FUNDAÇÃO
GETULIO VARGAS

EPGE
Escola de Pós-Graduação
em Economia

Ensaios Econômicos

Escola de

Pós Graduação

em Economia

da Fundação

Getulio Vargas

*On the Statistical Estimation of Diffusion
Processes – A Partial Survey (Revised
Version, Forthcoming Brazilian Review of
Econometrics)*

Rubens Penha Cysne

Outubro de 2004

On the Statistical Estimation of Diffusion Processes - A Partial Survey (Revised Version, Forthcoming, Brazilian Review of Econometrics).

Rubens Penha Cysne[*]

October 16, 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 employment of Martingale Estimating Functions, and the application of Generalized Method of Moments (GMM).

1   Introduction

A large number of models in economics and finance describe the time evolution of dynamic phenomena in a continuous-time stochastic framework. Interest-rate models, for instance, are nowadays frequently formulated in terms of nonlinear stochastic differential equations. This implies the need to estimate the parameters of such models. However, in practice, the data used for such inference is always of a discrete nature, sampled at discrete intervals of time. This leads us to a very specific statistical problem, that

---
[*]Professor at the Graduate School of Economics (EPGE/FGV) of the Getulio Vargas Foundation and, in 2004, a Visiting Scholar at the Department of Economics of the University of Chicago. E mail: rpcysne@uchicago.edu.

has been a subject of active academic research for many years, dating back to the seminal work of Phillips (1959).

The aim of this paper is to provide a partial survey of some of the techniques used in the statistical inference of diffusions. The qualification "partial" here is used to alert the reader that we review only a fraction of the many techniques that have been devised to deal with the problem: an improvement of the Euler-Maruyama discretization scheme, the employment of Martingale Estimating Functions, and the application of the Generalized Method of Moments (GMM).

The Euler-Maruyama approach employs a discrete-time approximation to the continuous system. The estimation of the discrete-time model is then accomplished by maximum likelihood. Martingale Estimating Functions, the second technique studied here, and whose prime example is the score function (the gradient of the likelihood function with respect to the parameters of interest), represent a particular type of estimating functions, distinguished by the nice property of allowing the use of all the available machinery of Martingale Theory. The utilization of this machinery is particularly helpful in the derivation of large-sample properties of the estimators, in which case the Martingale Central Limit Theorem [Billingsley (1961)] can be used.

Finally, a GMM estimator [Hansen (1982)] is a vector that minimizes a distance function, properly defined, of the sample moments from zero. Approximation results that justify the use of GMM in the estimation of diffusion processes are found in Hansen and Scheinkman (1995).

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 process with continuous sample paths which can be characterized by an infinitesimal generator (which we are going to define below). The simplest diffusion process is the Wiener process, the stochastic process that corresponds to the Brownian Motion.

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 $\mathbb{R}^d$:

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

$\theta \in \mathbb{R}^k$ denoting parameter of the model (we are considering that σ does not depend on θ). 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 $\{W_u, \ t \geq u \geq 0\}$. 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 $h : \mathbb{R}^k \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ 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:

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

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 tth generation, s_t for a random or fixed parameter at the tth 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 be obtained in closed-form only 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.

As pointed out before, the first two techniques on which we shall concentrate here, the Euler-Maruyama discretization and the 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 and Jasiak (2001)), are usually followed by additional steps involving simulation and calibration (in order to improve the quality of the estimators). The third method object of our analysis will be the Generalized Method of Moments (GMM).

2   Two Basic Applications in Finance

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 for more complicated models, which use other underlying diffusion processes, and 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 options 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 al. (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:
$$\frac{dP(t)}{dt} = k(t)P(t), \ P_0 \text{ given}$$

4

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

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

$W(t)$ 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

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

Using Itô's Lemma:

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

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

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:

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

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

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

Solving (4) with condition (5) gives the price of the option as a function of time and of the parameter (which must be estimated) σ.

5

- Cox, Ingersoll and Ross (CIR-SR), (1985)[1]: In this model the state variable follows a diffusion process given by:

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

In this case the parameters α, θ and σ are the purpose of the statistical estimation.

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 $(\Omega, \mathcal{F}, P)$. 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 $f(X_t) : R^n \to R$, $\mathcal{Q}$ -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:

$$\Gamma_t f(x_0) = E\left[f(X_t) \mid X_0 = x_0\right] \quad (\equiv E^0(f(X_t))) \tag{6}$$

It can be shown that these operators $(L^2(\mathcal{Q}) \to L^2(\mathcal{Q}))$ are well defined ($f = f^*$ $\mathcal{Q}$-a.e.$\to$ $\Gamma_t(f) = \Gamma_t(f^*)$ $\mathcal{Q}$-a.e. 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, $E^0(X_{t+s}) = E^0(E^t(X_{t+s}))$, 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:

$$\Lambda f(x_0) = \lim_{t\downarrow 0} \frac{\Gamma_t f(x_0) - f(x_0)}{t}, \quad t \geq 0 \tag{7}$$

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).

