 = x) &= x + \Delta Lx + \frac{\Delta^2}{2} L^2 x + \\ &+ \int_0^\Delta \int_0^{u_1} \int_0^{u_2} E(L^3 x)(X_z) du_1 du_2 du_3 \end{aligned}$$

X_z standing for $X_{(k-1)\Delta+u_s+1}$ in Lemma 1. Since $Lx = b(x, \theta)$, $L^2 x = Lb(x, \theta) = bb_x + \frac{1}{2} vb_{xx}$ we get (23). The remaining $O(\Delta^3)$ derives from the fact that the (absolute value of the) integrand in (22) is supposed to be bounded (Sorensen, 1995 provides sufficient conditions in some particular cases) by some $M \in \mathbb{R}_+$ and $u_i \leq \Delta, i = 1, 2$, in which case,

$$\int_0^\Delta \int_0^{u_1} \int_0^{u_2} |E(L^{s+1}x)(X_z) du_1 du_2 du_3| \leq M\Delta^3$$

To get (24) we need $E_\theta(X_\Delta^2 \mid X_0 = x)$. Again, using (22):

$$\begin{aligned} E_\theta(X_\Delta^2 \mid X_0 = x) &= x^2 + \Delta Lx^2 + \frac{\Delta^2}{2} L^2 x^2 + \\ &+ \int_0^\Delta \int_0^{u_1} \int_0^{u_2} E(L^{s+1} x^2)(X_z^2) du_1 \dots du_3 \end{aligned}$$

We have: $Lx^2 = 2bx + v$, $L^2 x^2 = L(2bx + v) = b(2b + 2xb_x + v_x) + \frac{1}{2}v(2b_x + 2b_{xx} + v_{xx})$. Hence,

$$\begin{aligned} E_\theta(X_\Delta^2 \mid X_0 = x) &= x^2 + \Delta(2bx + v) + \frac{\Delta^2}{2}(b(2b + 2xb_x + v_x) \\ &+ \frac{1}{2}v(2b_x + 2b_{xx} + v_{xx})) \end{aligned} \quad (25)$$

From (23) we get:

$$[E_\theta(X_\Delta \mid X_0 = x)]^2 = x^2 + \Delta 2bx + \frac{\Delta^2}{2}(2b^2 + 2xb b_x + xv b_{xx}) + O(\Delta^3) \quad (26)$$

By subtracting (26) from (25) one gets (24).

Following Sorensen (1995), suppose X is an ergodic diffusion with invariant probability μ_θ when θ is the true parameter. Assuming the process departs from the invariant measure, the expressions (21) and (23) imply a bias of the estimating function (21) given by:

$$E_\theta \tilde{H}_N(\theta) = \frac{1}{2} \Delta^2 n E_{\mu_\theta} \left\{ b_x(\theta) \left[b(\theta) b_x(\theta) / v(\theta) + \frac{1}{2} b_{xx}(\theta) \right] \right\} + O(n \Delta^3)$$

This expression can be obtained by expanding $\tilde{H}_N(\theta)$ in (21):

$$\tilde{H}_N(\theta_0) - \tilde{H}_N(\tilde{\theta}) = \tilde{H}'_N(\tilde{\theta})(\theta_0 - \tilde{\theta}) \quad (27)$$

Since $\tilde{H}_N(\tilde{\theta}) = 0$, we have

$$\begin{aligned} (\theta_0 - \tilde{\theta}) &= \frac{\tilde{H}_N(\theta_0)}{\tilde{H}'_N(\tilde{\theta})} \rightarrow \frac{\Delta E_{\mu_\theta} \left\{ b_x(\theta) \left[b(\theta) b_x(\theta) / v(\theta) + \frac{1}{2} b_{xx}(\theta) \right] \right\}}{2 E_{\mu_\theta} \{ b_x^2(\theta) / v(\theta) \}} \\ &+ O(\Delta^2) \end{aligned}$$

When the quadratic term in (20) is taken into consideration, the bias (when θ is the true parameter value) turns out to be:

$$\begin{aligned} E_\theta H_N(\theta) &= \frac{1}{2} \Delta n E_{\mu_\theta} \left\{ \partial_\theta \log v(\theta) \left[\frac{1}{2} b(\theta) \partial_x \log v(\theta) + \partial_x b(\theta) + \frac{1}{4} \partial_x^2 v(\theta) \right] \right\} + \\ &+ O(n \Delta^2) \end{aligned}$$

The important point to notice above is that the bias of the estimating function is of order $\Delta^2 n$, being therefore considerable even when Δ is small.

