

Statistical inference for diffusion processes with discrete data: a survey

Inferenza statistica per processi di diffusione da osservazioni discrete: una rassegna

Sergio Pastorello

Dipartimento di Scienze Economiche

Università di Bologna

pastorel@economia.unibo.it

Eduardo Rossi

Dipartimento di Economia Politica

e Metodi Quantitativi

Università di Pavia

erossi@eco.unipv.it

Riassunto: In questo lavoro presentiamo alcuni risultati recenti nell'ambito della stima dei parametri di processi di diffusione da osservazioni in tempo discreto. L'attenzione è concentrata sul caso scalare, in assenza di salti. Sono illustrati due approcci basati sulla verosimiglianza, rispettivamente su un'espansione di Hermite della densità di transizione ignota e sul calcolo di quest'ultima per integrazione Monte Carlo, e due strategie GMM, basate sul generatore infinitesimale e sullo score di un modello ausiliario opportunamente scelto. Infine, vengono brevemente richiamate le proprietà teoriche ed empiriche dei diversi approcci.

Keywords: Diffusion process, discrete sampling, maximum likelihood estimation, generalized method of moments estimation.

1. Introduction

This paper surveys some recent approaches to statistical inference on the parameters of an SDE from discrete data, i.e. when the process is sampled at discrete time intervals (instead of a continuous record of the trajectory). For the sake of simplicity, in this paper we mainly consider a univariate (x_t scalar), time-homogeneous (i.e., μ and σ do not depend on t) SDE without jump component:

$$dx_t = \mu(x_t; \theta)dt + \sigma(x_t; \theta)dw_t \quad (1)$$

Typically, some assumption is needed to ensure that (1) has a nonexplosive, unique weak solution. In the following, we assume that some set of sufficient conditions are verified.

We restrict our exposition to parametric approaches, where the drift and the diffusion functions are known up to an unknown parameter vector $\theta \in \Theta$, a compact subset of $\mathbb{R}^d$. Parametric methods may be further classified as likelihood or non-likelihood based. For the sake of brevity, we do not deal with nonparametric or semiparametric strategies, in which either μ or σ , or both, are assumed to be functions of the contemporaneous value of x_t .

A basic requirement for all these approaches to produce consistent and asymptotically normal estimates is that the sample size n grows to infinity. For most of them, however, this condition is not sufficient, because they are based on an approximation which has to vanish sufficiently rapidly to ensure that the estimators exhibit the desired properties.

Some further condition is required to ensure this result. In some cases, based on approximate discretizations of the parametric SDE of interest, it is necessary to impose that the length of the interval between consecutive observations (Δ) shrinks to zero. This kind of assumptions, labelled infill asymptotics (as opposed to sample size asymptotics), considerably simplifies the task, but is somewhat unappealing for two reasons. First, it is often the case that Δ is fixed, and more frequent observations cannot simply be collected. Second, frequently the SDE is just a mathematically convenient device to approximate the true continuous time process of the variable of interest, and the approximation is adequate only if Δ is not too small. For example, the trajectory over time of the price of a stock looks like the realization of a diffusion process if the observations are sampled daily; however, high frequency (tick by tick) data show that the trajectory is not continuous (as it should be was it a diffusion), but rather piecewise constant, and should consequently be modelled using a marked point process.

For parametric SDEs, an appealing alternative is represented by the approximations of the loglikelihood which can be made arbitrarily good by increasing either the number of terms in the analytical expansion of the true transition density (see section 2), or the number of simulation steps (section 3). Let us denote with M the order of the approximation. In this case, consistency (either towards the true value of the parameters θ , or the true maximum likelihood estimator) and asymptotic normality of the maximum approximate likelihood estimates requires that $M \rightarrow \infty$. This condition is in general more appealing than assuming $\Delta \rightarrow 0$.

While we limit our exposition to the case of the basic SDE (1), most of the inference strategies that we outline may be generalized in different interesting directions, such as models with jump components, time-inhomogeneous models, multivariate settings with or without partial observability on x_t (i.e. models with latent variables), and random sampling schemes whose process may or may not be dependent upon x_t or θ . While interesting, each of these extensions poses a number of analytical and numerical difficulties. Again, for simplicity we do not consider them in detail here. However, we briefly mention whether an extension is available to deal with the above issues.

In sections 2 and 3 we present two approaches based on the likelihood, while sections 4 and 5 focus on GMM approaches. Some concluding remarks are outlined in section 6.

2. Likelihood estimation with Hermite expansion

With a continuous record of observations, the likelihood could be computed using an absolutely continuous change of measure (see Basawa and Prakasa Rao (1980)). With discrete data, things are much tougher. Suppose for simplicity that the sample is given by equally spaced observations $\{x_{i\Delta}, i = 0, 1, \dots, n\}$. The object of interest is the unknown transition density function $p_x(x_\Delta, \Delta | x_0; \theta)$. In a few cases, $p_x(x_\Delta, \Delta | x_0; \theta)$ has a closed form expression. In general, however, some kind of approximation is required to compute the likelihood. In this section we present a strategy which provides a closed form approximation to the true likelihood, while section 3 outlines an alternative approach that approximates it numerically.