---

[1] This process is usually abbreviated by CIR-SR, with SR denoting square root (because of the $\sqrt{X_t}$ term). One usually denotes by CIR-VR the process, used in another work of these authors, in which the exponent of X_t is 3/2, rather than 1/2. We shall come back to this process later in this text.

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

$$dX_t = b(X_t; \theta)dt + \sigma(X_t; \theta)dW_t \tag{8}$$

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

$$L_\theta = b(x; \theta)\frac{d}{dx} + \frac{1}{2}\sigma^2(x; \theta)\frac{d^2}{dx^2} \tag{9}$$

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 (8) 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:

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

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

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

Using the definition of L_θ established by (9),

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

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) and definition (6):

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

7

VI- Now take limits with $t \to 0$ 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:

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

Using the Cauchy-Schwarz (Hölder) inequality:

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

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

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

Using Fubini again,

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

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

$$= (1/t) \int_0^t \| \Gamma_s L_\theta f(x_0) - L_\theta f(x_0) \|^2 ds$$

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[2] 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 concentrate our exposition here on Prakasa Rao (1999) and

---

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

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 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 $[S, S + T]$:*

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

(QV stands for quadratic variation).

Proof. We omit the ω. Make

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

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

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

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

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

and

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

By the independence of increments:

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

$$(10)$$

because $(\frac{X(\frac{jT}{2^n})-X(\frac{(j-1)T}{2^n})}{\sqrt{\sigma^2 \frac{T}{2^n}}})^2$ has a chi-square distribution with one degree of freedom. (10) 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

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

■

Proposition 2 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):

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

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 $f(\theta) = \theta X_T - \theta^2 \frac{T}{2}$, with solution $\hat{\theta} = X_T/T$.*

4.2 Discretely Observed Data

4.2.1 The Case in Which the Transition Densities are known

There are three well known cases in which the stochastic differential equation (1) is easily solvable, and the corresponding transition functions known[3]: 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.

Example 3 *(Maximum likelihood estimation when the transition functions are known): Consider the equation that describes the evolution of the price of the underlying stock in the Black and Scholes model[4]. This falls into the first case considered above:*

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

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,

$$\log P_t = \log P_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W(t) \tag{12}$$

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

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

---

[3]Wong (1964) investigates some other particular cases.

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

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 $\sigma^2 t$:

$$E(P_t \mid P_0) = P_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \frac{\sigma^2}{2}t} = P_0 e^{\mu t} \tag{13}$$

Since the transition densities of Y_t are known, the application of maximum likelihood in this case follows in a straightforward way. Assuming that the first observation of the Markov process are known, equation (13) leads to the estimators of the mean $(\hat{a})$ and variance $(\hat{b})$ of $log P_t - \log P_{t-1}$ given by:

$$\hat{a} = \frac{1}{N} \sum_{t=1}^{N} (\log P_t - \log P_{t-1})$$
$$\hat{b} = \frac{1}{N} \sum_{t=1}^{N} (\log P_t - \log P_{t-1} - \hat{a})^2$$

The Maximum Likelihood Estimators μ and σ^2 are then given, respectively, by $\hat{a} + \frac{\hat{b}}{2}$ and $\hat{b}$.

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 (12). 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 $h \to 0$ (with $hT \to k \in \mathbb{R}$)[5].

4.2.2 The Case in Which 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 estimations in more general settings.

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:

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

---

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

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 (although biased) estimators.

Sorensen (1995) is the main source of our analysis in this section and the next. In contrast with other methods which use approximate likelihood ratios [Kutoyants(1984) and Yoshida (1992)], the method below uses the exact likelihood of a discretized process[6].

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

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

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 of $\mathbb{R}^d$. We discretize this process by assuming that the drift and the diffusion are constant in the time interval[7] $\Delta_i = t_i - t_{i-1}$:

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

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:

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

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

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

The transition density of the discretized process then reads:

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

---

[6]Both methods, though, lead to the same estimators, when the diffusion coefficient is constant (see Shoji, 1995).

