N° 9405

KERNEL M-ESTIMATORS :

Non parametric diagnostics for structural models

Christian GOURIEROUX*
Alain MONFORT**
C. TENREIRO***

February 1994

* CREST-CEPREMAP

** CREST

*** Coïmbra University

We are grateful to E. Guerre for many helpful comments.

Kernel M-Estimators :
Non parametric diagnostics for structural models

Abstract

We consider kernel M-estimators, defined as a solution of an optimization problem of the kind :

$$\hat{\theta}_T(s) = Arg\, min_\theta \sum_{t=1}^{T} \frac{1}{h_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \psi(Y_t, \hat{\alpha}_T; s; \theta),$$

where $\hat{\alpha}_T$ tends to a limit α_∞, when T tends to infinity, and d is the dimension of the conditioning variable. We study the asymptotic properties, consistency and asymptotic normality, of such a functional estimator. Then we derive some approximation of $\hat{\theta}_T$ in a neighbourhood of the hypothesis $\{lim_T\, \hat{\theta}_T(s) = o, \forall s\}$, and explain how to use it for non parametric diagnostics in a structural framework.

M-Estimateurs à Noyau :
diagnostics non paramétriques pour des modèles structurels

Résumé

Nous considérons des M-estimateurs à noyau définis comme solution d'un problème d'optimisation du type :

$$\hat{\theta}_T(s) = Arg\, min_\theta \sum_{t=1}^{T} \frac{1}{h_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \psi(Y_t, \hat{\alpha}_T; s; \theta),$$

où $\hat{\alpha}_T$ tend vers une limite α_∞, lorsque T tend vers l'infini, et d est la dimension de la variable conditionante. Nous donnons les propriétés asymptotiques de tels estimateurs fonctionnels. Puis nous introduisons une approximation de $\hat{\theta}_T$ au voisinage de l'hypothèse $\{lim_T\, \hat{\theta}_T(s) = o, \forall s\}$, et nous expliquons comment l'utiliser pour effectuer des recherches non paramétriques de spécifications paramétriques.

Keywords : Kernel, functional estimation, M-estimators,

Mots clés : Noyau, estimation fonctionnelle, M-estimateurs

JEL : C14

I - INTRODUCTION

Functional estimation is generally concerned with the determination of probability density functions or regression functions from a set of observations. The classical non parametric estimators may be built using either kernels or bases of orthogonal functions. These nonparametric approaches are usually considered as useful either for pure descriptive purposes, or as a preliminary step before some more structural modelling. The aim of this paper is to explain how these nonparametric methods might be used jointly with structural models.

The idea of such a joint approach can be explained from the example of an option pricing formula. Let us recall that an european call option based on the asset S maturing at time T, with strike K, is characterized by its terminal value $Max(0, S_T - K)$ at time T. This option has a price P(t,T,K) at time t. Under a set of conditions coucerning the price evolution of the underlying asset (the price of the asset at time t, S_t, is a geometric brownian notion, i.e. is such that $dS_t/S_t = \mu\, dt + \sigma dW_t$), the evolution of the interest rate (assumed to be constant $r_t = r$), the completeness of the market (assumption of continuous trading, and dependence of the option price on the single factor (W_t)), it is known [Black-Scholes (1973)] that the option price is given by the Black-Scholes (B.S)formula :

$$
\begin{aligned}
(1.1)\ & P(t,T,K) \\
&= S_t \left\{ \Phi \left[\frac{1}{\sigma\sqrt{T-t}} Log\, \frac{S_t}{K\, exp - r(T-t)} + \sigma \frac{\sqrt{T-t}}{2} \right] \right. \\
&\left. + \frac{-K\, exp - r(T-t)}{S_t} \Phi \left[\frac{1}{\sigma\sqrt{T-t}} Log\, \frac{S_t}{K\, exp - r(T-t)} - \sigma \frac{\sqrt{T-t}}{2} \right] \right\}. \\
&= g(t,T,K,r,\sigma)\, (say),
\end{aligned}
$$

where Φ is the cumulative distribution function of N(0,1).

The set of conditions underlying this B.S. pricing formula is clearly very restrictive and is generally not satisfied in practice. However such a misspecification does not necessarily imply that the B.S. pricing formula itself, which is a byproduct of the underlying model, must be rejected. Indeed such a formula may be considered has a first way of comparing option prices corresponding to different residual maturities T-t, and different strikes K. With such an interpretation the normalized parameter σ is not necessarily a volatility, but has to be fixed in order to get the best pricing. For instance if we have observations on the prices $P[t_i, T_i, K_i]$ of different options, i=1...,n based on the same underlying asset, and observations on the prices S_{t_i}, and on the interest rates r_{t_i}, i=1,...,n, we may look for a pricing formula with a parameter $\hat{\sigma}_n$ solution of the optimisation problem :

$$(1.2) \quad \hat{\sigma}_n = Arg\ min_\sigma \sum_{i=1}^{n} [P[t_i, T_i, K_i] - g[t_i, T_i, K_i, r_{t_i}, \sigma]]^2$$

$$= Arg\ min_\sigma \sum_{i=1}^{n} \psi[Y_i, \sigma] \text{ (say)},$$

where Y_i denotes the set of observed variables $P[t_i, T_i, K_i], S_{t_i}, r_{t_i}$. With such an approach the pricing parameter σ is implicitely assumed to be constant.

Now this B.S. approach may be improved along two different lines. The first idea is to modify the underlying assumptions concerning the evolution of S_t, r_t... and to rederive a new pricing formula corresponding to this new underlying model. It is the usual approach considered in financial theory, which generally leads to analytically untractable pricing formula except in some unrealistic framework [Hull-White (1987), Wiggins (1987)]. The second idea is to directly modify the pricing formula, without considering the form of a potential underlying model. For instance, we may introduce in the B.S. formula a stochastic parameter funtion of a lagged value of the price of the underlying asset, i.e. consider a pricing formula of the kind :

$$(1.3) \qquad P[t, T, K] = g[t, T, K, r_t, \sigma(S_{t-1})].$$

If we have no a priori idea on the function $S_{t-1} \to \sigma(S_{t-1})$, we may look for it through a kernel method. Let us introduce a kernel K, and h_n a sequence of real numbers tending to zero, when n goes to infinity. We may estimate the "non constant" part of the pricing parameter evaluated at s by solving :

$$(1.4) \qquad \hat{\theta}_n(s) = Arg\ min_\theta \sum_{i=1}^{n} \frac{1}{h_n} K\left(\frac{S_{t_i-1} - s}{h_n}\right)$$

$$[P(t_i, T_i, K_i) - g(t_i, T_i, K_i, r_{t_i}, \hat{\sigma}_n + \theta)]^2$$

$$= Arg\ min_\theta \sum_{i=1}^{n} \frac{1}{h_n} K\left(\frac{S_{t_i-1} - s}{h_n}\right) \psi[Y_i, \hat{\sigma}_n + \theta].$$

Then we can test if the pricing parameter is independent of the past value of the price by checking if the function $\hat{\theta}_n$ is sufficiently close to zero. In summary the idea is first to exhibit a so-called implicit parameter (the implicit volatility in our example), then to directly analyse the structural evolution of this parameter. A similar example applied to implicit hedging parameters is discussed in Gouriéroux-Laurent (1994).

The main difference between the previous example and the classical applications of functional estimation is that the kernel estimation is not directly applied to a conditional regression, but is applied to a structural parameter, the (pricing or implicit) volatility parameter in the B.S. example. For such kernel estimators it will be necessary to develop an asymptotic theory taking into account jointly the nonparametric and the parametric (structural) features of the problem.

In section 2, we consider the so-called **kernel M-estimators**. They are defined as a solution of an optimisation problem of the kind :

$$(1.5) \qquad \hat{\theta}_T(s) = Arg\ min_\theta \sum_{t=1}^{T} \frac{1}{h_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \psi(Y_t, \hat{\alpha}_T; s; \theta),$$

where $\hat{\alpha}_T$ goes to a limit α_∞, when T goes to infinity, and d is the dimension of the conditioning variable S. In practice $\hat{\alpha}_T$ will often be a preliminary parametric estimator of a parameter of interest. This estimator may also appear in the "conditioning variable" S, for instance when we analyse the dependence of some parameters on error terms, which are replaced by first step residuals. Then we derive the asymptotic properties : consistency and asymptotic normality, of such functional estimators.