Improving the Estimators by Using Better Approximations for the Moments Under certain technical conditions, Kessler (1997) devised ways to reduce the bias described above. He retained the idea of approximating the transition densities by a gaussian distribution, but improved the approximation of the mean and of the variance. In order to follow Kessler's approach to the problem we need one definition.

Definition 4 *Make*

$$r_k(\Delta, x; \theta) = \sum_{i=0}^k \frac{\Delta^i}{i!} L_\theta^i f(x)$$

where $f(x)=x$, and where L_θ^i denotes the i -fold application of the differential operator L_θ .

Using the same ideas as the ones detailed in the previous section, but now dealing with expansions of order k (instead of order 2 only), Kessler obtained new approximations for the mean and for the variance of the transition function Ψ_k (now, with a remainder term of order $O(\Delta^{k+1})$): ⁸:

$$E_\theta(X_\Delta \mid X_0 = x) = r_k(\Delta, x; \theta) + O(\Delta^{k+1}) \quad (28)$$

$$\Psi_k(\Delta, x; \theta) = \sum_{j=0}^k \Delta^j \sum_{r=0}^{k-j} \frac{\Delta^r}{r!} L_\theta^r g_{x,}^j(x) \quad (29)$$

where $g_{x,}^j(y)$, $j = 0, 1, \dots, k$ is defined by the expression:

$$(y - r_k(\Delta, x; \theta))^2 = \sum_{i=0}^k \Delta^i g_{x,}^i(y) + O(\Delta^{k+1})$$

As shown in Sorensen (1995), the new approximation of (18) with (28) and (29) replacing (16) and (17) performs better than the previous one. With these modifications, one gets another score function (which replaces (20)) and other estimators. The estimators so obtained are only slightly biased [Sorensen, 1995], when Δ is not too large. Under additional conditions, Kessler (1997) shows that the new estimators are consistent and asymptotically normal.

⁸Remember that in the previous subsection we made an assumption about the boundedness of the integrand.

Measuring the Loss of Information Due to Discretization Dacunha-

Castelle (1986) assumes the sampling to be equidistant and provides a measure of the amount of information lost by discretization in the nonlinear case. The loss is measured, as a function of Δ , in terms of the asymptotic variance of the MLE estimator of the parameters. Such a procedure allows for a determination of how spaced in time the observations can be without leading to a meaningful problem. The author studies the model (15) [called model E*] and also the particular case when $\sigma(X_t, \theta) = \sigma$ (a constant) [called model E].

The method expresses the transition density of the Markov chain, p_Δ , as a combination of Brownian Bridge functionals⁹. This is achieved through the use of Girsanov's theorem and Itô's formula. The author concludes that, when σ is known, the loss of precision due to discretization is of order Δ^2 , whereas when σ is unknown the loss is of order Δ .

5 Martingale Estimating Functions (MEF)

We start this Section with a Proposition showing that the score function used to derive the MLE in the previous subsections are themselves Martingale estimating functions. The proofs are somewhat standard in the literature.

Proposition 5 *Under regularity conditions (URC), the score function is a Martingale.*

Proof. First we show that the likelihood function is a Martingale and then that the Score Function is a Martingale.

I- The likelihood function is a Martingale:

Let P and Q be two different probability measures on the space $(\Omega, \mathcal{F})$, and let $\mathcal{F}_{n,n \in \mathbb{N}}$ be a filtration defined in this space. For each n , let P_n and Q_n be the restrictions of P and Q to $\mathcal{F}_n$. Suppose Q_n is absolutely continuous with respect to P_n and make Z_n the likelihood function (Radon Nikodym derivative) dQ_n/dP_n . Then, for sets A in the σ -algebra $\mathcal{F}_{n-1}$ we have (because the restriction of P to $\mathcal{F}_n$ and the restriction of P to $\mathcal{F}_{n-1}$ must agree on sets in $\mathcal{F}_{n-1}$):

$$\int_A Z_{n-1} dP = Q(A) = \int_A Z_n dP$$

⁹If $\{X(t), t \geq 0\}$ is a Brownian process, a Brownian Bridge is the stochastic process $\{X(t), 0 \leq t \leq 1 \mid X(1) = 0\}$. It has mean zero and covariance function $\text{Cov}(X(s), X(t) \mid X(1) = 0, s \leq t \leq 1) = s(1-t)$. It can also be represented as $Z(t) = X(t) - tX(1)$ and is very useful in the study of empirical distribution functions.

