On the Statistical Estimation of Diffusion Processes - A Survey*

Rubens Penha Cysne

December 17, 2003

Abstract

This survey covers a (small) sample of the literature covering the parametric inference of diffusion processes using discretely-sampled data. We concentrate on some recent contributions to the literature based on three different approaches to the problem: approximated likelihood methods, Martingale estimating functions and 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.

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

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

*Preliminary version.

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 $\sigma-$algebra generated by $\{W_u - W_v, t \geq u \geq v \geq 0\}$. Denote by $\mathcal{F}_{0,t}$ the $\sigma-$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 $h : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ is $\mathcal{F}_t-$measurable and σ is a $(d \times m)$ $\mathcal{F}_t-$measurable matrix with $\sigma_{ij} : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$. Under a set of Lipschitz conditions (see, e.g., Prakasa Rao (1999) or Oksendal (2000, section 5.2)), the equation:

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

has a unique and continuous solution $X_t(t, \omega)$ and, for each $t \geq 0$, X_t is $(\mathcal{F}_t \mid Borel)-$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 $Var(X(t) - X(s)) = \sigma^2 |s - t|$.

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 difficulty with the ideal method, different alternatives have been proposed in the literature. One alternative is calculating the estimators departing from some approximation of the continuous-time likelihood function obtained using simulation methods and then define the estimators based on this function (Pedersen, 1995a, 1995b). Another alternative is based on

Gaussian approximations to the transition density. This strategy, however, usually leads to biased estimators. The bias tends to increase when the time between observations gets larger. A third approach uses general martingale estimating functions and, a fourth, GMM estimation. We will concentrate here solely on the three last methodologies.

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. We consider below some examples of applications in these areas.

- Black and Scholes (1973). We present here the version of Campbell et alli (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:

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

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

$$dP(t) = \mu P(t)dt + \sigma P(t)dW(t), \qquad t \geq 0 \tag{2}$$

$W(t)$ 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 $P(.)$) that an initial investment I is allocated in options and stocks according to

$$I = G(P, t) + \alpha P \tag{3}$$

Using Itô's Lemma:

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

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:

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

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

$$G(P(T), T) = \max(0, P(T) - X) \tag{5}$$

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 $E(ds/dt)$:

  $$s = m + v + \gamma E(ds/dt) \tag{6}$$

  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:

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

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 $\mathrm{A}_1 = -A_2$ 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 $f(.)$ be a bounded twice continuously differentiable function, with bounded derivatives and X_t a generic time-homogeneous diffusion process defined on the probability space $(\Psi, \mathcal{F}, P)$. Let $\mathcal{Q}$ be the probability measure induced by X_t on R^n and $L^2(\mathcal{Q})$ be the space of Borel measurable functions $f(X_t) : R^n \to R$ $\mathcal{Q}$ -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:

$$\Gamma_t f(x_0) = E\left[f(X_t) \mid X_0 = x_0\right] \quad (\equiv E^0(f(X_t))) \tag{7}$$

It can be shown that these operators ($L^2(\mathcal{Q}) \to L^2(\mathcal{Q})$) 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 os iterated expectations, $\mathrm{E}^0(X_{t+s}) = E^0(E^t(X_{t+s}))$, implying $\Gamma_{t+s} = \Gamma_t \Gamma_s$).

In the remaining of this text, we will some times refer to the generator of a diffusion process $f(X_t)$. 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:

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

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.

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

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \tag{9}$$

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

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

Then $\Lambda f = L_\theta f$.