To avoid the need to assume that $\Delta \rightarrow 0$, Aït-Sahalia (2002) has suggested to expand the unknown density of interest using Hermite polynomials. The approximation could be made arbitrarily good by choosing a sufficiently high order of the expansion, without any need of assumptions about Δ . A general result states that the Hermite expansion of a

given density converges only if the latter is “close” to normal, in the sense that its tails are sufficiently thin for a certain condition to hold. This condition is violated by the transition densities of most of the interesting specifications of the SDE (1).

Aït-Sahalia (2002), however, has shown that the required condition can be met by transforming (1) into another SDE, with the density sufficiently close to gaussian for its Hermite expansion to converge. This can be achieved by using the following two steps. First, transform x into y using:

$$y = G(x, \boldsymbol{\theta}) = \int^x \frac{1}{\sigma(u, \boldsymbol{\theta})} du \quad (2)$$

where the constant of integration is irrelevant. Notice that by the strict positivity of $\sigma(x, \boldsymbol{\theta})$ on the domain of x , $G(x, \boldsymbol{\theta})$ is monotonically increasing and invertible for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. By Itô’s Lemma, the SDE for y_t is given by:

$$dy_t = \mu_y(y_t, \boldsymbol{\theta}) dt + dw_t \quad (3)$$

where:

$$\mu_y(y_t, \boldsymbol{\theta}) = \frac{\mu[G^{-1}(y_t, \boldsymbol{\theta}), \boldsymbol{\theta}]}{\sigma[G^{-1}(y_t, \boldsymbol{\theta}), \boldsymbol{\theta}]} - \frac{1}{2} \sigma'[G^{-1}(y_t, \boldsymbol{\theta}), \boldsymbol{\theta}]$$

Notice that by the Jacobian formula:

$$p_x(x_\Delta, \Delta | x_0; \boldsymbol{\theta}) = \frac{1}{\sigma(x_\Delta, \boldsymbol{\theta})} p_y[G(x_\Delta, \boldsymbol{\theta}) | G(x_0, \boldsymbol{\theta}); \boldsymbol{\theta}] \quad (4)$$

Aït-Sahalia (2002, Proposition 2) shows that, under mild regularity conditions, the tails of $p_y(y_\Delta, \Delta | y_0; \boldsymbol{\theta})$ have a gaussian-like upper bound.

Unfortunately, this property does not represent a complete solution because it is also possible to verify that, as $\Delta \rightarrow 0$, $p_y(y_\Delta, \Delta | y_0; \boldsymbol{\theta})$ shrinks towards a Dirac measure on y_0 . To avoid this, a second step performs a pseudo-normalization of y , defined as:

$$z_\Delta = \Delta^{-1/2}(y_\Delta - y_0)$$

While this does not amount to an exact standardization of y , it is nonetheless sufficient to make the conditional density of z close enough to a $\mathcal{N}(0, 1)$ to admit a converging Hermite expansion. An application of the Jacobian formula immediately gives:

$$p_y(y_\Delta, \Delta | y_0; \boldsymbol{\theta}) = \Delta^{-1/2} p_z[\Delta^{-1/2}(y_\Delta - y_0) | y_0; \boldsymbol{\theta}] \quad (5)$$

We are now left with the problem of approximating $p_z(z_\Delta | y_0; \boldsymbol{\theta})$. We define the Hermite polynomials as:

$$H_j(z) = e^{\frac{1}{2}z^2} \frac{d^j}{dz^j} e^{-\frac{1}{2}z^2}, \quad \text{for } j \geq 0$$

and the Hermite expansion of $p_z(z_\Delta | y_0; \boldsymbol{\theta})$ as:

$$p_z^{(J)}(z_\Delta, \Delta | y_0; \boldsymbol{\theta}) = \phi(z_\Delta) \sum_{j=0}^J \eta_z^{(j)}(\Delta, y_0, \boldsymbol{\theta}) H_j(z_\Delta) \quad (6)$$

where $\phi_{\mathcal{N}}(Z)$ is the density of a $\mathcal{N}(0, 1)$ random variable. Given $p_z^{(J)}(z_\Delta, \Delta | y_0; \boldsymbol{\theta})$, a sequence of approximations $p_x^{(J)}(x_\Delta, \Delta | x_0; \boldsymbol{\theta})$ can be easily constructed by first applying

(5), and then (4). Aït-Sahalia (2002, Theorem 1) has shown that, under some regularity conditions, there exists $\bar{\Delta} > 0$ such that, for every $\Delta \in (0, \bar{\Delta})$, $\theta \in \Theta$ and x_Δ and x_0 in the domain of x ,

$$\lim_{J \rightarrow \infty} p_x^{(J)}(x_\Delta, \Delta | x_0; \theta) = p_x(x_\Delta, \Delta | x_0; \theta)$$

Moreover, this convergence is uniform in θ over Θ , and $\Delta = +\infty$ if $\mu_y(y, \theta)$ is negative (positive) near the right (left) boundary of the domain of y .