[7]For equidistant intervals, such approximation, usually called an Euler-Maruyama aproximation [Kloeden and Platen (1992)], can be written

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

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

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

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

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

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

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

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

Example 4 *The Cox-Ingersoll-and-Ross process which we have first seen in section 2 (CIR-SR), satisfies the stochastic differential equation:*

$$dX_t = (\alpha + \theta X_t)dt + \sigma\sqrt{X_t}dW_t \tag{21}$$

with $\alpha > 0$, $\theta < 0$ and $\sigma > 0$. Assume that the distance between observation times, Δ, is the same along the sample. As shown by Sorensen (1995), for this process (20) leads to the estimators:

$$\hat{\alpha}_n = \frac{(X_{t_n} - X_0)(\frac{1}{n}\sum_{i=1}^{n}X_{t_{i-1}})^{-1} - \sum_{i=1}^{n}X_{t_{i-1}}^{-1}(X_{t_i} - X_{t_{i-1}})}{\Delta\left[n^2(\sum_{i=1}^{n}X_{t_{i-1}})^{-1} - \sum_{i=1}^{n}X_{t_{i-1}}^{-1}\right]}$$

$$\hat{\theta}_n = \frac{\sum_{i=1}^{n}X_{t_{i-1}}^{-1}(X_{t_i} - X_{t_{i-1}}) - \frac{1}{n}(X_{t_n} - X_0)\sum_{i=1}^{n}X_{t_{i-1}}^{-1}}{\Delta\left[n - (\sum_{i=1}^{n}X_{t_{i-1}})(\sum_{i=1}^{n}X_{t_{i-1}}^{-1})/n\right]}$$

$$\hat{\sigma}_n^2 = \frac{1}{n\Delta}\sum_{i=1}^{n}X_{t_{i-1}}^{-1}\left[X_{t_i} - X_{t_{i-1}} - (\hat{\alpha}_n + \hat{\theta}_n X_{t_{i-1}}\Delta)\right]^2$$

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:

$$\tilde{H}_N(\theta) = \sum_{i=1}^{N}\left\{ \frac{b_\theta(X_{t_{i-1}};\theta)}{v(X_{t_{i-1}};\theta)}\left[(X_{(t_i)} - X_{(t_{i-1})} - b(X_{t_{i-1}};\theta)\Delta_i\right]\right\} \tag{22}$$

Consistency and Asymptotic Distribution:

This estimating function has been studied by Dorogovcev (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 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) proved asymptotic normality imposing $n\Delta_n^2 \to 0$, whereas Florens-Zmirou (1989) used 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 distance between observation times, Δ, is sufficiently small. Kloeden et al. (1992) confirmed this fact through simulation, whereas Pedersen (1995a) and Bibby and Sorensen (1995) have shown that if Δ is not small the bias can be severe.

Improving the Approximations for the Moments   Lemma 1 in Florens-Zmirou[8] (1989) provides an expansion of $E_\theta(X_\Delta \mid X_0 = x)$ 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:

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 $\sigma-$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 $k-1$:*

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

(23)

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 (22).

---

[8]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 (16) and (17) (writing b for $b(x; \theta)$ and v for $v(x; \theta)$) read:

$$E_\theta(X_\Delta \mid X_0 = x) = x + \Delta b + \frac{1}{2}\Delta^2 \left\{ bb_x + \frac{1}{2}vb_{xx} \right\} + O(\Delta^3) \qquad (24)$$

and :

$$Var_\theta(X_\Delta \mid X_0 = x) = v\Delta + \Delta^2 \left[\frac{1}{2}bv_x + v \left\{ b_x + \frac{1}{4}v_{xx} \right\} \right] + O(\Delta^3) \qquad (25)$$

Note that (16) and (17) are a particular case of these expressions, for the cases when $l = 1$ in (23).

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

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

X_z standing for $X_{(k-1)\Delta+\delta}$ in Lemma 1.

Since $Lx = b(x, \theta)$, $L^2x = Lb(x, \theta) = bb_x + \frac{1}{2}vb_{xx}$ we get (24). The remaining $O(\Delta^3)$ derives from the fact that the (absolute value of the) integrand in (23) 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,

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

To get (25) we need $E_\theta(X_\Delta^2 \mid X_0 = x)$. Again, using (23):

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

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

$$E_\theta(X_\Delta^2 \mid X_0 = x) = x^2 + \Delta(2bx + v) + \frac{\Delta^2}{2}(b(2b + 2xb_x + v_x) \qquad (26)$$
$$+ \frac{1}{2}v(2b_x + 2b_x + 2xb_{xx} + v_{xx}))$$

16

From (24) we get:

$$[E_\theta(X_\Delta \mid X_0 = x)]^2 = x^2 + \Delta 2bx + \frac{\Delta^2}{2}(2b^2 + 2xbb_x + xvb_{xx}) + O(\Delta^3) \quad (27)$$

By subtracting (27) from (26) one gets (25).