In section 3, we are interested in analyzing an hypothesis of parameter constancy, which after some normalization is equivalent to :

$$H_o : \{plim_T\, \hat{\theta}_T(.) = 0\}.$$

We first exhibit another functional estimator $\tilde{\theta}_T(.)$, which is asymptotically equivalent to $\hat{\theta}_T(.)$ under the null hypothesis and which admits an analytical expression. These estimators $\hat{\theta}_T$ and $\tilde{\theta}_T$ may then be compared to zero, in order to check the hypothesis H_o. Different applications are presented ; they concern the linearity of a regression function, the necessity to introduce an additional variable in a regression, the quadratic form of the volatility in ARCH models...

Section 4 is devoted to a simulation exercise to illustrate the usefulness of the approach. The nonparametric diagnostic is applied to check the presence of cross effects in a regression model, and to derive the form of these effects. It is in particular interesting to note that such a study may not be always performed from classical methods based on regression residuals.

The technical proofs of the asymptotic properties of the kernel M-estimators are gathered in the appendices.

II - KERNEL M-ESTIMATORS

These estimators are defined as a solution of the optimization problem :

$$(2.1) \qquad \hat{\theta}_T(s) = Arg\ min_\theta \sum_{t=1}^{T} \frac{1}{h_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \psi[Y_t, \hat{\alpha}_T; s; \theta],$$

where S and ψ are known functions with dimensions d and 1 defining the conditioning variables and the parametric objective function respectively, K is a kernel, Y_t t=1,...,T, are observed variables, and $\hat{\alpha}_T$ a sequence of random variables tending to a deterministic limit α_∞, when T tends to infinity.

Such kernel M-estimators have already been introduced for some particular criterion functions ψ. When $\psi_t = (Y_t - \theta)^2$, we get the usual kernel estimator of the regression function $E[Y_t/S_t = s]$ (see Hardle (1990) p. 9-20). When $\alpha = 1$, and $\psi_t = [Y_t - \theta_o - \theta_1(S_t - s)... - \theta_p(S_t - s)^p]^2$, we get the locally weighted least squares regression which locally fits a polynomial of degree p to the data [see Cleveland (1979) for the theoretical properties of this approach, Gouriéroux-Scaillet (1994) for applications to the term structure of interest rates]. Robinson (1984) considered the M-estimator of a location parameter for which $\psi_t = \rho(Y_t - \theta)$. Some applications to log-likelihood or pseudo log-likelihood criterion functions have been considered by Tibshirani-Hastie (1987), Staniswalis (1989), Fan-Heckman-Wang (1992), and applied to probit models and to survival data models.

The asymptotic properties of such M-estimators can be deduced in the usual way [Jennrich (1969), Huber (1981), Burguete-Gallant-Souza (1982), White (1981), Gallant (1987), Gouriéroux-Monfort (1989)] from the asymptotic properties of kernel regression estimators [see e.g. Nadaraya (1964), Watson (1964), Bierens (1987), Gasser-Muller (1979), Mack-Silverman (1982)] applied to the objective function ψ and to its first and second order derivatives with respect to θ. In this section we essentially describe the main steps of the proof and some set of regularity conditions. The details of the proofs are given in the appendices.

Consistency

There are different kinds of regularity conditions on the observable process, on the estimator $\hat{\alpha}_T$, on the kernel and on the bandwidth h_T in order to ensure consistency . They are basically the following ones (see appendix 1 for more precise and less restrictive conditions).

Conditions on the observable process :

$(Y_t, t \epsilon Z)$ is a strictly stationary process, satisfying some mixing conditions. Therefore the results will be valid for correlated observations, as soon as this correlation is not too large [see Altman (1988), Giorfi-Hardle-Sarda-Vieu (1989), Roussas (1969)].

Conditions on the kernel :

The kernel K is continuous, with bounded support, and assigns mass one to $\mathbb{R}^d$.

Conditions on the bandwidth :

$h_T \rightarrow 0$, $Th_T^d \rightarrow +\infty$, when $T \rightarrow \infty$.

Conditions on the estimator $\hat{\alpha}_T$:

$\hat{\alpha}_T$ converges almost surely to a deterministic value α_∞, and

$$\hat{\alpha}_T - \alpha_\infty = o\left(h_T^{1+\frac{d}{2}}\right).$$

Conditions on the objective function ψ :

ψ is continuously third order differentiable with respect to θ, and satisfies suitable integrability conditions. S, ψ and $\frac{\partial \psi}{\partial \theta}$ are continuously differentiable with respect to α.

Asymptotic identifiability conditions.

There exists a unique solution $\theta_\infty(s)$ to the optimization problem:

$$Min_\theta \; E[\psi(Y, \alpha_\infty; s; \theta)/S(Y, \alpha_\infty) = s].$$

Under such regularity conditions, we can first study the asymptotic behaviour of the normalized objective function :

$$L_T(s;\theta) = \sum_{t=1}^{T} \frac{1}{Th_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \psi(Y_t, \hat{\alpha}_T; s; \theta).$$

We have asymptotically :

$$\begin{aligned}
\lim_{as} L_T(s;\theta) &= \lim_{as} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{h_T^d} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \psi(Y_t,\alpha_\infty;s;\theta) \\
&= \lim E\left\{\frac{1}{h_T^d} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \psi(Y_t,\alpha_\infty;s;\theta)\right\} \\
&= \lim E\left\{\frac{1}{h_T^d} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] E[\psi(Y,\alpha_\infty;s;\theta)/S(Y,\alpha_\infty)]\right\} \\
&= E[\psi(Y,\alpha_\infty;s;\theta)/S(Y,\alpha_\infty)=s]\, f_S(s) \\
&= L_\infty(s;\theta)(\text{say}),
\end{aligned}$$

where f_S is the marginal density of $S(Y,\alpha_\infty)$.

If some additional regularity conditions are also introduced in order to ensure that this almost sure convergence is uniform in θ, for $\theta \in \Theta$, and if Θ is compact, we deduce from a classical Jennrich's argument [see Jennrich (1969)] that there asymptotically exists a solution $\hat{\theta}_T(s)$ of the finite sample problem $Min_\theta\, L_T(s;\theta)$, which is a consistent estimator of the unique solution $\theta_\infty(s)$ of the asymptotic optimization problem $Min_\theta\, L_\infty(s;\theta)$. These additional sufficient regularity conditions may either have the form of a Cramer's condition concerning the uniform integrability of the derivatives of the criterion function up to order 3 (see Staniswalis (1987)), or of a separation between parameters and observations in the criterion function [see Andrews (1990)]. The uniform convergence may also be directly verified in some particular applications where the criterion function is convex in the parameter. A set of such conditions is described in appendix 1.

Asymptotic normality

Then if the limit $\theta_\infty(s)$ is in the interior of the parameter set $\Theta, \hat{\theta}_T(s)$ will asymptotically satisfy the first order condition :

$$(2.2) \qquad \sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T^d}} K\left[\frac{S(Y_t,\hat{\alpha}_T)-s}{h_T}\right] \frac{\partial\psi}{\partial\theta}(Y_t,\hat{\alpha}_T;s;\hat{\theta}_T(s)) = 0.$$

This system may be expanded around $\hat{\alpha}_T \simeq \alpha_\infty$ and $\hat{\theta}(s) \simeq \theta_\infty(s)$. We just give here the main steps of the proof which is detailed in appendix 2. We get :

$$\sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T^d}} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial\psi}{\partial\theta}[Y_t,\alpha_\infty;s;\theta_\infty(s)]$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T^d}} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial^2\psi}{\partial\theta\partial\theta'}[Y_t,\alpha_\infty;s;\theta_\infty(s)][\hat{\theta}_T(s)-\theta_\infty(s)]$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T^d}} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial^2\psi}{\partial\theta\partial\alpha'}[Y_t,\alpha_\infty;s;\theta_\infty(s)](\hat{\alpha}_T-\alpha_\infty)$$

$$+\sum_{j=1}^{d}\sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T^d}} \frac{\partial\psi}{\partial\theta}[Y_t,\alpha_\infty;s;\theta_\infty(s)]\frac{1}{h_T}\frac{\partial K}{\partial u_j}\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right]$$

$$\times\frac{\partial S_j}{\partial\alpha'}[Y_t,\alpha_\infty](\hat{\alpha}_T-\alpha_\infty)\sim 0,$$

where the residual term has an order strictly smaller than the maximal order of the four terms of the L.H.S.