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:

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

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 :

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

V- Using Fubini's theorem (by assumption, the integrand is quasi-integrable w.r.t the product measure $ds \times dP$) 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 (in $\mathbb{L}^2$) $L_\theta f$. We need to show:

$$\int_{R^n} \left\{ [(1/t)\int_0^t (\Gamma_s L_\theta f(x_0) - L_\theta f(x_0))ds]^2 \right\} dx_0 \to 0$$

Using the Cauchy-Schwarz (Hölder) inequality:

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

Using Fubini again,

$$\begin{aligned} & (1/t) \int_{R^n} \left\{ \int_0^t [\Gamma_s L_\theta f(x_0) - L_\theta f(x_0)]^2 ds \right\} d\mathcal{Q}(x_0) \\ = & \ (1/t) \int_0^t \left\{ \int_{R^n} [\Gamma_s L_\theta f(x_0) - L_\theta f(x_0)]^2 dx_0 \right\} ds \\ = & \ (1/t) \int_0^t \left\| \Gamma_s L_\theta f(x_0) - L_\theta f(x_0) \right\|^2 ds \end{aligned}$$

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 $\phi \in L^2(\mathcal{Q})$, $\{\Gamma_t \phi,\ t \geq 0\}$ converges (in $\mathbb{L}^2(\mathcal{Q})$) to ϕ as $t \downarrow 0$.) ■

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

4 Maximum Likelihood Estimation (MLE)

4.1 Continuosly Observed Data

Likelihood methods for continuously observed diffusions are standard in the literature (Prakasa Rao, 1999; Basawa and Prakasa Rao, 1980). Consider the diffusion process (1). Note that the function $\sigma(s, X_t)$ does 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 sample path of the process by noticing that:

$$\sum_{i=1}^{2^n} \left[X_{t\wedge k/2^n} - X_{t\wedge (k-1)/2^n}\right]^2 \to \sigma^2 T \text{ a.s. as n} \to \infty$$

Therefore, we can 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 (functional) space $(C\,[0,T], \mathcal{B}_T)$, $\mathcal{B}$ corresponding to the Borel sigma-algebra defined in $C\,[0,T]$. By this we mean:

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

$P_\theta^T(B)$ 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 $C\,[0,T]$:

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

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 $\hat{\theta}_T : ((C\,[0,T], \mathcal{B}_T) \to (\Theta, \tau)$, such that:

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

where τ is the $\sigma-$algebra of Borel subsets of Θ.

Example 2 *Making* $h(\theta, s, X_t) = \theta$, *the above equation leads to the maximization of* $f(\theta) = \theta X_T - \theta T$, *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 cases when the stochastic differential equation (1) is easily solvable, and the corresponding transition functions known: i) $h(\theta, t, X_t) = \mu X_t$, $\sigma(t, X_t) = \sigma X_t$, called geometric brownian motion; ii) $h(\theta, t, X_t) = \alpha(\beta - X_t)$, $\sigma(t, X_t) = \sigma$, the Orstein-Uhlenbeck process and; iii) $h(\theta, t, X_t) = \mu(\beta - X_t)$, $\sigma(t, X_t) = \sigma\sqrt{X_t}$. 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 model[2], which falls into the first case considered above:

$$dP(t) = \mu P(t)dt + \sigma P(t)dW(t), \qquad t \geq 0 \tag{11}$$

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,

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

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

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 $(\mu - \frac{1}{2}\sigma^2)$ and variance $\sigma^2 t$:

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

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

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

4.2.2 The Case when the Transition Densities are not Known

The diffusion for the prices of stocks described above is usually not supported by the data. There is considerable evidence that the increments of Y_t, the logarithm of the price of the stock used to price the option, are neither independent nor Gaussian, as implied by equation (2). This leads to the necessity of more complex models in which the straight application of likelihood is usually not possible because the transition densities are not known.

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

$$p(X_0, X_1, ..., X_n) = p_0(X_0, \theta)\Pi_{j=1}^{n} p_k(X_{t_j}, t_j \mid X_{t_{j-1}}, t_{j-1}; \theta) \qquad (12)$$

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

represent the transition density functions. As we mentioned before, 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. Sorensen (1995) is the main source of our analysis in this Section.

When the distance between observations is sufficiently small, such approximations lead to reasonable estimators, although biased.