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 (22) and (24) imply a bias of the estimating function (22) 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 (22):

$$\tilde{H}_N(\theta_0) - \tilde{H}_N(\tilde{\theta}) = \tilde{H}'_N(\tilde{\theta})(\theta_0 - \tilde{\theta}) \quad (28)$$

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

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

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

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

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 ith 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 $O(\Delta^{k+1})$): [9]:

$$E_\theta(X_\Delta \mid X_0 = x) = r_k(\Delta, x; \theta) + O(\Delta^{k+1}) \tag{29}$$

$$\Psi_k(\Delta, x; \theta) = \sum_{j=0}^{k} \Delta^j \sum_{r=0}^{k-j} \frac{\Delta^r}{r!} L_\theta^r g_{x,}^j(x) \tag{30}$$

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 (18) with (29) and (30) replacing (16) and (17) leads to a new approximate score function and to new estimators that perform better than the previous one.

Example 5 *Suppose X_t is governed by:*

$$dX_t = \theta X_t dt + \sigma dW_t$$

Sorensen (1995) has shown that the estimating function based on the approximation detailed in this subsection (with equidistant observations) leads to the estimators (for $k = 2$):

$$\tilde{\theta}_{2,n} = \Delta^{-1}(\sqrt{2Q_n - 1} - 1) \tag{31}$$

$$\tilde{\sigma}_{2,n}^2 = \frac{\frac{1}{n} \sum_{i=1}^{n} (X_{t_i} - X_{t_{i-1}} Q_n)}{\Delta + \tilde{\theta}_{2,n}\Delta^2 + \frac{2}{3}\tilde{\theta}_{2,n}^2\Delta^3} \tag{32}$$

---

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

where

$$Q_n = \frac{\sum_{i=1}^n X_{t_i} X_{t_{i-1}}}{\sum_{i=1}^n X_{t_{i-1}}^2} \tag{33}$$

provided that $Q_n \geq 1/2$.

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.

Measuring the Loss of Information Due to Discretization   Dacunha-Castelle (1986) assumes that the sampling is 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 significant 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 functionals[10]. 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 on account of discretization is of order Δ^2, whereas when σ is unknown the loss is of order Δ.

5   Martingale Estimating Functions (MEF)

This section is based on Kessler and Sorensen (1999). 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 standard in the literature.

---

[10]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.

Proposition 5 *Under regularity conditions (u.r.c.), 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 $\sigma-$algebra $\mathcal{F}_{n-1}$ we have (because the restriction of P and Q to $\mathcal{F}_n$ and the restriction of P and Q 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 $\mathcal{F}_{n-1}$:

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

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

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 (but not eliminated) by the use of better approximations to the mean and to the variance of the transition density. 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:

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \tag{34}$$

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

$$\int g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta)\, p(\Delta, x, y; \theta) = 0 \tag{35}$$

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[11]. The approach followed by Kessler and Sorensen does not require that the functions $g(.)$ are polynomials, even though, in some cases, they happen to be. The example below details one such case.

Example 6 *Take the Cox-Ingersoll-and-Ross (21) model presented before in this text. As shown by Kessler and Sorensen, for $n = 0, 1, ...,$ this model leads to the spectrum:*

$$\Lambda_\theta = \{-n\theta\}$$

with eigenfunctions:

$$\phi_i(X) = \sum_{m=0}^{i} (-1)^m \begin{pmatrix} i + 2\alpha\sigma^{-2} - 1 \\ i - m \end{pmatrix} \frac{X^m}{m!} (-2\theta X \sigma^{-2})$$

They are based on the eigenfunctions of the generator of the diffusion process[12].

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.

---

[11] This 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.

[12] 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.

Note that (35) 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 (28), equations (34) and (35) 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 (34):

$$g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) = \sum_{j=1}^{N} \alpha_j(\Delta, x; \theta) h_j(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \qquad (36)$$

Since for each j, the event $[\alpha_j(\Delta, x; \theta) \in B, \text{B borelian in } \mathbb{R}^n] \in \mathcal{F}_{t_{i-1}}$ (where $x = X_{t_{i-1}}$), such functions satisfy (35) 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 (36). 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 (8). Let $\phi(x; \theta)$ be an eigenfunction and $\lambda(\theta)$ an eigenvalue of the operator L_θ. Then, under weak regularity conditions (u.r.c.):*

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

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:
$$Z_t = e^{\lambda t} \phi(X_t) \qquad (37)$$

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 (8) 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 $(-\lambda)$, $L\phi(X_t) + \lambda\phi(X_t) = 0$ 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 (37) 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, u.r.c., 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

As detailed in the seminal paper by Hansen (1982), a GMM estimator is obtained by minimizing a criterion function of sample moments which are derived from orthogonality conditions. 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). Also, when the processes are stationary these generators can be employed to derive moment conditions that can be used for estimation purposes by the application of Hansen's (1982) GMM.