This relation may also be written as :

$$(2.3)\quad \sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T^d}} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial\psi}{\partial\theta}[Y_t,\alpha_\infty;s;\theta_\infty(s)]$$

$$+\sum_{t=1}^{T} \frac{1}{Th_T^d} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial^2\psi}{\partial\theta\partial\theta'}[Y_t,\alpha_\infty;s;\theta_\infty(s)]\sqrt{T}\sqrt{h_T^d}(\hat{\theta}_T(s)-\theta_\infty(s))$$

$$+\sum_{t=1}^{T} \frac{1}{Th_T^d} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial^2\psi}{\partial\theta\partial\alpha'}[Y_t,\alpha_\infty;s;\theta_\infty(s)]\sqrt{T}\sqrt{h_T^d}(\hat{\alpha}_T-\alpha_\infty)$$

$$+\sum_{j=1}^{d}\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T^d}} \frac{\partial K}{\partial u_j}\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial\psi}{\partial\theta}[Y_t,\alpha_\infty;s;\theta_\infty(s)]$$

$$\times\frac{\partial S_j}{\partial\alpha'}(Y_t,\alpha_\infty)\frac{1}{h_T}(\hat{\alpha}_T-\alpha_\infty)\sim 0,$$

or:

$$\sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T^d}} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial\psi}{\partial\theta}\left[Y_t,\alpha_\infty;s;\theta_\infty(s)\right]$$

$$+ f_S(s) E\left[\frac{\partial^2\psi}{\partial\theta\partial\theta'}[Y,\alpha_\infty;s;\theta_\infty(s)]/S(Y,\alpha_\infty) = s\right] \sqrt{T}\sqrt{h_T^d}(\hat\theta_T(s)-\theta_\infty(s))$$

$$+ f_S(s) E\left[\frac{\partial^2\psi}{\partial\theta\partial\alpha'}[Y,\alpha_\infty;s;\theta_\infty(s)]/S(Y,\alpha_\infty) = s\right] \sqrt{h_T^d}\sqrt{T}(\hat\alpha_T-\alpha_\infty)$$

$$+ \sum_{j=1}^{d}\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T^d}} \frac{\partial K}{\partial u_j}\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right]$$

$$\frac{\partial\psi}{\partial\theta}\left(Y_t;\alpha_\infty;s;\theta_\infty(s)\right) \frac{\partial S_j}{\partial\alpha'}(Y_t;\alpha_\infty) \frac{1}{h_T}(\hat\alpha_T-\alpha_\infty) \sim 0$$

In the previous expansion we have distinguished the two terms (the first and fourth ones) for which the factors have a zero conditional expectations and the two other ones. Indeed for the first one we have :

$$E\left[K\left(\frac{S_t-s}{h_T}\right)\frac{\partial\psi}{\partial\theta}[Y_t;\alpha_\infty;s;\theta_\infty(s)]/S_t = s\right] = 0,$$

$$\Longleftrightarrow E\left[\frac{\partial\psi}{\partial\theta}[Y_t;\alpha_\infty;s;\theta_\infty(s)]/S_t = s\right] = 0$$

since $\theta_\infty(s)$ gives the optimum of the asymptotic criterion function. For the fourth one we have to develop a special treatment bacause the kernel only appears through its derivatives.

We now have to introduce some conditions implying that the replacement of $\hat\alpha_T$ by α_∞ has no effect on the asymptotic distribution of the kernel estimator, i.e. to eliminate the two last terms of the expansion. The third one is negligible if the speed of convergence of $\hat\alpha_T - \alpha_\infty$ to zero is large enough.

condition(i) on the speed of convergence of $\hat\alpha_T$

$$\hat\alpha_T - \alpha_\infty = o\left(\frac{1}{\sqrt{Th_T^d}}\right).$$

It is seen that this condition is also sufficient to eliminate the last term (see appendix 2). This result is different from what is usually obtained in a parametric framework, in which an additional condition is needed to eliminate the effects of first step estimator.

Another condition has to be introduced to eliminate another term (see appendix 2).

condition(ii) on the speed of convergence of $\hat{\alpha}_T$

$$\hat{\alpha}_T - \alpha_\infty = O(h_T^{d+2}).$$

Under these additional conditions the previous expansion reduces to :

$$\sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T^d}} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial\psi}{\partial\theta}[Y_t,\alpha_\infty;s;\theta_\infty(s)]$$

$$+ f_S(s) E\left[\frac{\partial^2\psi}{\partial\theta\partial\theta'}[Y,\alpha_\infty;s;\theta_\infty(s)]/S(Y,\alpha_\infty) = s\right] \sqrt{T}\sqrt{h_T^d}(\hat{\theta}_T(s)-\theta_\infty(s)] \sim 0.$$

We can now apply the usual kernel regression estimation theory, which ensures that :

$$\sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T^d}} K\left[\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right] \frac{\partial\psi}{\partial\theta}[Y_t,\alpha_\infty;s;\theta_\infty(s)],$$

is of order 1, asymptotically normal with zero mean, and a covariance matrix given by :

$$\Omega(s) = \int K^2(v)dv\, f_S(s)C_o(s), \tag{2.8}$$

where :

$$C_o(s) = E\left[\frac{\partial\psi}{\partial\theta}[Y,\alpha_\infty;s;\theta_\infty(s)]\frac{\partial\psi}{\partial\theta'}[Y,\alpha_\infty;s;\theta_\infty(s)]/S(Y,\alpha_\infty) = s\right].$$

This result is valid under a set of regularity conditions discussed in appendix 2 ; among them the most important one concerns the bandwidth.

condition on the bandwidth :

$$h_T = o(T^{-1/(d+4)}).$$

condition on the kernel :

the kernel is of order one :

$$\int u_j \, K(u)du = o, \forall j.$$

It has to be noted that the asymptotic variance-covariance matrix does not contain any covariance term associated with lagged values of the observable process. Indeed these terms are negligible with respect to $C_o(s)$. [see the appendix].

$\sqrt{T}\sqrt{h_T^d}[\hat{\theta}_T(s) - \theta_\infty(s)]$ is of order 1 and and we have the following result.

Property 1 : The kernel M-estimator is consistent, asymptotically normal :

$$\sqrt{T}\sqrt{h_T^d}[\hat{\theta}_T(s) - \theta_\infty(s)] \xrightarrow{d} N[0, \sum(s)],$$

where :

$$\sum(s) = \frac{J(s)^{-1}C_o(s)J(s)^{-1}}{f_S(s)} \int K^2(v)dv,$$

$$J(s) = E\left[\frac{-\partial^2\psi}{\partial\theta\partial\theta'}[Y, \alpha_\infty; s; \theta_\infty(s)]/S(Y, \alpha_\infty) = s\right].$$

We obtain the classical expression $J(s)^{-1}\, C_o(s)J(s)^{-1}$ of the asymptotic covariance matrix of a parametric M-estimator (see White (1981)), but in which each matrix is evaluated conditionally to $S(Y, \alpha_\infty) = s$, and this term is multiplied by the kernel factor $\int K^2(v)dv$ and the density effect.

Since all the matrices J, C_o are evaluated conditionally to S = s, we immediately deduce the following corollary, for criterion functions of the form:

$$\psi[Y_t, \hat{\alpha}_T; s; \theta] = \tilde{\Psi}[Y_t, S(Y_t, \hat{\alpha}_T); \hat{\alpha}_T; s; \theta].$$

Corollary 2 : The estimators $\hat{\theta}_T(s)$ and $\hat{\hat{\theta}}_T(s)$ solutions of :

$$\hat{\theta}_T(s) = Arg\, min_\theta \sum_{t=1}^{T} \frac{1}{Th_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \tilde{\psi}(Y_t, S(Y_t, \hat{\alpha}_T); \hat{\alpha}_T; s; \theta),$$

$$\hat{\hat{\theta}}_T(s) = Arg\, min_\theta \sum_{t=1}^{T} \frac{1}{Th_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \tilde{\psi}[Y_t, s, \hat{\alpha}_T; s; \theta),$$

have the same asymptotic properties.

This result is natural since the kernel focusses on the value S = s, and therefore it is possible to replace $S(Y_t, \hat{\alpha}_T)$ by this value in the objective function.