By the definition of conditional expectation, since A is in $\mathcal{F}_{n-1}$:

$$\int_A E^{\mathcal{F}_{n-1}} Z_n dP = \int_A Z_n dP$$

Since the probability measure is finite, these equalities imply $E^{\mathcal{F}_{n-1}} Z_n = Z_{n-1}$ P -ae.

II- The score function is a Martingale:

Now working with densities defined with respect to the Lebesgue measure, consider the likelihood function $\Lambda_n = \exp(l_n(\theta) - l_n(\theta_0))$. Taking the first derivative with respect to theta yields $d\Lambda_n/d\theta = \Lambda_n dl_n/d\theta$. Assuming the derivative can be passed through the integral:

$$\begin{aligned} E_{\theta_0}^{\mathcal{F}_{n-1}} \Lambda_n dl_n/d\theta &= E_{\theta_0}^{\mathcal{F}_{n-1}} d\Lambda_n/d\theta = (d/d\theta) E_{\theta_0}^{\mathcal{F}_{n-1}} (\Lambda_n) \\ &= (d/d\theta) \Lambda_{n-1} = \Lambda_{n-1} dl_{n-1}/d\theta \end{aligned}$$

The demonstration is concluded by setting $\theta_0 = \theta$. ■

We have seen that the use of Gaussian approximations of the transition function leads to biased estimators. We have also seen the biases of such estimators can be somewhat reduced by the use of better approximations to the mean and variance of the transition density, but not eliminated. Such a problem can be avoided by the use of more general MEF. By more general we mean MEF that are not necessarily based on Gaussian approximations to the transition densities of the diffusion processes.

Trying to mimic the score function, such estimating functions $G_n(\theta)$ are usually of the form:

$$G_n(\theta) = \sum_{i=1}^n g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \quad (30)$$

where the functions $g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta)$ satisfy:

$$\int g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) p(\Delta, x, y; \theta) = 0 \quad (31)$$

Here x stands for $X_{t_{i-1}}$ and y for X_{t_i} . Part of the literature considers functions $g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta)$ as polynomials in y ¹⁰. The approach followed by Kessler

¹⁰This was the case, for instance, of the score function (20). However, we have seen that the approximation given by (20) was biased when the time intervals between observations were bounded away from zero.

and Sorensen does not require the functions $g(\cdot)$ to be polynomials. They are based on the eigenfunctions of the generator of the diffusion process¹¹.

An important property of an estimating function is being unbiased and being able to identify the correct value of the parameter. Formally, if θ_0 stands for the true value of the parameter, one must have:

$$E_\theta G_n(\theta) = 0 \Leftrightarrow \theta = \theta_0$$

$p(\Delta, x, y; \theta)$, the transition density from state x to state y , is usually not known.

Note that (31) means $E_\theta^{\mathcal{F}_{t_{i-1}}} g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) = 0$, implying that $G_n(\theta)$ is a (difference) martingale and, by the law of iterated expectations, $E_\theta G_n(\theta) = 0$. By an analysis following the same first order-expansion used in (27), equations (30) and (31) imply that the estimator $\hat{\theta}$ obtained by making $G_n(\theta) = 0$ is unbiased.

It remains, though, choosing the most adequate MEF according to some optimizing criterion. In order to do so, consider a class of MEF given by making, in (30):

$$g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) = \sum_{j=1}^N \alpha_j(\Delta, x; \theta) h_j(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \quad (32)$$

Since for each j , the event $[\alpha_j(\Delta, x; \theta) \in B, B \text{ borelian in } \mathbb{R}^n] \in \mathcal{F}_{t_{i-1}}$ (where $x = X_{t_{i-1}}$), such functions satisfy (31) if $h_j(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta)$ does. Godambe and Heide (1987) proposed two possible criteria for the choice of the MEF. The first, called *fixed sample criterion*, minimizes the distance to the (usually not explicitly known) score function. The second, called *asymptotic criterion*, chooses the MEF that has the smallest asymptotic variance.

Kessler and Sorensen (1999) provide an analysis of the fixed-sample-criterion type. Under this technique, the estimating function can be viewed as a projection of the score function onto a set of estimating functions of the form (32). Such estimating functions are defined by using the eigenfunctions and eigenvalues of the generator L of the underlying diffusion process (which was the object of our analysis in Section 3). An important part of their analysis is showing that MEF can be so obtained. This is done in their equation 2.4, which we present below as a Proposition.