Note that, by the law of iterated expectations:

$$\int_{R^n} f d\mathcal{Q} = \int_{R^n} \Gamma_t f d\mathcal{Q} \tag{38}$$

Using the framework developed in section 3, since $\frac{1}{t}(\Gamma_t f - f)$ converges in $L^2(\mathcal{Q})$ to $\Lambda_t f$, (38) implies:

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

and the (first set of) moments conditions[13]:

$$E\left[\Lambda f(X_t)\right] = 0 \text{ for all } f \in \Psi \tag{39}$$

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:

$$E\left[\Lambda f(X_{t+1})\check{f}(X_t) - f(X_{t+1})\check{\Lambda}\check{f}(X_t)\right] = 0 \text{ for all } f \in \Psi, \check{f} \in \check{\Psi} \tag{40}$$

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.

In practice, GMM estimation usually starts with the employment of an Euler-Maruyama discretization. Using the notation here developed, take, for instance, the diffusion process:

$$dX_t = g(X_t)dt + \sigma dW_t$$

By using the Eulyer-Maruyama approximation:

$$X_t - X_{t-1} = g(X_{t_{i-1}}) + \sigma(X_{t_i} - X_{t_{i-1}})$$

Next, define $e_{t_i} = X_{t_i} - X_{t_{i-1}} - g(X_{t_{i-1}})$, $f_{t_i} = (X_{t_i} - X_{t_{i-1}} - g(X_{t_{i-1}}))^2 - \sigma^2 \Delta t$ and $z_t = (e_{t_i}, f_{t_i}, X_{t_{i-1}} e_{t_i}, X_{t_{i-1}} f_{t_i})$. Under the null, $E_{t-1} z_{t_i} = 0$. Replace this conditional expectation by its sample counterpart to obtain a quadratic form, the maximization of which leads to the parameter estimates (see the second example below).

---

[13]Remember the definition of Ψ from Section 3.

Example 7 *Consider the diffusion process:*

$$dp = -a(p - \mu)dt + \sigma dW_t, \ p(0) = p_0 > 0, \ a > 0 \tag{41}$$

Campbell et al. (1997) present a heuristic development of this example. The direct use of the results we proved in section 3, though, allows for a direct formal approach. Taking f as the identity function and applying the generator Λ to (41):

$$\Lambda(dp) = -a(p - \mu) \tag{42}$$

Using (39) leads to the first moment condition:

$$Ep = \mu \tag{43}$$

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 (39) once more:

$$E \left[-f'(p)a(p - \mu) + \frac{1}{2}f''(p)\sigma^2 \right] = 0 \tag{44}$$

Using (40):

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

Equations (39) and (40) define an infinite number of moment conditions, depending on the choice of f. Under the regularity conditions provided by the authors, GMM can then be applied.

Example 8 *As a second example, and also for the purpose of comparison with one of the other estimation procedures we have seen in this paper, let's go back to the CIR-SR model of interest rates presented in section 2 and in example 4. Chan et al. (1992) estimate this process using GMM. Denoting by X_t the interest rate at time t, these authors estimate a discrete time version of the CIR-SR model given by:*

$$X_{t+1} - X_t = \alpha + \theta X_t + \epsilon_{t+1}$$

in which $E \; \epsilon_{t+1} = 0$ and $E(\epsilon_{t+1})^2 = \sigma^2 X_t$. The estimation goes as follows. Make $\beta = (\alpha, \theta, \sigma^2)$. Next, define:

$$f_t(\beta) = \begin{bmatrix} \epsilon_{t+1} \\ \epsilon_{t+1} X_t \\ \epsilon_{t+1}^2 - \sigma^2 X_t \\ (\epsilon_{t+1}^2 - \sigma^2 X_t) X_t \end{bmatrix}$$

Under the null, $E f_t(\beta) = 0$. The estimators associated with the GMM method are then found simply by replacing $E f_t(\beta)$ with its sample counterpart. The process leads to a minimization of a quadratic form.

Example 9 *(Brazilian Financial Time Series) For an estimation of the*

CIR-SR model (and other interest-rate models as well) with Brazilian financial time series, the reader can refer to Barrossi-Filho and Genaro Dario (2003). In another section of the paper, these authors also use Monte Carlo simulation methods to compare finite-sample distribution properties of the GMM and of the Euler-Maruyama approximation.

7 Comparing Different Estimators

Comparisons of the estimators studied in this (partial) survey are relatively scarce in the literature. Changes of the time interval used for the simulations is one reason for this. The unavoidable aliasing phenomenon (the fact that distinct continuous-time processes may look identical when sampled at discrete points of time), can lead to different results, making it extremely difficult to make general statements about relative efficiencies of one or another estimator.