Remark 2.8 : In particular the estimators obtained by :

$$\hat{\theta}_T(s) = Arg\, min_\theta \sum_{t=1}^{T} \frac{1}{Th_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \omega(S(Y_t, \hat{\alpha}_T))\psi[Y_t, \hat{\alpha}_T; s; \theta],$$

$$\begin{aligned} \hat{\hat{\theta}}_T(s) &= Arg\, min_\theta \sum_{t=1}^{T} \frac{1}{Th_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \omega(s)\psi[Y_t, \hat{\alpha}_T; s; \theta] \\ &= Arg\, min_\theta \sum_{t=1}^{T} \frac{1}{Th_T^d} K\left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right] \psi[Y_t, \hat{\alpha}_T; s; \theta], \end{aligned}$$

have the same asymptotic properties. Therefore it is asymptotically useless to introduce a weighting function only depending on S.

Example 2.9 : Estimation of a regression function.

In a parametric framework it is known [Gouriéroux-Monfort-Trognon (1984), Gouriéroux-Monfort (1993)] that the parameters appearing in a conditional mean can be consistently estimated by a weighted pseudo-maximum likelihood method based on a linear exponential family. The same result is valid for kernel M-estimators. Let us consider a family of distributions with p.d.f. :

$$f(y; m) = exp[A(m) + B(y) + C(m)y],$$

where m is the mean of the distribution. This implies $\frac{dA}{dm}(m) + m\,\frac{dC}{dm}(m) = 0.$

We may introduce a functional estimator of the regression function of Y given another variable X (univariate in the example) by :

$$(2.9) \qquad \hat{m}_T(x) = Arg\,max_m \sum_{t=1}^{T} \frac{1}{h_T} K\left[\frac{X_t - x}{h_T}\right] \omega(X_t)\, Log\, f(Y_t, m)$$

$$= Arg\,max_m \sum_{t=1}^{T} \frac{1}{h_T} K\left[\frac{X_t - x}{h_T}\right] \omega(X_t)\, [A(m) + C(m)Y_t].$$

We have :

$$\frac{\partial \psi(Y,m)}{\partial m} = \frac{dC(m)}{dm} [Y_t - m]\omega(X_t),$$

$$\frac{\partial^2 \psi(Y,m)}{\partial m^2} = \left[\frac{d^2C(m)}{dm^2}(Y_t - m) - \frac{dC(m)}{dm}\right] \omega(X_t).$$

We deduce that :

$$C_o(x) = E\left[\left(\frac{\partial \psi}{\partial m}(Y, m(x))\right)^2 / X = x\right] = \omega(x)^2 V(Y/X = x) \left(\frac{dC}{dm}(m(x))\right)^2,$$

$$J(x) = E\left[-\frac{\partial^2 \psi}{\partial m^2}(Y, m(x))/X = x\right] = \omega(x) \frac{dC}{dm}(m(x)).$$

Therefore the different kernel P.M.L. estimators have the same asymptotic distribution, which is independent of the chosen linear exponential family :

$$\sqrt{T}\sqrt{h_T}[\hat{m}_T(x) - m(x)] \xrightarrow{d} N\left[0, \frac{V(Y/X = x)}{f(x)} \int K^2(v)dv\right],$$

and corresponds to the usual asymptotic distribution of the kernel regression estimator (since the normal family of distributions is a linear exponential family). Of course their finite sample properties are not the same.

III - LOCAL KERNEL M-ESTIMATORS

In this section we want to use kernel M-estimators in order to ckeck the constraint defined by :

$$(3.1) \qquad H_o = \{\theta_\infty(s) = 0, \forall s\}.$$

Such a constraint will be the basis for testing the constancy of a functional parameter θ with respect to a variable S. In practice this null hypothesis corresponds to the fact that a given parametric model of interest is well specified.

It is of course possible to compare the estimator $\hat{\theta}_T(.)$, introduced in the previous section to the zero benchmark, and such a descriptive approach, which allows to detect the values of s which possibly cause the rejection of the hypothesis, seems to be preferable to the use of an aggregated test statistics, such as a mean square error based on $\hat{\theta}_T \, : \, \int_s \hat{\theta}_T(s)^2 ds$. However the determination of function $\hat{\theta}_T$ may be difficult in practice since it requires the optimisation of the objective function for a large number of s values. Therefore it is useful to propose an estimator easy to compute and asymptotically equivalent to $\hat{\theta}_T(.)$ under the null. Such an estimator is directly deduced from the expansions of the previous section.

Property 3 : The estimator :

$$\tilde{\theta}_T(s) = \left\{ \sum_{t=1}^{T} K \left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T} \right] \frac{\partial^2 \psi}{\partial \theta \partial \theta'} (Y_t, \hat{\alpha}_T; s; 0) \right\}^{-1}$$

$$\sum_{t=1}^{T} K \left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T} \right] \frac{\partial \psi}{\partial \theta} (Y_t, \hat{\alpha}_T; s; 0),$$

is asymptotically such that :

$$\tilde{\theta}_T(s) - \hat{\theta}_T(s) = o\left(\frac{1}{\sqrt{T}\sqrt{h_T^d}} \right),$$

under the null hypothesis.

The advantage of this modified estimator is to have an analytical expression, and to avoid the optimizations. Moreover this estimator is equal to :

$$\tilde{\theta}_T(s) = \left\{ \sum_{t=1}^{T} K \left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T} \right] \frac{\partial^2 \psi}{\partial \theta \partial \theta'} (Y_t, \hat{\alpha}_T; s; 0) \right\}^{-1}$$

$$\sum_{t=1}^{T} K \left[\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T} \right] \frac{\partial^2 \psi}{\partial \theta \partial \theta'} (Y_t, \hat{\alpha}_T; s; 0) \hat{v}_t(s),$$

$$\text{where : } \hat{v}_t(s) = \left[\frac{\partial^2 \psi}{\partial\theta\partial\theta'}(Y_t, \hat{\alpha}_T; s; 0)\right]^{-1} \frac{\partial\psi}{\partial\theta}(Y_t, \hat{\alpha}_T; s; 0),$$

is a normalized score. Therefore the function $\tilde{\theta}_T$ appears as a kernel smoothing of these residuals $\hat{v}(s)$ (especially when s do not appear in the ψ function). To check if $\tilde{\theta}_T$ is close to zero is equivalent to check if the estimated conditional expectation $E(\hat{v}_t/S_t = s)$ is close to zero.

We shall now describe some problems for which such a nonparametric diagnostic tool should be useful.

Example 3.2 : Linearity of a conditional regression.

To check if a regression function is linear, i.e. $H_o = \{\exists a, b : E(Y/X) = aX + b\}$, we may first estimate the linear model :

$$Y_t = aX_t + b + u_t,$$

and obtain some parametric estimators $\hat{a}_T, \hat{b}_T$ of a and b. Then we may consider the problem :

$$Min_\theta \sum_{t=1}^{T} K\left[\frac{X_t - x}{h_T}\right][Y_t - \hat{a}_T X_t - \hat{b}_T - \theta]^2,$$

i.e compute the kernel regression estimator of the residual term $\hat{u}_t = Y_t - \hat{a}_T X_t - \hat{b}_T$ on the explanatory variable X_t. In such a problem the two estimators $\hat{\theta}_T(.)$ and $\tilde{\theta}_T(.)$ coincide, and converge to zero under the null hypothesis $H_o = \{\theta_\infty(.) = 0\}$. Diagnostics based on this two steps estimation approach, the first one being parametric, and the second one nonparametric, have to be compared with the solutions usually proposed for this problem. Such solutions consist in estimating directly the nonlinear regression function by :

$$\frac{\sum_{t=1}^{T} K\left[\frac{X_t - x}{h_T}\right] Y_t}{\sum_{t=1}^{T} K\left[\frac{X_t - x}{h_T}\right]} = \hat{\gamma}_T(x)\ [say].$$

Then two approaches have been considered. The first one consists in determining the second order derivative $\frac{\partial^2 \hat{\gamma}_T(x)}{\partial x^2}$ in order to compare it to zero. The second one is based on a comparison between $\hat{\gamma}_T(x)$ and $\hat{a}_T x + \hat{b}_T$.

This last approach is equivalent to the computation of :

$$\hat{\hat{\theta}}_T(x) = Arg\, min_\theta \sum_{t=1}^{T} \frac{1}{h_T} K\left[\frac{X_t - x}{h_T}\right] [Y_t - \hat{a}_T x - \hat{b}_T - \theta]^2,$$

(since $\hat{\hat{\theta}}_T(x) = \hat{\gamma}_T(x) - \hat{a}_T x - \hat{b}_T$), and the comparison to zero of this function. Therefore from corollary 2 we deduce that the approaches based on $\hat{\theta}_T$ and $\hat{\hat{\theta}}_T$ are asymptotically equivalent.