Proposition 6 *Consider the diffusion process (9). Let $\phi(x; \theta)$ be an eigenfunction and $\lambda(\theta)$ an eigenvalue of the operator L_θ . Then, under weak regu-*

¹¹To get some intuition linking the eigenfunctions to the estimators of the diffusion process, remember (e.g., Karlin and Taylor, 1981) that the transition density of a diffusion process can be expressed as a series expansion using the eigenfunctions.

larity conditions (URC):

$$E_\theta [\phi(y; \theta) \mid X_{t-1} = x] = e^{-\lambda(\theta)\Delta} \phi(x; \theta)$$

for all x in the state space of X under P_θ , implying that

$$g(y, x; \theta) = \alpha(x; \theta) \{ \phi(y; \theta) - e^{-\lambda(\theta)\Delta} \phi(x; \theta) \}$$

is a martingale-difference estimating function.

Proof. Make:

$$Z_t = e^{\lambda t} \phi(X_t) \tag{33}$$

Then, by Itô's formula:

$$dZ_t = e^{\lambda t} \left[\lambda \phi(X_t) dt + \phi'(X_t) dX_t + \frac{1}{2} \phi''(X_t) (dX_t)^2 \right]$$

Taking into consideration that $(dX_t)^2 = \sigma^2 dt$, and using (9) one gets:

$$dZ_t = e^{\lambda t} [(\lambda \phi(X_t) + L\phi(X_t)) dt + \phi'(X_t) \sigma dW_t]$$

Since by assumption $\phi(X_t)$ is an eigenfunction of the operator L with eigenvalue $(-\lambda)$, $L\phi(X_t) + \lambda \phi(X_t) = 0$ and we have:

$$dZ_t = e^{\lambda t} [\phi'(X_t) \sigma(X_t) dW_t]$$

Integrating this expression,

$$Z_t = Z_0 + \int_0^t e^{\lambda s} (\phi'(X_s) \sigma(X_s)) dW_s$$

Since $\int_0^t e^{\lambda s} (\phi'(X_s) \sigma(X_s)) dW_s$ is a martingale, Z_t is a Martingale (for $u < t$, $E^u Z_t = Z_u$). Using this fact in (33) and the definition of Z_t one concludes that $E^{t-1} \phi(X_t; \theta) = e^{-\lambda(\theta)\Delta} \phi(X_{t-1}; \theta) = 0$, as required. ■

Kessler and Sorensen show that the estimators so obtained are, URC, consistent and asymptotically normal (by using the Martingale Central Limit Theorem, (Billingsley, 1961)).

The consistency and asymptotic normality of the estimators derived by Kessler and Sorensen do not require the assumption, as the analysis in Section 4 did, that the time between observations tends to zero. This is an important advantage of such estimators, since $\Delta \rightarrow 0$ is usually not observed by real data.

6 GMM Estimation

Hansen and Scheinkman (1995) show how to generate moment conditions for continuous-time Markov processes with discrete-time sampling. The basic idea pursued by the authors is that such processes can be characterized by means of forward or backward infinitesimal generators (see Section 3). And that when the processes are stationary these generators can be used to derive moment conditions that can be used for estimation purposes by the application of Hansen's (1982) GMM (generalized method of moments).

Note that, by the law of iterated expectations:

$$\int_{R^n} f d\mathcal{Q} = \int_{R^n} \Gamma_t f d\mathcal{Q} \quad (34)$$

Using the framework developed in Section 3, since $\frac{1}{t}(\Gamma_t f - f)$ converges in $L^2(\mathcal{Q})$ to $\Lambda_t f$, (34) implies:

$$\int_{R^n} \Lambda f d\mathcal{Q} = \lim_{t \rightarrow 0} \frac{1}{t} \int_{R^n} (\Gamma_t f - f) d\mathcal{Q} = 0$$

and the (first set of) moments conditions¹²:

$$E[\Lambda f(X_t)] = 0 \text{ for all } f \in \Psi \quad (35)$$

This is a well-known link between the generator and the stationary distribution. A second set of moment conditions is derived by the authors using the reverse-time process:

$$E[\Lambda f(X_{t+1})\check{f}(X_t) - f(X_{t+1})\check{\Lambda}\check{f}(X_t) = 0] \text{ for all } f \in \Psi, \check{f} \in \check{\Psi} \quad (36)$$

where $\check{f}$ and $\check{\Psi}$ are defined as in Section 3, but now with respect to the reverse process. Note that only the second set of moments, by depending on the variables measures in two consecutive points of time, directly captures the Markovian features of the model.