Given the closed form approximation $p_x^{(J)}(x_\Delta, \Delta | x_0; \theta)$, it is natural to estimate θ by maximizing the approximate sample loglikelihood. Notice first that, for finite J , it is possible that the approximated transition density be negative for some x . Before taking the logarithm, hence, a trimming procedure has to be applied to insure the positiveness of the argument. Let us now denote by $\hat{\theta}^{(J)}$ the estimator of θ obtained by maximizing the trimmed approximate loglikelihood function:

$$\hat{\theta}^{(J)} = \arg \max_{\theta \in \Theta} \sum_{i=0}^{n-1} \log p_x^{(J)}(x_{(i+1)\Delta}, \Delta | x_{i\Delta}; \theta)$$

and by $\hat{\theta}$ the exact (in general uncomputable) maximum likelihood estimator:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \sum_{i=0}^{n-1} \log p_x(x_{(i+1)\Delta}, \Delta | x_{i\Delta}; \theta)$$

Theorem 2 in Aït-Sahalia (2002) shows that, under suitable conditions, and for $\Delta \in (0, \bar{\Delta})$, $\hat{\theta}^{(J)}$ inherits the properties of $\hat{\theta}$ in the following sense:

1. As $J \rightarrow \infty$, and under the probability law induced by the true value θ_0 of θ , $\hat{\theta}^{(J)} \xrightarrow{p} \hat{\theta}$
2. As $n \rightarrow \infty$, it is possible to choose a sequence $J_n \rightarrow \infty$ such that $\hat{\theta}^{(J_n)}$ and $\hat{\theta}$ share the same asymptotic distribution.

The coefficients $\eta_Z^{(j)}(\Delta, y_0, \theta)$ in the density expansion (6) can be computed in a variety of ways. A particularly convenient one is based on the observation that:

$$\eta_z^{(j)}(\Delta, y_0, \theta) = \frac{1}{j!} \int_{\mathbb{R}} H_j(z_\Delta) p_z(z_\Delta | y_0; \theta) dz_\Delta = \frac{1}{j!} \mathbb{E} \{ H_j [\Delta^{-1/2}(y_\Delta - y_0)] | y_0; \theta \}$$

Since the coefficients of the Hermite expansion are moments of the conditional distribution of y_Δ given y_0 , they can be computed in a number of ways, including Monte Carlo integration. However, a particularly appealing alternative is to approximate them using the first $K + 1$ terms of the following formula:

$$\mathbb{E}[g(y_\Delta) | y_0] = \sum_{k=0}^K \mathcal{A}^k g(y_0) \frac{\Delta^k}{k!} + \mathbb{E}[\mathcal{A}^{K+1} g(y_\Delta) | y_0] \frac{\Delta^{K+1}}{(K+1)!}$$

where we denote with $\mathcal{A}$ the infinitesimal generator associated to the diffusion process (3) for y , defined by:

$$\mathcal{A}g(y_0) = \lim_{h \rightarrow 0} \frac{\mathbb{E}(y_h | y_0) - y_0}{h} \quad (7)$$

For this limit to exist, some condition has to be imposed on $g(\cdot)$. In most cases it is sufficient that $g \in \mathcal{C}^2$. Let us recall that for a scalar diffusion process such as (1), the infinitesimal generator is a function of the drift and the diffusion coefficient and the first two derivatives of g :

$$\mathcal{A}g(x) = \mu(x, \boldsymbol{\theta})g'(x) + \frac{1}{2} \sigma(x, \boldsymbol{\theta})g''(x)$$

This leaves us with the problem of choosing the orders J and K of the Hermite-Taylor approximation. Some numerical experiments have shown that the accuracy obtained with $J = 6$ and $K = 3$ is very high for the values of Δ commonly encountered in empirical applications. The closed form approximation outlined in this section has been applied in Aït-Sahalia (1999) to a variety of specifications of the short term rate and other diffusions. The results are extremely encouraging; the approximation error, measured by $\max_x |p^{(J)}(x|x_0; \boldsymbol{\theta}) - p(x|x_0; \boldsymbol{\theta})|$, is in most cases negligible already with $J = 2$, and, roughly speaking, increasing by 1 the order J of the approximation produces additional improvements by a factor of a least 10.

3. Simulated likelihood

Pedersen (1995) and Brandt and Santa Clara (2002) have suggested an alternative and much more general approximating strategy for the likelihood function. In general, the original SDE may be discretized according to a variety of approximation schemes with differing order of strong convergence (see Kloeden and Platen (1992) for an introduction to discretization schemes). Suppose to fix an approximation scheme, a denote with $p_x^{(1)}(x_\Delta, \Delta|x_0; \boldsymbol{\theta})$ the transition density (the “subdensity”) implied by such a choice. As an example, for the Euler scheme we know that:

$$p_x^{(1)}(x_\Delta, \Delta|x_0; \boldsymbol{\theta}) = \phi_{\mathcal{N}}[x_\Delta; x_0 + \mu(x_0, \boldsymbol{\theta})\Delta, \sigma(x_0, \boldsymbol{\theta})\Delta] \quad (8)$$

where we denote with $\phi_{\mathcal{N}}(x; m, v^2)$ the $\mathcal{N}(m, v^2)$ density.