Example 3.3 : Omitted variable in a regression function

Let us consider a linear regression : $Y_t = aX_t + b + u_t$, estimated by O.L.S. The estimators are $\hat{a}_T, \hat{b}_T$ and the residuals $\hat{u}_t = Y_t - \hat{a}_T X_T - \hat{b}_T$. Is it necessary to introduce some additional variable Z in this regression ? This variable may a priori be introduced with an additive effect (i.e. through parameter b) or with a cross effect (i.e. through parameter a). These two possibilities may be jointly analyzed by solving the problem :

$$Min_{a,b} \sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right) [Y_t - \hat{a}_T X_t - \hat{b}_T - aX_t - b]^2.$$

The solutions are :

$$\left[\hat{a}_T(z), \hat{b}_T(z)\right]' = \left\{\sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right) \begin{bmatrix} X_t^2 & X_t \\ X_t & 1 \end{bmatrix}\right\}^{-1}$$

$$\sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right) \begin{bmatrix} X_t \hat{u}_t \\ \hat{u}_t \end{bmatrix}.$$

The null hypothesis of no cross effect is $(a_\infty(z) = 0, \forall z)$, and the null hypothesis of no additive effect is $(b_\infty(z) = 0, \forall z)$.

Example 3.4 : Quadratic ARCH specification

Let us consider an AR(1) model with ARCH errors : $Y_t = aY_{t-1} + b + \varepsilon_t$, where $V(Y_t/\underline{Y_{t-1}}) = k_t = c_o + c_1 \varepsilon_{t-1}^2$.

We can estimate the different parameters by the pseudo maximum likelihood method. These estimators are denoted by $\hat{a}_T, \hat{b}_T, \hat{c}_{oT}, \hat{c}_{1T}$ respectively, and the residuals by $\hat{\varepsilon}_t = Y_t - \hat{a}_T Y_{t-1} - \hat{b}_T$.

It is known that financial series often show a leverage effect, i.e. an asymmetric reaction of the volatility to positive or negative values of the ε_{t-1} [Black (1976), Schwert (1989), Gouriéroux-Monfort (1992), Engle (1993), Bollerslev-Engle-Nelson (1993), Engle-Ng (1993)]. The problem is to check $\{k_t = c_o + c_1\varepsilon_{t-1}^2\}$ against $\{k_t = k(\varepsilon_{t-1})\}$.

For such a diagnostic, we can solve the problems :

$$Min_k \sum_{t=1}^{T} K\left(\frac{\hat{\varepsilon}_{t-1} - \varepsilon}{h_T}\right)\left\{-\frac{1}{2} Log(\hat{c}_{oT} + \hat{c}_{1T}\hat{\varepsilon}_{t-1}^2 + k) - \frac{1}{2}\frac{\hat{\varepsilon}_t^2}{(\hat{c}_{oT} + \hat{c}_{1T}\hat{\varepsilon}_{t-1}^2 + k)}\right\},$$

and compare to zero the solution $\hat{k}_T(\varepsilon)$. It has to be noted that the conditioning variable $\hat{\varepsilon}$ depends on the preliminary estimation step, and that we do not obtain an analytical expression of $\hat{k}_T(\varepsilon)$. Therefore it may be preferable to use the modified estimator. It is given by :

$$\tilde{k}_T(\varepsilon) = \left\{\sum_{t=1}^{T} K\left(\frac{\hat{\varepsilon}_{t-1} - \varepsilon}{h_T}\right)\left[\frac{1}{2}\frac{1}{(\hat{c}_{oT} + \hat{c}_{1T}\hat{\varepsilon}_{t-1}^2)^2} - \frac{\hat{\varepsilon}_t^2}{(\hat{c}_{oT} + \hat{c}_{1T}\hat{\varepsilon}_{t-1}^2)^3}\right]\right\}^{-1}$$

$$\left\{\sum_{t=1}^{T} K\left(\frac{\hat{\varepsilon}_{t-1} - \varepsilon}{h_T}\right)\left[-\frac{1}{2}\frac{1}{(\hat{c}_{oT} + \hat{c}_{1T}\hat{\varepsilon}_{t-1}^2)} + \frac{\hat{\varepsilon}_t^2}{2(\hat{c}_{oT} + \hat{c}_{1T}\hat{\varepsilon}_{t-1}^2)^2}\right]\right\}.$$

Finally a similar approch may be applied in the case of an ARCH-M model, in which the conditional variance k_t has an effect on the conditional mean. In this case the k parameter also appears in the expression of $\hat{\varepsilon}_t^2$.

IV - AN ILLUSTRATION

The implementation of the previous approach may be illustrated by a regression model with cross effects.

The data generating processes (DGP)

The data are generated in the following way :

$$(4.1) \qquad Y_t = a_o(Z_t)X_t + b_o + u_t,$$

X_t, Z_t, u_t are independent standard white noise processes, $a_o(.)$ is a function and b_o a scalar.

In the Monte Carlo study we used two data generating processes:
DGP I : $a_o(Z_t) = Z_t, b_o = 1$,
DGP II : $a_o(Z_t) = Z_t^2, b_o = 1$

Linear regression

Suppose now that we estimate a linear model of the form :

$$(4.2) \qquad Y_t = aX_t + b + v_t.$$

A straightforward computation shows that the O.L.S. estimators $\hat{a}_T, \hat{b}_T$ converge respectively to the pseudo-true values $Ea_o(Z_t)$ and b_o ; in the Monte Carlo studies these values are :
DGP I : $EZ_t = 0, b_o = 1$,
DGP II :$E(Z_t^2) = 1, b_o = 1$.

The residuals $\hat{u}_t = Y_t - \hat{a}_T X_t - \hat{b}_T$, converge to $u_t^o = [a_o(Z_t) - Ea_o(Z_t)]X_t + u_t$; this implies that $Eu_t^o Z_t = E[Z_t a_o(Z_t) - EZ_t Ea_o(Z_t)]EX_t = 0$ since $EX_t = 0$ and, therefore, a usual test on a possible omission of Z_t as an exogenous variable is likely to reject this omission. Moreover since u_t^o is a white noise and since $E[u_t^o/Z_t]$ is equal to zero, a residual plot of $\hat{u}_t$ against t or against Z_t will probably not detect any anomaly.

These predictions are confirmed by finite sample experiments ; for instance we give below (see figure 4.3) the residual plots against t and Z_t, for DGP I and II, and for $T = 50$ and $T = 200$.

Note that, if $EX_t = c \neq 0$, we have, in the case of DGP II for instance, $Eu_t^o Z_t = 0$ and $E(u_t^o/Z_t) = c[Z_t^2 - 1]$; therefore a residuals plot of $\hat{u}_t$ against Z_t could lead to the introduction of Z_t^2 as explanatory variable but, in this new model, the O.L.S. residuals would not detect any anomaly even though the model is still misspecified.

Kernel M-estimation of the regression coefficient

Kernel M-estimators of the discrepancy parameters θ_1, θ_2 are given by :

$$Min_{\theta_1,\theta_2} \sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right) [Y_t - \hat{a}_T X_t - \hat{b}_T - \theta_1 X_t - \theta_2]^2,$$

i.e.

$$\left[\hat{\theta}_{1T}(z), \hat{\theta}_{2T}(z)]'\right]' = \left\{\sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right) \begin{pmatrix} X_t^2 & X_t \\ X_t & 1 \end{pmatrix}\right\}^{-1}$$

$$\sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right) \begin{pmatrix} X_t \hat{u}_t \\ \hat{u}_t \end{pmatrix},$$

whose limit is, under (4.1) :

$$(4.4) \qquad [\theta_{1\infty}(z), \theta_{2\infty}(z)]' = [a_o(z) - Ea_o(Z_t), 0]'.$$

The values of the estimators $\hat{\theta}_{1T}(z), \hat{\theta}_{2T}(z)$ depend on the kernel K, on the bandwidth h_T and on the number of observations T.