Example 3 *Consider the diffusion process:*

$$dp = -a(p - \mu)dt + \sigma dW_t, \quad p(0) = p_0 > 0, \quad a > 0 \quad (37)$$

¹²Remember the definition of Ψ from Section 3.

Campbell et alli (1997) present a heuristic development of this example. The direct use of the results we proved in Section 3, though, allow for a direct formal approach. Taking f to be the identity function and applying the generator Λ to (37):

$$\Lambda(dp) = -a(p - \mu) \quad (38)$$

Using (35) leads to the first moment condition:

$$Ep = \mu \quad (39)$$

Now, instead of taking $f = I$ (Identity) in the above procedure, take a generic test function f (for instance, $f(\cdot) = (\cdot)^n, n \in N$). Make $Y_t = f(p_t)$ and apply Itô's Lemma:

$$df(p_t) = dY_t = \left[-f'(p)a(p - \mu) + \frac{1}{2}f''(p)\sigma^2 \right] dt + f'(p)\sigma dW$$

Using (35) once more:

$$E \left[-f'(p)a(p - \mu) + \frac{1}{2}f''(p)\sigma^2 \right] = 0 \quad (40)$$

Using (36):

$$E \left\{ \left[-a(p - \mu)\phi'(X_{t+1}) + \frac{1}{2}\sigma^2\phi''(X_t) \right] \check{\phi}(X_t) - \phi(X_{t+1}) \left[-a(p - \mu)\check{\phi}'(X_t) + \frac{1}{2}\sigma^2\check{\phi}''(X_t) \right] \right\} = 0$$

Equations (35) and (36) define an infinite number of moment conditions, depending on the choice of f . Under regularity conditions provided by the authors, GMM can then be applied.

References

- [1] Ait-Sahalia, Y. (2002) “Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach”. *Econometrica*, Vol. 70, N. 1, 223-262.
- [2] Basawa I. V. and Prakasa Rao, B.L.S. (1980): *Statistical Inference for Stochastic Processes*, Academic Press, London.
- [3] Bibby, B.M. and Sorensen, M. (1995): Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1 , 17-39.
- [4] Billingsley, P. (1961): The Lindberg-Levy theorem for martingales. *Proc. of the American Math. Soc.*, 12, 788-792.
- [5] Black, F. and Scholes, M. (1973): The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-654.
- [6] Campbell, J. Y, Lo, A. W. and MacKinlay, A. C. (1997): *The Econometrics of financial markets*. Princeton University Press.
- [7] Cox, J., Ingersoll, J. and Ross, S. (1985): A Theory of the term structure of interest rates. *Econometrica*, 53, 385-408.
- [8] Dacunha-Castelle, D. (1986): Estimation of the coefficients of a diffusion from discrete observations, *Stochastics*, vol. 19, pp. 263-284.
- [9] Dacunha-Castelle, D. and Duflo, M. (1982): *Probability and Statistics*, Vol 2, Springer Verlag, New York.
- [10] Dorogovcev, A. J. (1976): The Consistency of an estimate of a parameter of a stochastic differential equation. *Theor. Probability and Mathematical Statistics*. 10, 73-82.
- [11] Florens-Zmirou, D. (1989): Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20, 547-557.
- [12] Godambe, V. P. and Heyde, C. C. (1987): Quasi likelihood and optimal estimation. *Int. Statist. Rev.* 55, 231-244.
- [13] Gouriéroux, Christian, Joann Jasiak and Christia Gouriéroux (2002). *Financial Econometrics: Problems, Models, and Methods* (2001). Princeton Series in Finance.