Direct estimation of $\boldsymbol{\theta}$ based on the approximate loglikelihood implied by these discretization schemes is likely to work well only for time intervals shorter than those usually available in applications. To overcome this problem, we can partition the interval $(0, \Delta)$ in M subintervals as $0 = \tau_0 < \tau_1 < \dots < \tau_M = \Delta$, and define:

$$p_x^{(M)}(x_\Delta, \Delta|x_0; \boldsymbol{\theta}) = \int_{\mathbb{R}^{M-1}} \prod_{m=1}^M p_x^{(1)}(u_m, \Delta_m|u_{m-1}; \boldsymbol{\theta}) du_1 \dots du_{M-1} \quad (9)$$

where $u_0 = x_0$, $u_M = x_\Delta$, and $\Delta_m = \tau_m - \tau_{m-1}$. Pedersen (1995, Theorems 1,2,3) has shown that under (mild) regularity conditions, $p_x^{(M)}(x_\Delta, \Delta|x_0; \boldsymbol{\theta})$ exists and converges in L^1 towards $p_x(x_\Delta, \Delta|x_0; \boldsymbol{\theta})$ as $M \rightarrow \infty$. Furthermore, Pedersen (1995, Theorem 4) has shown that this result implies that, under the true probability measure, the approximate loglikelihood $\ell_n^{(M)}(\boldsymbol{\theta})$ converges in probability, as $M \rightarrow \infty$, to the true loglikelihood $\ell_n(\boldsymbol{\theta})$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $n \in \mathbb{N}$.

These results provide a sensible strategy to approximate the transition density and the loglikelihood. Notice that it is not necessary that $\Delta \rightarrow 0$ for the approximation to work; for Δ fixed, it is possible to improve the quality of the approximation by increasing the

number of subintervals M , because this reduces the length of the intervals over which the approximation has to be computed. The major difficulty in this approach is represented by the evaluation of the $(M - 1)$ -dimensional integral in equation (9). In general, a solution can be found using Monte Carlo integration methods. Formally, if $q(u_1, \dots, u_{M-1})$ is a density on $\mathbb{R}^{M-1}$, then (9) can be rewritten as:

$$p_x^{(M)}(x_\Delta, \Delta | x_0; \theta) = \int_{\mathbb{R}^{M-1}} \frac{\prod_{m=1}^M p_x^{(1)}(u_m, \Delta_m | u_{m-1}; \theta)}{q(u_1, \dots, u_{M-1})} q(u_1, \dots, u_{M-1}) du_1 \dots du_{M-1} \quad (10)$$

If $(u_1^{(k)}, \dots, u_{M-1}^{(k)})$, $k = 1, \dots, K$, denote K independent draws from $q(u_1, \dots, u_{M-1})$, then we can approximate $p_x^{(M)}(x_\Delta, \Delta | x_0; \theta)$ with:

$$p_x^{(M,K)}(x_\Delta, \Delta | x_0; \theta) = \frac{1}{K} \sum_{k=1}^K \frac{\prod_{m=1}^M p_x^{(1)}(u_m^{(k)}, \Delta_m | u_{m-1}^{(k)}; \theta)}{q(u_1^{(k)}, \dots, u_{M-1}^{(k)})} \quad (11)$$

where again $u_0^{(k)} = x_0$ and $u_M^{(k)} = x_\Delta$, $\forall k$. Following the literature on Monte Carlo integration, we refer to $q(u_1, \dots, u_{M-1})$ as an importance sampler. If:

$$\mathbb{E} \left[\frac{\prod_{m=1}^M p_x^{(1)}(u_m, \Delta_m | u_{m-1}; \theta)}{q(u_1, \dots, u_{M-1})} \right] < \infty \quad (12)$$

where the expectation is taken under $q(u_1, \dots, u_{M-1})$, then, by the strong law of large numbers:

$$\lim_{K \rightarrow \infty} |p_x^{(M,K)}(x_\Delta, \Delta | x_0; \theta) - p_x^{(M)}(x_\Delta, \Delta | x_0; \theta)| = 0 \quad \text{a.s.}$$

To approximate the unknown discrete sample loglikelihood $\ell_n(\theta)$, hence, four elements have to be fixed: (i) The subdensity $p_x^{(1)}(u_m, \Delta_m | u_{m-1}; \theta)$; (ii) The importance sampler $q(u_1, \dots, u_{M-1})$; (iii) The number of subintervals M ; and (iv) The number of trajectories used in Monte Carlo integration K . In general, using the approximate transition density $p_x^{(1)}(u_m, \Delta_m | u_{m-1}; \theta)$ induces a bias in $p_x^{(M,K)}(x_\Delta, \Delta | x_0; \theta)$ which can be reduced by increasing M . Apart from being computationally costly, this has the unpleasant effect of increasing the variance of $p_x^{(M,K)}(x_\Delta, \Delta | x_0; \theta)$, unless K is also increased, which makes the whole procedure even more computationally cumbersome. The issue is then to efficiently choose the subdensity and the importance sampler in order to obtain unbiased and sufficiently precise estimates even with low values of M and K .