We have considered three kernels : the Epanechnikov kernel $[K(u) = 3/4(1-u^2)\mathbb{I}_{\{|u|<1\}}]$, the Quartic kernel $[K(u) = 15/16(1-u^2)^2\mathbb{I}_{\{|u|<1\}}]$ and the Gaussian kernel $[K(u) = (2\pi)^{-\frac{1}{2}}exp(-u^2/2)]$; these kernels provide fairly simular results and in the sequel we retain the Epanechnikov kernel.

Moreover, we are interested in the null hypothesis $H_o : \{\theta_\infty(.) = 0\}$ and, under H_o, the asymptotic covariance matrix of $\sqrt{Th_T}\hat{\theta}_T(z)$ is:

$$\Sigma(z) = \frac{J^{-1}(z)C_o(z)J^{-1}(z)}{f_z(z)} \int K^2(u)du \text{ (see property 1).}$$

It is easily seen that J(z) and $C_o(z)$ are consistently estimated by $2\frac{X'X}{T}$ and $4\,\hat{\sigma}_T^2 \left(\frac{X'X}{T}\right)$ respectively, where X is the matrix whose rows are $(X_t, 1)$ and $\hat{\sigma}_T^2$ is the usual O.L.S. estimator of the error variance. Moreover, $\int K^2(u)du = 3/5$ for the Epanechnikov kernel, and $f_Z(z)$, the p.d.f. of Z, can be estimated by using the same kernel and the same bandwidth as in the estimation of $\hat{\theta}_T(z)$. Finally, under H_o, the distribution of $\hat{\theta}_T(z)$ is approximated by :

$$(4.5) \qquad N\left[0, \frac{3\,\hat{\sigma}_T^2}{5\,\hat{f}_Z(z)h_T}(X'X)^{-1}\right].$$

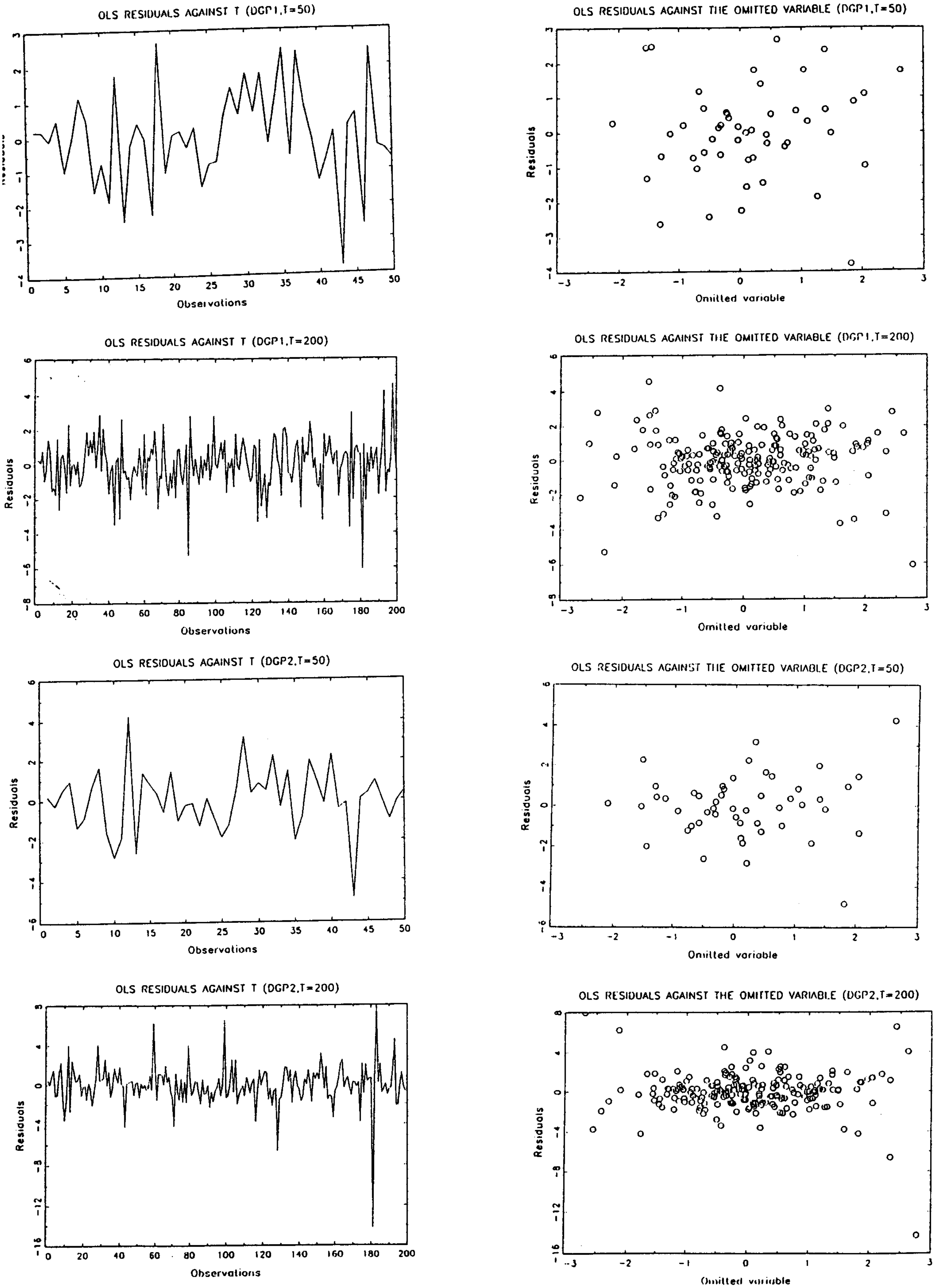

Figure 4.3 : Residuals plots

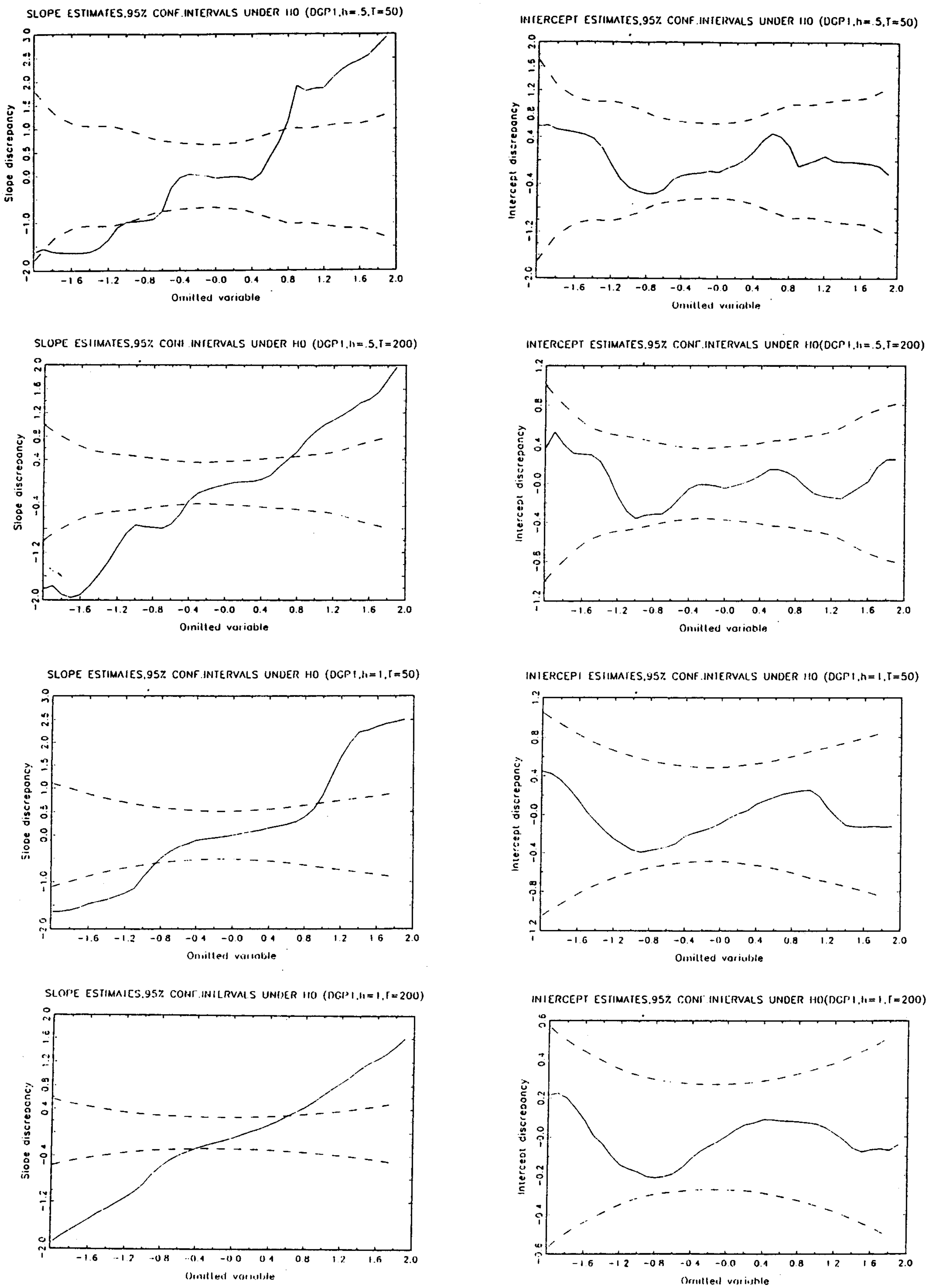

Figure 4.6 : Slope and intercept discrepancies (DGP I)