- [14] Hansen, L.P. (1982): Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-1054.
- [15] Hansen, L. P. and Scheinkman, J. A. (1995): Back to the future: Generating moment implications for continuous-time markov processes. *Econometrica*, Vol 63, N. 4, 767-804.
- [16] Kessler, M. (1997): Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, Volume 24 Issue 2 Page 211.
- [17] Kessler, M. and Sorensen, M. (1999): Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* 5(2), 1999, 299-314.
- [18] Kloeden P.E. and Platen, E. (1992): *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, New York.
- [19] Kloeden, P.E, Platen, E., Schurz, H. and Sorensen, M. (1992): On the effects of discretization on estimators of drift parameters for diffusion processes. Research Report No. 249, Department of Theoretical Statistics, University of Aarhus.
- [20] Karatzas, I., Shreve, S.E. (1991): *Brownian motion and stochastic calculus*. Second Edition, Spring Verlag.
- [21] Karlin, S. and Taylor, H. M. (1981): *A second course in stochastic process*. Orlando: Academic Press
- [22] Krugman. P. R. (1991): Target zones and exchange rate dynamics. *The Quarterly Journal of Economics*, 106, 669-682.
- [23] Krylov, N.Z. (1980): *Controlled diffusion processes*. Springer Verlag.
- [24] Oksendal, B. (2000): *Stochastic differential equations*, 5th Edition, Springer Verlag.
- [25] Pedersen, A.R. (1995a): A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.* 22, 55-71.
- [26] Pedersen, A.R. (1995b): Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* 1, 257-279.

- [27] Prakasa Rao, B.L.S. (1983): Asymptotic theory of non-linear least squares estimator for diffusion processes. *Math. Operationsforsch. u. Statist.*, Ser.
- [28] Prakasa Rao, B.L.S. (1988): Statistical inference from sampled data for stochastic processes. *Contemporary Mathematics* 80, 249-284.
- [29] Prakasa Rao, B.L.S (1999): Statistical inference for diffusion-type processes. *Kendall's Library of Statistics*, number 8.
- [30] Sorensen, M. (1995): Estimating functions for discretely observed diffusions: A review.
- [31] Sorensen, H (2002). Parametric inference for diffusion processes observed at discrete points in time: a survey. Mimeo, University of Copenhagen.
- [32] Stokey, N. (2000). "Continuous Dynamic Optimization", Mimeo, The University of Chicago.
- [33] Vasicek, O. (1977): An equilibrium characterization of the term structure. *Journal of financial economics*, 5, 177-188.
- [34] Wong, E. (1964): The Construction of a class of stationary markov processes. *Stochastic Processes in Mathematical Physics and Engineering. Proceedings of Symposia in Applied Mathematics*, 16, ed. by R. Bellman. Providence, RI: American Mathematical Society, pp. 264-276.
- [35] Yoshida, N. (1992): Estimation for diffusion processes from discrete observation, *Journal of Multivariate Analysis*, 41, 220-242.

ENSAIOS ECONÔMICOS DA EPGE

490. IDADE, INCAPACIDADE E A INFLAÇÃO DO NÚMERO DE PESSOAS COM DEFICIÊNCIA - Marcelo Neri ; Wagner Soares – Julho de 2003 – 54 págs.
491. FORECASTING ELECTRICITY LOAD DEMAND: ANALYSIS OF THE 2001 RATIONING PERIOD IN BRAZIL - Leonardo Rocha Souza; Lacir Jorge Soares – Julho de 2003 – 27 págs.
492. THE MISSING LINK: USING THE NBER RECESSION INDICATOR TO CONSTRUCT COINCIDENT AND LEADING INDICES OF ECONOMIC ACTIVITY - JoãoVictor Issler; Farshid Vahid – Agosto de 2003 – 26 págs.
493. REAL EXCHANGE RATE MISALIGNMENTS - Maria Cristina T. Terra; Frederico Estrella Carneiro Valladares – Agosto de 2003 – 26 págs.
494. ELASTICITY OF SUBSTITUTION BETWEEN CAPITAL AND LABOR: A PANEL DATA APPROACH - Samuel de Abreu Pessoa ; Silvia Matos Pessoa; Rafael Rob – Agosto de 2003 – 30 págs.
495. A EXPERIÊNCIA DE CRESCIMENTO DAS ECONOMIAS DE MERCADO NOS ÚLTIMOS 40 ANOS – Samuel de Abreu Pessoa – Agosto de 2003 – 22 págs.
496. NORMALITY UNDER UNCERTAINTY – Carlos Eugênio E. da Costa – Setembro de 2003 – 08 págs.
497. RISK SHARING AND THE HOUSEHOLD COLLECTIVE MODEL - Carlos Eugênio E. da Costa – Setembro de 2003 – 15 págs.
498. REDISTRIBUTION WITH UNOBSERVED 'EX-ANTE' CHOICES - Carlos Eugênio E. da Costa – Setembro de 2003 – 30 págs.
499. OPTIMAL TAXATION WITH GRADUAL LEARNING OF TYPES - Carlos Eugênio E. da Costa – Setembro de 2003 – 26 págs.
500. AVALIANDO PESQUISADORES E DEPARTAMENTOS DE ECONOMIA NO BRASIL A PARTIR DE CITAÇÕES INTERNACIONAIS - João Victor Issler; Rachel Couto Ferreira – Setembro de 2003 – 29 págs.
501. A FAMILY OF AUTOREGRESSIVE CONDITIONAL DURATION MODELS - Marcelo Fernandes; Joachim Grammig – Setembro de 2003 – 37 págs.
502. NONPARAMETRIC SPECIFICATION TESTS FOR CONDITIONAL DURATION MODELS - Marcelo Fernandes; Joachim Grammig – Setembro de 2003 – 42 págs.
503. A NOTE ON CHAMBERS'S "LONG MEMORY AND AGGREGATION IN MACROECONOMIC TIME SERIES" – Leonardo Rocha Souza – Setembro de 2003 – 11págs.
504. ON CHOICE OF TECHNIQUE IN THE ROBINSON-SOLOW-SRINIVASAN MODEL - M. Ali Khan – Setembro de 2003 – 34 págs.
505. ENDOGENOUS TIME-DEPENDENT RULES AND THE COSTS OF DISINFLATION WITH IMPERFECT CREDIBILITY - Marco Bonomo; Carlos Viana de Carvalho – Outubro de 2003 – 27 págs.