A very simple strategy has been proposed Pedersen (1995) and Brandt and Santa Clara (2002). Basically, both studies use the subdensity implied by the Euler approximation (8), and choose

$$q(u_1, \dots, u_{M-1}) = \prod_{m=1}^{M-1} p_x^{(1)}(u_m, \Delta_m | u_{m-1}; \theta) \quad (13)$$

With these settings, (10) and (11) simplify considerably. The former becomes:

$$p_x^{(M)}(x_\Delta, \Delta | x_0; \theta) = \int_{\mathbb{R}} p_x^{(1)}(x_\Delta, \Delta_M | u_{M-1}; \theta) p_x^{(M-1)}(u_{M-1}, \Delta - \tau_{M-1} | x_0; \theta) du_{M-1}$$

which is easily interpreted as the expectation of $p_x^{(1)}(x_\Delta, \Delta_M | u_{M-1}; \theta)$ taken over u_{M-1} and w.r.t. the approximation of the distribution conditional on x_0 . Consequently, (11)

becomes:

$$p_x^{(M,K)}(x_\Delta, \Delta | x_0; \theta) = \frac{1}{K} \sum_{k=1}^K p_x^{(1)}(x_\Delta, \Delta_M | u_{M-1}^{(k)}; \theta) \quad (14)$$

where $u_{M-1}^{(k)}$, $k = 1, \dots, K$, is the last $-(M-1)$ -th component drawn from (13).

Durham and Gallant (2002) show that, while straightforward to implement, this approach does not provide acceptable results in terms of unbiasedness and precision, unless M and K (and the resulting computational burden) are very high. In their investigation on diffusion models of the short term interest rate, however, Durham and Gallant (2002) and Durham (2000) show that alternative choices of the subdensity and of the importance sampler allow a large reduction in computational effort. While they explore several alternatives, the most effective strategy, which roughly attains the same degree of unbiasedness and precision with a 10^4 -fold reduction in computation time, entails dropping the Euler subdensity in favor of that implied by a discretization scheme outlined by Shoji and Ozaki (1998), and using an importance sampler labelled “modified Brownian bridge”. While the former is basically an effective bias reduction technique, the latter lowers the Monte Carlo variance by drawing points with higher probability in regions where the integrand is larger. This can be accomplished by recursively simulating K paths according to the following rule:

$$u_m = u_{m-1} + \tilde{\mu}_{m-1} \Delta_m + \tilde{\sigma}_{m-1} \Delta_m^{1/2} \epsilon_m$$

for $m = 1, \dots, M-1$, with:

$$\tilde{\mu}_{m-1} = \frac{x_\Delta - u_{m-1}}{\Delta - \tau_{m-1}} \quad \text{and} \quad \tilde{\sigma}_{m-1} = \left(\frac{M-m}{M-m+1} \right)^{1/2} \sigma(u_{m-1}, \theta)$$

It is easily shown that the importance sampler based on these recursions draws u_m from the approximate distribution based on the Euler discretization of (1) conditional on u_{m-1} and x_Δ .

4. Moment conditions based on the infinitesimal generator

Hansen and Scheinkman (1995) derive moment conditions for estimating and testing diffusion models using discrete time data. Using the infinitesimal generator, it is possible to construct two sets of moments conditions, which can be used to obtain GMM estimators of the parameter vector and diagnostic tests.

Let us start by simplifying the notation by assuming that $\Delta = 1$, and that the diffusion process is stationary. Further, let ϕ be a function in the domain $\mathcal{D}_\mathcal{A}$ of the infinitesimal generator $\mathcal{A}$ of (1). By the law of iterated expectations, it is easy to verify that:

$$\mathbb{E}\{\mathbb{E}[\phi(x_h) | x_0] - \phi(x_0)\} = 0$$

If we divide by h and take the limit for $h \rightarrow 0$, we obtain the first set of moment conditions:

$$\mathbb{E}[\mathcal{A}\phi(x_0)] = 0 \quad (15)$$

Intuitively, this is equivalent to the observation that, under stationarity, $\mathbb{E}[\phi(x_h)]$ is independent of h , and hence its derivative w.r.t. h is zero.

The second set of moment conditions exploits the same property for $E[\phi(x_1)\phi^*(x_0)]$, for $\phi, \phi^* \in \mathcal{D}_{\mathcal{A}}$. To derive it, start by noticing that $\mathcal{A}$ and the conditional expectation operator commute:

$$E[\mathcal{A}\phi(x_1)|x_0] = \mathcal{A}E[\phi(x_1)|x_0]$$

Let us consider the reverse time process, and its associated conditional expectation operator $E[\phi^*(x_0)|x_h]$. Since this operator has the same properties (in reverse time) as the usual conditional expectation operator, we can define the reverse time infinitesimal generator as:

$$\mathcal{A}^*\phi^* = \lim_{h \rightarrow 0} \frac{E[\phi^*(x_0)|x_h] - \phi^*(x_h)}{h}$$

It is easy to verify that the reverse time conditional expectation operator is the adjoint of the usual one, since by the law of iterated expectations we have:

$$E\{\phi^*(x_0)E[\phi(x_1)|x_0]\} = E\{\phi(x_1)E[\phi^*(x_0)|x_1]\}$$

It can be shown that $\mathcal{A}^*$ is also the adjoint of $\mathcal{A}$. This means that:

$$E\{\mathcal{A}E[\phi(x_1)|x_0]\phi^*(x_0)\} = E\{\phi(x_1)\mathcal{A}^*E[\phi^*(x_0)|x_1]\}$$