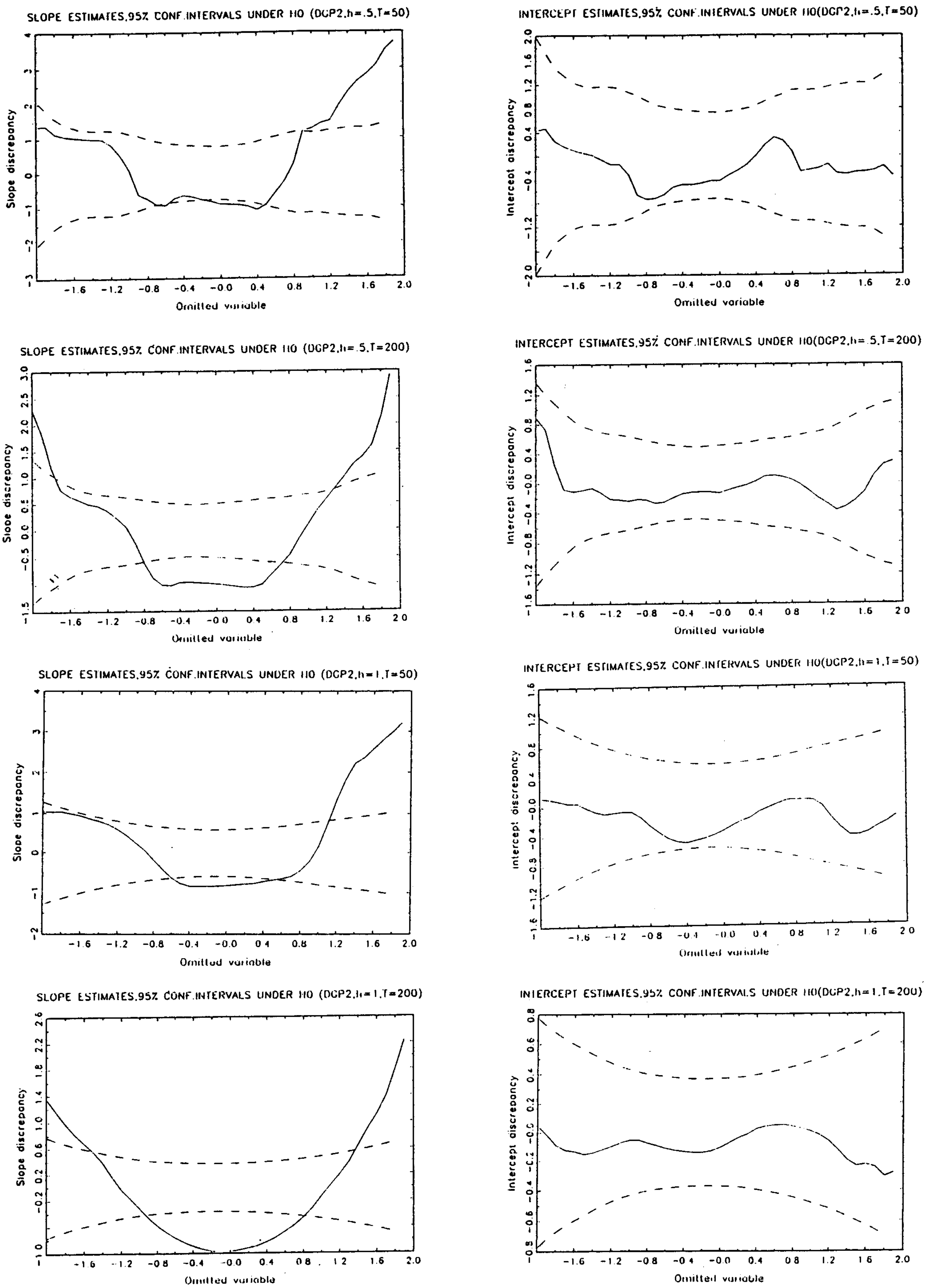

Figure 4.7 : Slope and intercept discrepancies (DGP II)

This provides, for any z, asymptotic confidence intervals for θ_1 (the slope discrepancy) and θ_2 (the intercept discrepancy) under H_o.

Figures (4.6) and (4.7) provide the graphs of the slope and intercept discrepancies $\hat{\theta}_1(z), \hat{\theta}_{2T}(z)$, as well as their 95% asymptotic confidence intervals under H_o, for DGP I and DGP II respectively; for each DGP we consider two values of h_T : 0.5 and 1, and two values of T : 50 and 200.

From these figures it is seen that all the intercept discrepancies are inside the confidence intervals, whereas some slope discrepancies are outside the intervals for all the values of h_T and T. It is also seen that, as expected, the curves are much smoother for $h_T = 1$ than for $h_T = .5$; the smoothness is also increasing with T.

V - FURTHER DEVELOPMENTS

In the previous sections, we have explained how to estimate a functional parameter, and how to check for its constancy. From a theoretical point of view, the results may be extended in several directions : we might look for some consistency results uniform in the values of the conditioning variable, and for the corresponding functional limit theorems, allowing for the determination of functional confidence intervals. We might also study the optimal choice of the bandwidth [see e.g. Rice (1984), Hardle-Marion (1985)] and of the kernel. However in our modelling strategy these choices are likely to depend on the alternative which is considered, and, since this alternative has an unspecified functional form, we may be led to introduce some minimax aspects.

Finally some other developments may concern the applications. In particular in financial problems, the basic models are generally misspecified. Then it is necessary to correct them for such a misspecification. The idea of this paper consists in directly introducing the correction through the parameters which are no longer assumed to be constant ; it may be applied to various kinds of parameters : (implicit) volatilities, (implicit) hedging parameters, parameters defining the term structure...

Appendix 1
Consistency of the kernel M-estimator

We shall prove the uniform a.s. convergence of the objective function, which implies the asymptotic existence and the strong convergence of the associated M-estimator (see Jennrich (1969), Robinson (1989), (1990)). The uniform aspect of the convergence is specific of this appendix [see appendix 2 lemma 1 for a non uniform consistency result]. Note that we mean uniform in the parameter and not with respect to s, as usual in nonparametric estimation. For this purpose we first decompose this objective function in three parts :

$$L_T(\theta) = [L_T(\theta) - \tilde{L}(\theta)] + [\tilde{L}_T(\theta) - E\,\tilde{L}_T(\theta)] + E\,\tilde{L}_T(\theta),$$

where : $\tilde{L}_T(\theta) = \sum_{t=1}^{T} \frac{1}{Th_T^d} K\left[\frac{S(Y_t,\alpha_\infty) - s}{h_T}\right] \psi(Y_t,\alpha_\infty;s;\theta).$

Then we prove that :

i) E $\tilde{L}_T(\theta)$ converge a.s uniformly to $L_\infty(\theta) = E[\psi(Y,\alpha_\infty;s;\theta)/S(Y,\alpha_\infty) = s]f_S(s)$.

ii) $\tilde{L}_T(\theta) - E\,\tilde{L}_T(\theta)$ converges a.s to zero.

iii) $\tilde{L}_T(\theta) - E\,\tilde{L}_T(\theta)$ satisfies some equicontinuity property, which implies the uniform a.s. convergence of $\tilde{L}_T(\theta) - E\,\tilde{L}_T(\theta)$ to zero.

iv) $sup_{\theta\epsilon\Theta} \mid L_T(\theta) - \tilde{L}_T(\theta) \mid$ converges a.s to zero.