506. CAPITAIS INTERNACIONAIS: COMPLEMENTARES OU SUBSTITUTOS? - Carlos Hamilton V. Araújo; Renato G. Flôres Jr. – Outubro de 2003 – 24 págs.
507. TESTING PRODUCTION FUNCTIONS USED IN EMPIRICAL GROWTH STUDIES - Pedro Cavalcanti Ferreira; João Victor Issler; Samuel de Abreu Pessoa – Outubro de 2003 – 8 págs.
508. SHOULD EDUCATIONAL POLICIES BE REGRESSIVE ? Daniel Gottlieb; Humberto Moreira – Outubro de 2003 – 25 págs.
509. TRADE AND CO-OPERATION IN THE EU-MERCOSUL FREE TRADE AGREEMENT - Renato G. Flôres Jr. – Outubro de 2003 – 33 págs.
510. OUTPUT CONVERGENCE IN MERCOSUR: MULTIVARIATE TIME SERIES EVIDENCE - Mariam Camarero; Renato G. Flôres Jr; Cecílio Tamarit – Outubro de 2003 – 36 págs.
511. ENDOGENOUS COLLATERAL - Aloísio Araújo; José Fajardo Barbachan; Mario R. Páscoa – Novembro de 2003 – 37 págs.
512. NON-MONOTONE INSURANCE CONTRACTS AND THEIR EMPIRICAL CONSEQUENCES - Aloísio Araujo; Humberto Moreira – Novembro de 2003 – 31 págs.
513. EQUILIBRIA IN SECURITY MARKETS WITH A CONTINUUM OF AGENTS - A. Araujo; V. F. Martins da Rocha; P. K. Monteiro – Novembro de 2003 – 17 págs.
514. SPECULATIVE ATTACKS ON DEBTS AND OPTIMUM CURRENCY AREA: A WELFARE ANALYSIS - Aloisio Araujo; Márcia Leon – Novembro de 2003 – 50 págs.
515. O MÉTODO GENERALIZADO DOS MOMENTOS(MGM): CONCEITOS BÁSICOS - Renato G. Flôres Jr – Novembro de 2003 – 27 págs.
516. VARIÁVEIS INSTRUMENTAIS E O MGM: USO DE MOMENTOS CONDICIONAIS - Renato G. Flôres Jr – Novembro de 2003 – 27 págs.
517. O VALOR DA MOEDA E A TEORIA DOS PREÇOS DOS ATIVOS - Fernando de Holanda Barbosa – Dezembro de 2003 – 17 págs.
518. EMPRESÁRIOS NÁUTICOS, GARANTIAS E ACESSO À CRÉDITO - Marcelo Côrtes Néri; Fabiano da Silva Giovanini - Dezembro de 2003 – 23 págs.
519. DESENHO DE UM SISTEMA DE METAS SOCIAIS - Marcelo Côrtes Néri; Marcelo Xerez - Dezembro de 2003 – 24 págs.
520. A NEW INCIDENCE ANALYSIS OF BRAZILIAN SOCIAL POLICIES USING MULTIPLE DATA SOURCES - Marcelo Côrtes Néri - Dezembro de 2003 – 55 págs.
521. AN INTRA-HOUSEHOLD APPROACH TO THE WELFARE COSTS OF INFLATION - Rubens Penha Cysne – Janeiro de 2004 – 16 págs.
522. CENTRAL LIMIT THEOREM FOR ASYMMETRIC KERNEL FUNCTIONALS - Marcelo Fernandes; Paulo Klinger Monteiro – Fevereiro de 2004 – 23 págs.
523. THE TRADE-OFF BETWEEN INCENTIVES AND ENDOGENOUS RISK - Aloísio Araujo; Humberto Moreira; Marcos H. Tsuchida – Fevereiro de 2004 – 21 págs.