However, a stationary scalar diffusion process is reversible, and hence $\mathcal{A} = \mathcal{A}^*$; by taking the unconditional expectation of the above expression, we finally have:

$$E[\mathcal{A}\phi(x_1)\phi^*(x_0) - \phi(x_1)\mathcal{A}\phi^*(x_0)] = 0 \quad (16)$$

Given the moment conditions (15) and (16), an estimate of the parameter vector θ in μ and σ may be obtained using a GMM approach. Let us consider for simplicity just the first moment condition. If we select a vector of $m \geq d$ test functions $\phi(x) = [\phi_1(x), \dots, \phi_m(x)]'$, and denote:

$$\phi'(x) = [\phi'_1(x), \dots, \phi'_m(x)]^\top \quad \phi''(x) = [\phi''_1(x), \dots, \phi''_m(x)]^\top$$

then (15) implies that there exists $\theta_0 \in \Theta$ such that:

$$E \left[\mu(x, \theta_0)\phi'(x) + \frac{1}{2} \sigma(x, \theta_0)\phi''(x) \right] = \mathbf{0} \quad (17)$$

The idea is to estimate θ_0 using the empirical counterpart of this expression. The estimation problem can be formalized as:

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \left\| \sum_{i=1}^n \left[\mu(x_i, \theta)\phi'(x_i) + \frac{1}{2} \sigma(x_i, \theta)\phi''(x_i) \right] \right\|_{\Omega}^2$$

where Ω is a positive definite weighting matrix. The second set of moment conditions may be used separately or jointly with the first one along the same lines.

There is still a major point to be made about this approach. The two moment conditions above involve the marginal distribution of x_i (15) and the conditional distribution of x_{i+1} given x_i (16). They seem quite general, because they hold for any functions $\phi, \phi^* \in \mathcal{D}_{\mathcal{A}}$ (which is essentially equivalent to choose $\phi, \phi^* \in \mathcal{C}^2$). However, these conditions are not sufficient to separately identify all the parameters in μ and σ ; instead, Hansen and Scheinkman (1995, Section 5) have shown that (15) and (16) allows to identify the

marginal distribution, while the two set of restrictions jointly allow to identify θ up to a scale factor.

While this approach has not been applied extensively in the literature, it is nevertheless important because it presents several advantages. First, it can be extended to the case of irregular or stochastic sampling; second, unlike many of the previous strategies, it is exact (i.e. it does not require any approximation step); and third, it has paved the way to a stream of subsequent studies which have tried to solve the identification problem and to extend the approach to more general settings using some further properties of the infinitesimal generator.

5. Efficient Method of Moments

The theory of EMM was developed in Gallant and Tauchen (1996) and extended to non-Markovian data with latent variables in Gallant and Long (1997). The EMM approach to the estimation of the parameters of a model of interest (“structural”) suggests to apply GMM, with moment conditions given by the expectation under the structural model of the score of an another (“auxiliary”) model. Such expectation can in general be easily computed by simulation.

In the context of the estimation of θ in the SDE (1), the steps involved in EMM can be outlined as follows. Let us assume that the process defined by (1) is stationary; we simplify the notation as in the previous section by posing $\Delta = 1$. Further, define $\mathbf{z}_{i-1} = (x_{i-1}, \dots, x_{i-l})'$, and suppose that the score generator is given by the parameterized family of conditional transition densities $\{f_x(x_i|\mathbf{z}_{i-1}; \beta); \beta \in \mathbf{B} \subset \mathbb{R}^p\}$, with $p \geq d$. Given a time series of observations $\{x_i; i = 1, \dots, n\}$, the quasi-maximum likelihood estimate $\hat{\beta}$ of the parameters of the score generator is defined by:

$$\hat{\beta} = \arg \max_{\beta \in \mathbf{B}} \frac{1}{n} \sum_{i=l+1}^n \log f_x(x_i|\mathbf{z}_{i-1}; \beta)$$

Recall that we have denoted by $p_x(x_i|x_{i-1}; \theta)$ the transition density implied by the SDE (1). From quasi-maximum likelihood theory, we know that (under regularity conditions) $\hat{\beta} \xrightarrow{a.s.} \beta_0$, where β_0 is defined by the solution of the asymptotic problem:

$$\beta_0 = \arg \max_{\beta \in \mathbf{B}} \int \log f_x(x_i|\mathbf{z}_{i-1}; \beta) p_x(x_i|x_{i-1}; \theta_0) dx_i \quad (18)$$

Furthermore,

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{J}_0^{-1} \mathbf{I}_0 \mathbf{J}_0^{-1})$$

with:

$$\begin{aligned} \mathbf{J}_0 &= \int \frac{\partial^2}{\partial \beta \partial \beta'} \log f_x(x_i|\mathbf{z}_{i-1}; \beta_0) p_x(x_i|x_{i-1}; \theta_0) dx_i \\ \mathbf{I}_0 &= \int \frac{\partial}{\partial \beta} \log f_x(x_i|\mathbf{z}_{i-1}; \beta_0) \frac{\partial}{\partial \beta'} \log f_x(x_i|\mathbf{z}_{i-1}; \beta_0) p_x(x_i|x_{i-1}; \theta_0) dx_i \end{aligned}$$

This expression of $\mathbf{I}_0$ is valid under the assumption that the expectation of the cross product of auxiliary scores at different dates is zero. This condition is met if $f_x(x_i|\mathbf{z}_{i-1}; \beta)$

closely approximates $p_x(x_i|x_{i-1}; \boldsymbol{\theta})$. Since the same condition is necessary for the asymptotic efficiency of the estimator derived below, we assume it is verified. We come back to this point later.