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

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t, \quad X_0 = x_0 \qquad (13)$$

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 of $\mathbb{R}^d$. We discretize this process by assuming that the drift and the diffusion are constant in the time interval[3] $\Delta_i = t_i - t_{i-1}$:

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

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:

$$E_\theta(X_{t_i} \mid X_{t_{i-1}}) = b(X_{t_{i-1}}, \theta)\Delta_i + X_{t_{i-1}} \qquad (14)$$

[3] For equidistant intervals, such approximation, usually called an Euler-Maruyama aproximation, can be written

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

$$E_\theta\left[(X_{t_i} - E_\theta(X_{t_i} \mid X_{t_{i-1}}))^2 \mid X_{t_{i-1}}\right] = \sigma^2(X_{t_{i-1}}, \theta)\Delta_i \tag{15}$$

Note that there are two types of approximation going on 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:

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

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

$$L_n(\theta) = \prod_{i=1}^{n} p(X_{t_i} \mid X_{t_{i-1}}) p(X_{t_0}) \tag{16}$$

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

$$l_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N}\left\{\frac{(X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}}, \theta)\Delta_i)^2}{\sigma^2(X_{t_{i-1}}, \theta)\Delta_i} + \log(2\pi\sigma^2(X_{t_{i-1}}, \theta)\Delta_i\right\}] + \log p(X_{t_0}) \tag{17}$$

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

$$H_N(\theta) = \sum_{i=1}^{N}\left\{\begin{array}{c} \frac{b_\theta(X_{t_{i-1}},\theta)}{\sigma^2(X_{t_{i-1}},\theta)}\left[(X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}}, \theta)\Delta_i)\right] + \\ + \frac{\sigma^2_\theta(X_{t_{i-1}},\theta)}{2(\sigma^2(X_{t_{i-1}},\theta))^2\Delta_i}\left[(X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}}, \theta)\Delta_i)^2 - \sigma^2(X_{t_{i-1}}, \theta)\Delta_i\right] \end{array}\right\} \tag{18}$$

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 (18), $\sigma^2(X_{t_{i-1}}, \theta) = v(X_{t_{i-1}}, \theta)$ 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:

$$\tilde{H}_N(\theta) = \sum_{i=1}^{N}\left\{\frac{b_\theta(X_{t_{i-1}};\theta)}{v(X_{t_{i-1}};\theta)}\left[(X_{(t_i)} - X_{(t_{i-1})} - b(X_{t_{i-1}};\theta)\Delta_i\right]\right\} \tag{19}$$

This estimating function has been studied by Dorogovicev (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, theses authors showed 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 (17). Florens-Zmirou imposed $n\Delta_n^2 \to 0$, and Yoshida the less restrictive assumption $n\Delta_n^3 \to 0$.

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) show that if Δ is not small the bias can be severe[4].

Improving the Approximations for the Moments Lemma 1 in Florens-Zmirou[5] (1989) provides an expansion of $E_\theta(X_\Delta \mid X_0 = x)$ which can be used to improve (14) and (15) 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:

Lemma 2 *(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 $\sigma-$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 $k-1$:*

$$E^{k-1}f(X_{k\Delta}) = \sum_{l=0}^{s} \frac{\Delta^l}{l!} L^l f(X_{(k-1)\Delta}) + \int_0^\Delta \int_0^{u_1} \ldots \int_0^{u_s} E^{k-1}(L^{s+1}f)(X_{(k-1)\Delta+u_s+1}) du_1 \ldots du_{s+1} \tag{20}$$

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 (19).

[4] Estimators (18) and (19) 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.