524. DO DIVIDENDS SIGNAL MORE EARNINGS ? - Aloísio Araujo; Humberto Moreira; Marcos H. Tsuchida – Fevereiro de 2004 – 26 págs.
525. Biased managers, organizational design, and incentive provision - Cristiano M. Costa; Daniel Ferreira; Humberto Moreira – Fevereiro de 2004 – 11 págs.
526. Land taxes in a Latin American context - Juliano J. Assunção; Humberto Moreira – Fevereiro de 2004 - 19 págs.
527. Indicadores coincidentes de atividade econômica e uma cronologia de recessões para o Brasil - Angelo J. Mont'alverne Duarte; João Victor Issler; Andrei Spacov - Fevereiro de 2004 – 41 págs.
528. TESTING UNIT ROOT BASED ON PARTIALLY ADAPTIVE ESTIMATION - Zhijie Xiao; Luiz Renato Lima – Março de 2004 – 27 págs.
529. DO SHOCKS PERMANENTLY CHANGE OUTPUT? LOCAL PERSISTENCY IN ECONOMIC TIME SERIES - Luiz Renato Lima; Zhijie Xiao – Março de 2004 – 21 págs.
530. A NEW PERSPECTIVE ON THE PPP HYPOTHESIS - Soyoung Kim; Luiz Renato Lima – Março de 2004 – 36 págs.
531. TRADE LIBERALIZATION AND INDUSTRIAL CONCENTRATION: EVIDENCE FROM BRAZIL - Pedro Cavalcanti Ferreira; Giovanni Facchini – Março de 2004 - 25 págs.
532. REGIONAL OR EDUCATIONAL DISPARITIES? A COUNTERFACTUAL EXERCISE - Angelo José Mont'Alverne; Pedro Cavalcanti Ferreira; Márcio Antônio Salvato – Março de 2004 – 25 págs.
533. INFLAÇÃO: INÉRCIA E DÉFICIT PÚBLICO – Fernando de Holanda Barbosa – Março de 2004 – 16 págs.
534. A INÉRCIA DA TAXA DE JUROS NA POLÍTICA MONETÁRIA – Fernando de Holanda Barbosa – Março de 2004 – 13 págs.
535. DEBT COMPOSITION AND EXCHANGE RATE BALANCE SHEET EFFECTS IN BRAZIL: A FIRM LEVEL ANALYSIS - Marco Bonomo; Betina Martins ; Rodrigo Pinto – Março de 2004 – 39 págs.
536. THE SET OF EQUILIBRIA OF FIRST-PRICE AUCTIONS - Paulo Klinger Monteiro – Março de 2004 – 6 págs.
537. OPTIMAL AUCTIONS WITH MULTIDIMENSIONAL TYPES AND THE DESIRABILITY OF EXCLUSION - Paulo Klinger Monteiro ; Benar Fux Svaiter; Frank H. Page Jr – Março de 2004 – 8 págs.
538. INCOME INEQUALITY IN A JOB-SEARCH MODEL WITH HETEROGENEOUS TIME PREFERENCES – Rubens Penha Cysne – Março de 2004 – 35 págs.
539. IMPOSTO INFLACIONÁRIO E TRANSFERÊNCIAS INFLACIONÁRIAS NO BRASIL: 1947-2003 - Rubens Penha Cysne; Paulo C. Coimbra Lisboa – Março de 2004 – 7 págs.
540. ON THE STATISTICAL ESTIMATION OF DIFFUSION PROCESSES – A survey - Rubens Penha Cysne – Abril de 2004 – 26 págs.