If we define:

$$\mathbf{m}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}) = \int \frac{\partial}{\partial \boldsymbol{\beta}} \log f_x(x_i|\mathbf{z}_{i-1}; \hat{\boldsymbol{\beta}}) p_x(x_i|x_{i-1}; \boldsymbol{\theta}) dx_i$$

then $\mathbf{m}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}) \xrightarrow{a.s.} \mathbf{m}(\boldsymbol{\beta}_0, \boldsymbol{\theta})$ by the dominated convergence theorem and the consistency property of $\hat{\boldsymbol{\beta}}$. Moreover, inspection of the first order conditions of the asymptotic problem (18) reveals that:

$$\mathbf{m}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \frac{\partial}{\partial \boldsymbol{\beta}} \int \log f_x(x_i|\mathbf{z}_{i-1}; \boldsymbol{\beta}_0) p_x(x_i|x_{i-1}; \boldsymbol{\theta}_0) dx_i = \mathbf{0} \quad (19)$$

so that $\mathbf{m}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathbf{0}$. Finally, from the Taylor series expansion of $\mathbf{m}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}_0)$ w.r.t. $\hat{\boldsymbol{\beta}}$ around $\boldsymbol{\beta}_0$, and exploiting the asymptotic normality of $\hat{\boldsymbol{\beta}}$:

$$\sqrt{n} \mathbf{m}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}_0) \quad (20)$$

Equations (19) and (20) allow us to easily devise a GMM-type estimator of $\boldsymbol{\theta}_0$ by looking for the value of the parameters that minimizes the length of the score vector $\mathbf{m}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta})$ w.r.t. to the distance metric associated with $\mathbf{I}_0^{-1}$, the inverse of the asymptotic variance of $\mathbf{m}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}_0)$. More specifically, we define the estimator as the solution to:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \left\| \mathbf{m}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}) \right\|_{\mathbf{I}_0^{-1}}^2 \quad (21)$$

From the asymptotic properties of GMM estimators, we know that $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normal, with:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{M}_0' \mathbf{I}_0^{-1} \mathbf{M}_0)$$

where

$$\mathbf{M}_0 = \frac{\partial}{\partial \boldsymbol{\theta}'} \mathbf{m}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \int \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}'} \log f_x(x_i|\mathbf{z}_{i-1}; \boldsymbol{\beta}_0) p_x(x_i|x_{i-1}; \boldsymbol{\theta}_0) dx_i$$

The implementation of (21) requires two further steps. First, we may consistently estimate $\mathbf{I}_0$ with the corresponding sample average:

$$\hat{\mathbf{I}} = \frac{1}{n} \sum_{i=l+1}^n \frac{\partial}{\partial \boldsymbol{\beta}} \log f_x(x_i|\mathbf{z}_{i-1}; \boldsymbol{\beta}_0) \frac{\partial}{\partial \boldsymbol{\beta}'} \log f_x(x_i|\mathbf{z}_{i-1}; \boldsymbol{\beta}_0)$$

Second, $\mathbf{m}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta})$ can be evaluated by simulation. Specifically, for any given $\boldsymbol{\theta}$, we can simulate a trajectory $\{\tilde{x}_i(\boldsymbol{\theta}); i = 1, \dots, sn\}$ using a discretized version of the SDE (1) and a random number generator, and compute the approximation:

$$\widetilde{\mathbf{m}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}) = \frac{1}{sn} \sum_{i=1}^{sn} \frac{\partial}{\partial \boldsymbol{\beta}} \log f_x[\tilde{x}_i(\boldsymbol{\theta})|\tilde{\mathbf{z}}_{i-1}(\boldsymbol{\theta}); \hat{\boldsymbol{\beta}}]$$

To avoid any discretization bias, the simulated trajectory is initially created by applying the Euler or the Milstein scheme with a small step length, say $1/h$, and then by sampling one every h points. The result is a simulated trajectory of length sn with negligible discretization bias. Let us for simplicity assume that s is integer – i.e., the simulated series is s times longer than the observed one. If h tends to zero at a certain rate as n grows to infinity, then it can be shown that the estimator obtained by minimizing the chi square criterion derived from (21), with $\mathbf{I}_0$ and $\mathbf{m}(\hat{\beta}, \theta)$ respectively substituted by $\hat{\mathbf{I}}$ and $\widetilde{\mathbf{m}}(\hat{\beta}, \theta)$, is consistent and has the same asymptotic distribution as $\hat{\theta}$, except that the asymptotic variance is multiplied by the factor $(1 + 1/s) > 1$. Since s can be chosen arbitrarily large, we assume in the following that s is infinite, so that the variance inflating factor equals 1. Finally, $\mathbf{M}_0$ can be consistently estimated with:

$$\widetilde{\mathbf{M}} = \frac{\partial}{\partial \theta'} \widetilde{\mathbf{m}}(\hat{\beta}, \hat{\theta}) = \frac{1}{sn} \sum_{i=l+1}^{sn} \frac{\partial^2}{\partial \beta \partial \theta'} \log f_x[\tilde{x}_i(\hat{\theta}) | \tilde{z}_{i-1}(\hat{\theta}); \hat{\beta}]$$