Checking the influence of changes in discrete-time sampling interval requires repeating the experiments several times (once the number of sample paths has been set, depending upon the discrete time intervals, the number of sample points will, of course, vary accordingly).

There is still the problem that, when the diffusion term is state dependent, it is possible that the numerical values of the sample path diverges, because the variance becomes large. This usually implies the necessity of a transformation of variables (in order to obtain a stochastic differential equation with a constant diffusion coefficient), and of a reversal of the transformation later.

Shoji and Osaki (1997) is one of the few examples performing such comparisons. This paper develops Monte Carlo experiments using five different

methods of obtaining MLE estimators (including the Euler-Maruyama approach and GMM, both of which we have reviewed here). Two different diffusion processes are considered, the first with linear drift and state-dependent diffusion coefficients and the other with nonlinear drift and constant diffusion coefficients:

$$dX_t = (\alpha + \beta X_t)dt + \sigma X_t dW_t \tag{45}$$

and

$$dX_t = \alpha X_t^3 dt + \sigma dW \tag{46}$$

In the experiments, the number of sample paths is fixed and the number of sample points varies depending upon the discrete time interval. Note, in (45), that the diffusion term is state dependent, thereby causing the type of problem we have mentioned above in this section, requiring a change of variable.

After making Monte Carlo simulations, these authors concluded, regarding (45), that the GMM performed somewhat inferior to the Euler-Maruyama approach (a poorer performance of GMM estimators is also found by Barrossi-Filho and Genaro Dario (2004)). The GMM, though, did a little better in (46) than in (45).

The interested reader can refer to Jiang and Knight (1999) and Mykland and Ait-Sahalia (2000) for further study of the comparison among different estimators. Jiang and Knight (1999) use Monte Carlo simulation to investigate the finite-sample properties of various estimators, including GMM estimators and some others which we have not reviewed here. Ait-Sahalia and Mykland (2000) provide a general method to compare the performance of a variety of estimators of diffusion processes, when the data are not only discretely sampled in time but, in addition, the time separating successive observations may possibly be random. GMM and the Euler-Maruyama approximation are among the methods assessed by these authors.

8 Conclusion

In this paper, we have provided a partial review of the literature regarding the statistical estimation of diffusion processes by investigating three of the available estimation methods: an improvement of the Euler-Maruyama discretization scheme, the employment of Martingale Estimating Functions and the application of Generalized Method of Moments (GMM).

In order to interest the reader in this important area of statistical research, we have provided several examples, a final section with a brief review of a comparison between the performance of two of the methods that we

have investigated, as well as detailed formalizations of some of the analytical developments made in the original texts. This is an area of research that is presently very active and of particular interest to economists, particularly, at the present stage, because of its several applications in finance and macroeconomics.

A research parallel to this one involves studying the statistical consequences of the random sampling (as opposed to the effects on which we have concentrated here, of the discrete sampling) of the diffusion processes. Ait-Sahalia and Mykland (2003) are a seminal reference in this area, where a new body of research is expected to emerge.

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] Ait-Sahalia, Y. and Mykland, Per (2000): "A Comparison of Estimators of Continuous-Time Diffusions". Annals of the Wold Congress of the Econometric Society, Seattle, 2000.

[3] Ait-Sahalia, Y. and Mykland, Per (2003): "The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions.". Econometrica, Vol 71, N. 2: 483-549.

[4] Barrossi-Filho, M. and Dario, A. de Genaro (2004): "Estimação de Modelos em Tempo Contínuo: Taxa de Juros de Curto Prazo e Simulações de Monte Carlo". Mimeo, FEA/USP.

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

[6] Bibby, B.M. and Sorensen, M. (1995): "Martingale Estimation Functions for Discretely Observed Diffusion Processes". Bernoulli, 1 , 17-39.

[7] Billingsley, P. (1961): "The Lindberg-Levy Theorem for Martingales". Proc. of the American Math. Soc., 12, 788-792.

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

[9] Campbell, J. Y, Lo, A. W. and MacKinlay, A. C. (1997): "The Econometrics of Financial Markets". Princeton University Press.

[10] Chan, K; Karolyi, G; Longstaff, Francis & Sanders, Anthony (1992). "An Empirical Comparison of Alternative Models of Short-Term Interest Rate". Journal of Finance, v. XLVII: 1209-1227.

[11] Cox, J., Ingersoll, J. and Ross, S. (1985): "A Theory of the Term Structure of Interest Rates". Econometrica, 53, 385-408.

[12] Dacunha-Castelle, D. (1986): "Estimation of the Coefficients of a Diffusion from Discrete Observations", Stochastics, vol. 19, pp. 263-284.

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

[14] 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.