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

The new equations for the conditional average and variance in (14) and (15) (writing b for $b(x;\theta)$ and v for $v(x;\theta)$) read:

$$E_\theta(X_\Delta \mid X_0 = x) = x + \Delta b + \frac{1}{2}\Delta^2 \left\{ bb_x + \frac{1}{2} v b_{xx} \right\} + O(\Delta^3) \qquad (21)$$

and :

$$Var_\theta(X_\Delta \mid X_0 = x) = v\Delta + \Delta^2 \left[\frac{1}{2} b v_x + v \left\{ b_x + \frac{1}{4} v_{xx} \right\} \right] + O(\Delta^3) \qquad (22)$$

Note that (14) and (15) are a particular case of these expressions, for the cases when $l = 1$ in (20).

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

$$\begin{aligned} E_\theta(X_\Delta \quad | \quad X_0 = x) &= x + \Delta L x + \frac{\Delta^2}{2} L^2 x + \\ &+ \int_0^\Delta \int_0^{u_1} \int_0^{u2} E(L^3 x)(X_z) du_1 du_2 du_3 \end{aligned}$$

X_z standing for $X_{(k-1)\Delta + u_s + 1}$ in Lemma 1. Since $Lx = b(x, \theta)$, $L^2 x = Lb(x, \theta) = bb_x + \frac{1}{2} v b_{xx}$ we get (21). The remaining $O(\Delta^3)$ derives from the fact that the (absolute value of the) integrand in (20) 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,

$$\int_0^\Delta \int_0^{u_1} \int_0^{u2} \left| E(L^{s+1} x)(X_z) du_1 du_2 du_3 \right| \leq M \Delta^3$$

To get (22) we need $E_\theta(X_\Delta^2 \mid X_0 = x)$. Again, using (20):

$$\begin{aligned} E_\theta(X_\Delta^2 \quad | \quad X_0 = x) &= x^2 + \Delta L x^2 + \frac{\Delta^2}{2} L^2 x^2 + \\ &+ \int_0^\Delta \int_0^{u_1} \int_0^{u2} E(L^{s+1} x^2)(X_z^2) du_1 ... du_3 \end{aligned}$$

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

$$\begin{aligned} E_\theta(X_\Delta^2 \quad | \quad X_0 = x) &= x^2 + \Delta(2bx + v) + \frac{\Delta^2}{2}(b(2b + 2x b_x + v_x) \qquad (23) \\ &+ \frac{1}{2} v (2b_x + 2b_x + 2x b_{xx} + v_{xx})) \end{aligned}$$

From (21) we get:

$$[E_\theta(X_\Delta \mid X_0 = x)]^2 = x^2 + \Delta 2bx + \frac{\Delta^2}{2}(2b^2 + 2xbb_x + xvb_{xx}) + O(\Delta^3) \quad (24)$$

By subtracting (24) from (23) one gets (22).

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 (19) and (21) imply a bias of the estimating function (19) 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 (19):

$$\tilde{H}_N(\theta_0) - \tilde{H}_N(\tilde{\theta}) = \tilde{H}'_N(\tilde{\theta})(\theta_0 - \tilde{\theta}) \quad (25)$$

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

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

When the quadratic term in (18) is taken into consideration, the bias turns out to be:

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

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 3 *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 $O(\Delta^{k+1})$): [6]:

$$E_\theta(X_\Delta \mid X_0 = x) = r_k(\Delta, x; \theta) + O(\Delta^{k+1}) \tag{26}$$

$$\Psi_k(_k(\Delta, x; \theta) = \sum_{j=0}^{k} \Delta^j \sum_{r=0}^{k-j} \frac{\Delta^r}{r!} L_\theta^r g_{x,}^j (x) \tag{27}$$

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

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

As shown in Sorensen (1995), the new approximation of (16) with (26) and (27) replacing (14) and (15) performs better than the previous one. With these modifications, one gets another score function (which replaces (18)) 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.

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

[6] Remember that in the previous subsection we made an assumption about the boundedness of the integrand.

can be without leading to a meaningful problem. The author studies the model (13) [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 functionals[7]. 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.