Hypothesis testing and model assessment can be conducted using standard GMM recipes. A natural question is how to choose the auxiliary density $f_x(x_i | z_{i-1}, \beta)$. This is a crucial issue, because Gallant and Tauchen (1996) have shown that if the score generator encompasses $p_x(x_i | x_{i-1}; \theta)$, then $\hat{\theta}$ is as efficient as maximum likelihood. Essentially, this means that the closer the fitted auxiliary score is to the unknown true transition density, the smaller the difference between the asymptotic variances of $\hat{\theta}$ and of the ML estimator, with the difference being zero when $f_x(x_i | z_{i-1}, \beta)$ and $p_x(x_i | x_{i-1}; \theta)$ coincide. Gallant and Long (1997) have developed a systematic strategy to meet this condition, based on a particular family of auxiliary models, named SNP (for SemiNonParametric). The SNP density is based on the Hermite expansion of the density associated to the innovations in x_i .

6. Conclusions

Recently, huge advances have been made towards a complete solution of the problem of estimating the parameters of a continuous time diffusion process from a discretely sampled trajectory. In this paper, we have briefly surveyed four of them that appear more promising of future developments. The differences between them are both theoretical and practical. For example, the likelihood approximating strategy outlined in section 2 is probably the best in term of precision and computational burden for scalar diffusions without jump components (see Jensen and Poulsen, 2000, for a numerical experiment), but it seems difficult to devise an extension to multivariate settings with partial observability of the state variables. This kind of problem can, at least in part, be solved by the SML approach of section 3 (see the applications in Durham and Gallant (2002)), although there are some computational and theoretical aspects that need further work. The most complete strategy is probably the EMM approach of section 5, which has already been successfully applied to models much more complicated than the scalar diffusion we considered in this paper. Finally, the GMM approach based on the infinitesimal generator presented in section 4 is the only one which does not require any of the approximation steps characterizing the other strategies, and it has been extended to more general settings by Hansen, Scheinkman and Touzi (1998), Chen, Hansen and Scheinkman (2000) and Darolles, Florens and Renault (1998). Other available strategies involve MCMC Bayesian

estimation, or nonparametric and semiparametric approaches. Research on this topic is ongoing, and further interesting results will no doubt appear in the near future.

References

- Aït-Sahalia Y. (1999) Transition densities for interest rate and other nonlinear diffusion, *Journal of Finance*, 54, 1361–1395.
- Aït-Sahalia Y. (2002) Maximum likelihood estimation of discretely sampled diffusions: A closed-form approach, *Econometrica* (forthcoming).
- Basawa I. and Prakasa Rao B. (1980) *Statistical Inference for Stochastic Processes*, Academic Press, London.
- Brandt M. and Santa Clara P. (2002) Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets, *Journal of Financial Economics* (forthcoming).
- Chen X., Hansen L. and Scheinkman J. (2000) Principal components and the long run, *mimeo*, Dept. of Economics, University of Chicago.
- Darolles S., Florens J. and Renault E. (1998) Nonlinear principal components and inference on a conditional expectation operator, *mimeo*, CREST, Paris.
- Durham G. (2000) Likelihood-based specification analysis of continuous-time models of the short-term interest rate, *mimeo*, Department of Economics, University of Iowa.
- Durham G. and Gallant R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes, *Journal of Business and Economic Statistics*, forthcoming.
- Gallant R. and Long J. (1997) Estimating stochastic differential equations efficiently by minimum chi square, *Biometrika*, 84, 125–141.
- Gallant R. and Tauchen G. (1996) Which moments to match?, *Econometric Theory*, 12, 657–681.
- Hansen L. and Scheinkman J. (1995) Back to the future: Generating moment implications for continuous-time markov processes, *Econometrica*, 63, 767–804.
- Hansen L., Scheinkman J. and Touzi N. (1998) Spectral methods for identifying scalar diffusions, *Journal of Econometrics*, 86, 1–32.
- Jensen B. and Poulsen R. (2000) Transition densities of diffusion processes: a review and numerical comparison of approximation techniques, *mimeo*, University of Copenhagen.
- Kloeden P. and Platen E. (1992) *Numerical Solutions to Stochastic Differential Equations*, Springer, New York.
- Pedersen A. (1995) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scandinavian Journal of Statistics*, 22, 55–71.
- Shoji I. and Ozaki T. (1998) Estimation for non-linear stochastic differential equations by a local linearization method, *Stochastic Analysis and Applications*, 16, 733–752.