[15] Florens-Zmirou, D. (1989): "Approximate Discrete-Time Schemes for Statistics of Diffusion Processes". Statistics, 20, 547-557.

[16] Godambe, V. P. and Heyde, C. (1987): "Quasi Likelihood and Optimal Estimation". Int. Statist. Rev. 55, 231-244.

[17] Gourieroux, Christian and J. Jasiak (2001): "Financial Econometrics: Problems, Models, and Methods". Princeton Series in Finance.

[18] Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica 50, 1029-1054.

[19] 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.

[20] Jiang. G and John L. Knight (1999): "Finite Sample Comparison of Alternative Estimators of Itô Diffusion Processes: a Monte Carlo Study". Journal of Computational Finance, Volume 2, N. 30.

[21] Karatzas, I., Shreve, S.E. (1991): "Brownian Motion and Stochastic Calculus". Second Edition, Spring Verlag.

[22] Karlin, S. and Taylor, H. M. (1981): "A Second Course in Stochastic Process". Orlando: Academic Press.

[23] Kessler, M. (1997): "Estimation of an Ergodic Diffusion from Discrete Observations". Scandinavian Journal of Statistics, Volume 24 Issue 2 Page 211.

[24] Kessler, M. and Sorensen, M. (1999): "Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Process". Bernoulli **5**(2), 1999, 299-314.

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

[26] 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.

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

[28] Kutoyants, Y. A. (1984): "Parameter Estimation for Stochastic Processes" (trans. and ed. B. L. S. Prakasa Rao). Berlin, Heldmann.

[29] Oksendal, B. (2000): "Stochastic Differential Equations", 5th Edition, Springer Verlag.

[30] 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.

[31] Pedersen, A.R. (1995b): "Consistency and Asymptotic Normality of an Approximate Maximum Likelihood Estimator for Discretely Observed Diffusion Processes". Bernoulli 1, 257-279.

[32] Phillips, A. W. (1959): "The estimation of Parameters in Systems of Stochastic Differential Equations". Biometrika 46, 67–76.

[33] Prakasa Rao, B.L.S. (1983): "Asymptotic Theory of Non-Linear Least Squares Estimator for Diffusion Processes". Math. Operationsforsch. Statist.

[34] Prakasa Rao, B.L.S. (1988): "Statistical Inference from Sampled Data for Stochastic Processes". Contemporary Mathematics 80, 249-284.

[35] Prakasa Rao, B.L.S (1999): "Statistical Inference for Diffusion-Type Processes". Kendall's Library of Statistics, number 8.

[36] Shoji, I. (1995): "Consistency of the Estimator Derived from the Euler Method". Research Memo. 587, Institute of Statistical Mathematics, Tokyo..

[37] Shoji, I and Tohru Ozaki (1997): "Comparative Study of Estimation Methods for Continuous Time Stochastic Processes", Journal of time series analysis vol. 18, no. 5.

[38] Sorensen, M. (1995): "Estimating Functions for Discretely Observed Diffusions: A Review".

[39] Sorensen, H (2002). "Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: a Survey". Mimeo, University of Copenhagen.

[40] Stokey, N. (2000). "Continuous Dynamic Optimization", Mimeo, The University of Chicago.

[41] Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure of Interest Rates". Journal of financial economics, 5, 177-188.

[42] Wong, E. (1964): "The Construction of a Class of Stationary Markov Processes". Stochastic Processes in Mathematical Physics and Engineering. Proceedings of Simposia in Applied Mathematics, 16, ed. by R. Bellman. Providence, RI: American Mathematical Society, pp. 264-276.

[43] Yoshida, N. (1992): "Estimation for Diffusion Processes from Discrete Observations", Journal of Multivariate Analysis, 41, 220-242.

Últimos Ensaios Econômicos da EPGE

[544] Samuel de Abreu Pessôa e Rafael Rob. *The implications of embodiment and Putty–Clay to economic development*. Ensaios Econômicos da EPGE 544, EPGE–FGV, Abr 2004.

[545] Henrique Dezemone Forno e Ricardo de Oliveira Cavalcanti. *Money with Bank Networks*. Ensaios Econômicos da EPGE 545, EPGE–FGV, Jun 2004.

[546] Henrique Dezemone Forno e Ricardo de Oliveira Cavalcanti. *Cyclical Interest Payments on Inside Money*. Ensaios Econômicos da EPGE 546, EPGE–FGV, Jun 2004.

[547] Fernando de Holanda Barbosa. *Dois experimentos de política monetária no modelo Novo–Keynesiano*. Ensaios Econômicos da EPGE 547, EPGE–FGV, Abr 2004.