Proposition 4 *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 $\mathcal{F}_{n,n\in N}$ 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 $\sigma-$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 $\mathcal{F}_{n-1}$):

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

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

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

[7]If $\{X(t),\ t \geq 0\}$ 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 $\mathrm{Cov}(X(s), X(t) \mid X(1) = 0, s \leq t \leq 1) = s(1-t)$. It can also be represented as $Z(t) = X(t) - tX(1)$ and is very useful in the study of empirical distribution functions.

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:

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \tag{28}$$

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

$$\int g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta)\ p(\Delta, x, y; \theta) = 0 \tag{29}$$

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[8]. The approach followed by Kessler and Sorensen does not require the functions $g(.)$ to be polynomials. They are based on the eigenfunctions of the generator of the diffusion process[9].

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:

[8] This was the case, for instance, of the score function (18). However, we have seen that the approximation given by (18) was biased when the time intervals between observations were bounded away from zero.

[9] To 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.

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

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

Note that (29) means $E_\theta^{\mathcal{F}_{t_{i-1}}} g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) = 0$, implying that $\mathrm{G}_n(\theta)$ 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 (25), equations (28) and (29) 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 (28):

$$g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) = \sum_{j=1}^{N} \alpha_j(\Delta, x; \theta) h_j(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \tag{30}$$

Since for each j, $\alpha_j(\Delta, x; \theta) \in \mathcal{F}_{t_{i-1}} \left[x = X_{t_{i-1}} \right]$, such functions satisfy (29) 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 (30). Such estimating functions are defined by using the eigenfuntions 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 5 *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 regularity conditions (URC):*

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

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

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

is a martingale-difference estimating function.

Proof. Make:

$$Z_t = e^{\lambda t}\phi(X_t)) \tag{31}$$

Then, by Itô's formula:

$$dZ_t = e^{\lambda t}\left[\lambda\phi(X_t)dt + \phi\prime(X_t)dX_t + \frac{1}{2}\phi\prime\prime(X_t)(dX_t)^2\right]$$

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

$$dZ_t = e^{\lambda t}\left[(\lambda\phi(X_t) + L\phi(X_t)dt + \phi\prime(X_t)\sigma dW_t\right]$$

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:

$$dZ_t = e^{\lambda t}\left[\phi\prime(X_t)\sigma(X_t)dW_t\right]$$

Integrating this expression,

$$Z_t = Z_0 + \int_0^t e^{\lambda s}(\phi\prime(X_s)\sigma(X_s))dW_s$$

Since $\int_0^t e^{\lambda s}(\phi\prime(X_s)\sigma(X_s))dW_s$ is a martingale, Z_t is a Martingale (for $u < t$, $E^u Z_t = Z_u$).Using this fact in (31) 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 $\Delta \to 0$ 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:

$$\int_{R^n} f d\mathcal{Q} = \int_{R^n} \Gamma_t f d\mathcal{Q} \tag{32}$$

Using the framework developed in Section 3, since $\frac{1}{t}(\Gamma_t f - f)$ converges in $\mathrm{L}^2(\mathcal{Q})$ to $\Lambda_t f$, (32) implies:

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

and the moments conditions:

$$E\left[\Lambda_t f(X_t\right] = 0 \text{ for all } f \in \Psi \tag{33}$$

This is a well-known link between the generator and the stationary distribution.

Example 3 *Consider the diffusion process:*

$$dp = -a(p - \mu)dt + \sigma dW_t, \;\; p(0) = p_0 > 0, \;\; a > 0 \tag{34}$$

Taking f to be the identity function and applying the generator Λ to (34):

$$\Lambda(dp) = -a(p - \mu) \tag{35}$$

Using (33) leads to the first moment condition:

$$Ep = \mu \tag{36}$$

–Now, instead of taking $f = I$ (Identity) in the above procedure, take a generic test function f (for instance, $f(.) = (.)^n, n \in N$). 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 (33) once more:

$$E\left[-f'(p)a(p - \mu) + \frac{1}{2} f''(p)\sigma^2 \right] = 0 \tag{37}$$

Equation (33) defines an infinite number of moment conditions, depending on the choice of f.