[548] Samuel de Abreu Pessôa, Pedro Cavalcanti Gomes Ferreira, e Fernando A. Veloso. *The evolution of international output differences (1960–2000): From factors to productivity*. Ensaios Econômicos da EPGE 548, EPGE–FGV, Jun 2004.

[549] Márcio Antônio Salvato, Fábio Augusto Reis Gomes, e João Victor Issler. *Principais características do consumo de duráveis no Brasil testes de separabilidade entre duráveis e não–duráveis*. Ensaios Econômicos da EPGE 549, EPGE–FGV, Jun 2004.

[550] João Victor Issler e Rachel Couto Ferreira. *Avaliando pesquisadores e departamentos de economia no Brasil a partir de citações internacionais*. Ensaios Econômicos da EPGE 550, EPGE–FGV, Jun 2004.

[551] Humberto Luiz Ataide Moreira e David Martimort. *Common agency with informed principals*. Ensaios Econômicos da EPGE 551, EPGE–FGV, Jun 2004.

[552] Luiz Renato Regis de Oliveira Lima e Zhijie Xiao. *Purchasing power parity and the unit root tests: A robust analysis*. Ensaios Econômicos da EPGE 552, EPGE–FGV, Jul 2004.

[553] Daniel Gottlieb, Aloisio Pessoa de Araújo, e Humberto Luiz Ataide Moreira. *A model of mixed signals with applications to countersignaling an the GED*. Ensaios Econômicos da EPGE 553, EPGE–FGV, Jul 2004.

[554] Lucas Jóver Maestri e Carlos Eugênio Ellery Lustosa da Costa. *The risk–properties of human capital and the design of government policies*. Ensaios Econômicos da EPGE 554, EPGE–FGV, Jul 2004.

[555] Daniel Gottlieb e Lucas Jóver Maestri. *Banning information as a redistributive device*. Ensaios Econômicos da EPGE 555, EPGE–FGV, Jul 2004.

[556] Leonardo Pio Perez e Pedro Cavalcanti Gomes Ferreira. *Efeitos macroe-conômicos e custos sociais de uma transição entre regimes de previdência no Brasil*. Ensaios Econômicos da EPGE 556, EPGE–FGV, Jul 2004.

[557] Rubens Penha Cysne. *Inflation and income inequality: A link through the job–search process*. Ensaios Econômicos da EPGE 557, EPGE–FGV, Ago 2004.

[558] Rubens Penha Cysne. *A search–theoretic explanation for the negative correla-tion between labor income and impatience*. Ensaios Econômicos da EPGE 558, EPGE–FGV, Ago 2004.

[559] Rubens Penha Cysne. *Income inequality:The role of impatience in a job–search process*. Ensaios Econômicos da EPGE 559, EPGE–FGV, Ago 2004.

[560] Rubens Penha Cysne. *Towards a measure of income inequality freed from the volatility caused by variations in the rate of unemployment*. Ensaios Econômicos da EPGE 560, EPGE–FGV, Ago 2004.

[561] Rubens Penha Cysne. *On the positive correlation between income inequality and unemployment*. Ensaios Econômicos da EPGE 561, EPGE–FGV, Ago 2004.

[562] Rubens Penha Cysne. *A general–equilibrium closed–form solution to the wel-fare costs of inflation (Forthcoming, Revista Brasileira de Economia)*. Ensaios Econômicos da EPGE 562, EPGE–FGV, Ago 2004.

[563] Marcelo Côrtes Neri e Marcelo Casal Xerez. *Aspectos dinâmicos de um sistema de metas sociais*. Ensaios Econômicos da EPGE 563, EPGE–FGV, Ago 2004.

[565] Marcelo Côrtes Neri e Marcelo Casal Xerez. *Desenho de um sistema de metas sociais*. Ensaios Econômicos da EPGE 565, EPGE–FGV, Set 2004.

[566] Paulo Klinger Monteiro, Rubens Penha Cysne, e Wilfredo Maldonado. *Inflation and Income Inequality: A Shopping–Time Aproach (Forthcoming, Journal of Development Economics)*. Ensaios Econômicos da EPGE 566, EPGE–FGV, Set 2004.

[567] Rubens Penha Cysne. *Solving the Non–Convexity Problem in Some Shopping–Time and Human–Capital Models*. Ensaios Econômicos da EPGE 567, EPGE–FGV, Set 2004.

[568] Paulo Klinger Monteiro. *First–Price auction symmetric equlibria with a general distribution*. Ensaios Econômicos da EPGE 568, EPGE–FGV, Set 2004.

[569] Samuel de Abreu Pessôa, Fernando A. Veloso, e Pedro Cavalcanti Ferreira. *On The Tyranny of Numbers: East Asian Miracles in World Perspective*. Ensaios Econômicos da EPGE 569, EPGE–FGV, Out 2004.