Another set of moments is derived by Hansen and Scheinkman (1995) using the properties of the reverse-time diffusion. Under regularity conditions provided by the authors, GMM can then be applied.

References

[1] Basawa I. V. and Prakasa Rao, B.L.S. (1980): Statistical Inference for Stochastic Processes, Academic Press, London.

[2] Bibby, B.M. and Sorensen, M. (1995): Martingale estimation functions for discretely observed diffusion processes. Bernoulli, 1 , 17-39.

[3] Billingsley, P. (1961): The Lindberg-Levy theorem for martingales. Proc. of the American Math. Soc., 12, 788-792.

[4] Black, F. and Scholes, M. (1973): The pricing of options and corporate liabilities, Journal of Political Economy, 81, 637-654.

[5] Campbell, J. Y, Lo, A. W. and MacKinlay, A. C. (1997): The Econometrics of financial markets. Princeton University Press.

[6] Cox, J., Ingersoll, J. and Ross, S. (1985): A Theory of the term structure of interest rates. Econometrica, 53, 385-408.

[7] Dacunha-Castelle, D. (1986): Estimation of the coefficients of a diffusion from discrete observations, Stochastics, vol. 19, pp. 263-284.

[8] Dacunha-Castelle, D. and Duflo, M. (1982): Probability and Satistics, Vol 2, Springer Verlag, New York.

[9] 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.

[10] Florens-Zmirou, D. (1989): Approximate discrete-time schemes for statistics of diffusion processes. Statistics, 20, 547-557.

[11] Godambe, V. P. and Heyde, C. C. (1987): Quasi likelihood and optimal estimation. Int. Statist. Rev. 55, 231-244.

[12] Hansen, L.P. (1982): Large sample properties of generalized method of moments estimators, Econometrica 50, 1029-1054.

[13] 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.

[14] Kessler, M. (1997): Estimation of an ergodic diffusion from discrete observations. Scandinavian Journal of Statistics, Volume 24 Issue 2 Page 211.

[15] Kessler, M. and Sorensen, M. (1999): Estimating equations based on eigenfunctions for a discretely observed diffusion process. Bernoulli **5**(2), 1999, 299-314.

[16] Kloeden P.E. and Platen, E. (1992): Numerical Solution of Stochastic Differential Equations. Springer-Verlag, New York.

[17] 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.

[18] Karatzas, I., Shreve, S.E. (1991): Brownian motion and stochastic calculus. Second Edition, Spring Verlag.

[19] Karlin, S. and Taylor, H. M. (1981): A second course in stochastic process. Orlando: Academic Press

[20] Krugman. P. R. (1991): Target zones and exchange rate dynamics. The Quarterly Journal of Economics, 106, 669-682.

[21] Krylov, N.Z. (1980): Controlled diffusion processes. Springer Verlag.

[22] Oksendal, B. (2000): Stochastic differential equations, 5th Edition, Springer Verlag.

[23] 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.

[24] Pedersen, A.R. (1995b): Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. Bernoulli 1, 257-279.

[25] Prakasa Rao, B.L.S. (1983): Asymptotic theory of non-linear least squares estimator for diffusion processes. Math. Operationsforsch. u. Statist., Ser.

[26] Prakasa Rao, B.L.S. (1988): Statistical inference from sampled data for stochastic processes. Contemporary Mathematics 80, 249-284.

[27] Prakasa Rao, B.L.S (1999): Statistical inference for diffusion-type processes. Kendall's Library of Statistics, number 8.

[28] Sorensen, M. (1995): Estimating functions for discretely observed diffusions: A review.

[29] Sorensen, H (2002). "Parametric inference for diffusion processes observed at discrete points in time: a survey." Mimeo, University of Copenhagen.

[30] Yoshida, N. (1992): Estimation for diffusion processes from discrete observation, Journal of Multivariate Analysis, 41, 220-242.