For notational convenience and without loss of generality the proof is given for d=1. We denote $S_t = S(Y_t;\alpha_\infty)$, $\psi_t(\theta) = \psi(Y_t;\alpha_\infty;s,\theta)$. If ψ is a mapping of R^p on R^q, $x \to (\psi_1(x),...,\psi_q(x))$: $\| \psi(x) \| = Max\{|\psi_1(x)|,...,|\psi_q(x)|\}$.

A - The set of assumptions

Conditions on the observable process

i) $(Y_t, t\epsilon\mathbb{Z})$ is a strictly stationary process.

ii) It satisfies the geometric mixing condition :

$$sup \mid P(A\cap B) - P(A)P(B) \mid < a\rho^K,\ K\epsilon N^*, A\epsilon F_{-\infty}^{0}, B\epsilon F_K^{+\infty},$$

with $a \geq 0$, $1 > \rho \geq 0$, where : $F_{-\infty}^{0} = \sigma(Y_s, s \leq 0), F_K^{+\infty} = \sigma(Y_s, s \geq K)$.

Conditions on S

i) S_i has a marginal continuous distribution with a strictly positive p.d.f. f_S.

ii) $S(Y_t, \alpha)$ is continuously differentiable with respect to α.

Conditions on the kernel

i) K is a differentiable function, with bounded support.

ii) It satisfies $\int K(u)du = 1$.

iii) $\frac{dK(u)}{du}$ is integrable.

Conditions on the bandwidth

i) $h_T \rightarrow 0$, when $T \rightarrow \infty$.

ii) There exists $\beta \in]0,1[$ such that $\frac{T^{\frac{(1-\beta)}{2}} h_T}{Log\, T} \rightarrow +\infty$, when $T \rightarrow \infty$.

Conditions on the estimator $\hat{\alpha}_T$

i) $\hat{\alpha}_T$ converges a.s to a deterministic value α_∞.

ii) $\hat{\alpha}_T - \alpha_\infty = o(h_T^{3/2})$.

Condition on the parameter space

Θ is a compact set

Conditions on the objective function ψ

i) The mapping $\theta \rightarrow \psi(Y, \alpha; s; \theta)$ is continuous, $\forall Y, \alpha, s$.

ii) The mapping $Y \rightarrow \psi(Y, \alpha; s; \theta)$ is measurable, $\forall \theta$,.

iii) The mapping $\alpha \rightarrow \psi(Y, \alpha; s; \theta)$ is differentiable at α_∞, $\forall Y, \theta, s$.

iv) The mapping $\theta \rightarrow \psi(Y, \alpha_\infty; s; \theta)$ is continuously differentiable, $\forall Y, s$.

Asymptotic identifiability condition

There exists a unique solution $\theta_\infty(s)$ to the optimization problem:

$$Min_{\theta \in \Theta} L_\infty(\theta; s),$$

where:

$$L_\infty(\theta; s) = E[\psi_t(\theta)/S_t = s] f_S(s).$$

More technical conditions

(A.1) For any $\varepsilon > o$, there exists a neighbourhood W(s) of s such that:

$$\forall z \epsilon W(s) : \sup_{\theta \in \Theta} \mid E[\psi(Y, \alpha_\infty; z; \theta)/S(Y, \alpha_\infty) = z] f_S(z)$$
$$- E[\psi(Y, \alpha_\infty; s; \theta)/S(Y, \alpha_\infty) = s] f_S(s) \mid < \varepsilon.$$

(A.2) $\exists \delta > 0,\ E \left| \sup_{\theta \in \Theta} |\psi(Y, \alpha_\infty; z; \theta)| \right|^{\frac{2}{\beta} + \delta} < +\infty,$
where β is the parameter appearing in condition ii) concerning the bandwith.

(A.3) $E \left[\sup_{\theta \in \Theta} || \frac{\partial \psi}{\partial \theta}(Y, \alpha_\infty; z; \theta) || \right] < \infty.$

(A.4) i) We have in a neighbourhood of α_∞,

$$\left| K\left(\frac{S(Y,\alpha) - s}{h_T}\right) \psi(Y, \alpha; s; \theta) - K\left(\frac{S(Y,\alpha_\infty) - s}{h_T}\right) \psi(Y, \alpha_\infty; s; \theta) \right.$$
$$- \frac{\partial K}{du}\left(\frac{S(Y,\alpha_\infty) - s}{h_T}\right) \psi(Y, \alpha_\infty; s; \theta) \frac{\partial S}{\partial \alpha'}(Y, \alpha_\infty)(\alpha - \alpha_\infty)$$
$$\left. - K\left(\frac{S(Y,\alpha_\infty) - s}{h_T}\right) \frac{\partial \psi}{\partial \alpha'}(Y, \alpha_\infty; s; \theta)(\alpha - \alpha_\infty) \right|$$
$$\leq \frac{\| \alpha - \alpha_\infty \|^2}{h_T^2} M(Y, s; \theta), \forall Y, s, \theta.$$

ii) $E\left[\sup_{\theta\in\Theta} M(Y_o, s;\theta)\right] < +\infty.$

iii) The following functions are bounded in a neihbourhood of s:

$$E\left[\sup_{\theta\in\Theta} |\psi(Y_o, \alpha_\infty; s;\theta)| \| \frac{\partial S}{\partial \alpha'}(Y_o, \alpha_\infty)\| / S_o = .\right] f_S(.),$$

$$E\left[\sup_{\theta\in\Theta} \| \frac{\partial S}{\partial \alpha'}(Y_o, \alpha_\infty; s;\theta)\| / S_o = .\right] f_S(.).$$

B - Uniform consistency of $E\,\tilde{L}_T(\theta)$

Lemma 1 : $\lim_{T\to\infty} \sup_{\theta\epsilon\Theta} \mid E\tilde{L}_T(\theta) - L_\infty(\theta) \mid = 0.$

Proof of lemma 1 :

Since $h_T \to 0$, K sums to one and has a bounded support, we know that for any neighbourhood W(s) of s, there exists T_o such that :

$$\forall T \geq T_o, \sup_{\theta\in\Theta} \mid E\tilde{L}_T(\theta) - L_\infty(\theta) \mid$$

$$\leq \sup_{\theta\in\Theta} \int \mid E[\psi(Y), \alpha_\infty; s - yh_T;\theta)/S(Y,\alpha_\infty) = s - yh_T] f_S(s - yh_T)$$

$$-E[\psi(Y, \alpha_\infty; s;\theta)/S(Y,\alpha_\infty) = s] f_S(s) \mid\mid K(y) \mid dy$$

$$\leq \sup_{z\in W(s)} sup_{\theta\in\Theta} \mid E[\psi(Y,\alpha_\infty; z;\theta)/S(Y,\alpha_\infty) = z] f_S(z)$$

$$-E\psi[(Y,\alpha_\infty; s;\theta)/S(Y,\alpha_\infty) = s] f_S(s) \mid \int \mid K(u) \mid du.$$

Therefore lemma 1 is a direct consequence of assumption A.1.

□

C - $\tilde{L}_T(\theta) - E\tilde{L}_T(\theta)$ converges a.s to zero

We first consider the following decomposition of $\tilde{L}_T$:

$$\begin{aligned} \tilde{L}_T(\theta) &= \tilde{L}_T^+(\theta) + \tilde{L}_T^-(\theta) \\ &= \frac{1}{T}\sum_{t=1}^{T} \frac{1}{h_T} K\left(\frac{S_t - s}{h_T}\right) \psi_t \mathbb{1}_{|\psi_t|<M_T} \end{aligned}$$

$$+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_T}K\left(\frac{S_t-s}{h_T}\right)\psi_t\mathbb{1}_{|\psi_t|\geq M_T},$$

where $M_T = T^\gamma$, with $\gamma(\frac{2}{\beta}+\delta)>1$ and $\beta\epsilon]0,1[$ is the value given in condition ii) on the bandwidth and δ is the value appearing in assumption A.2.

Lemma 2 : $|\tilde{L}_T^-(\theta)-E\tilde{L}_T^-(\theta)|\to 0$ a.s, when $T\to\infty$.

Proof of lemma 2 :

Following the idea of the proof of proposition 1 given in Mack-Silverman (1982), we conclude that:

$$|\tilde{L}_T^-(\theta)-E\tilde{L}_T^-(\theta)|=0 \text{ a.s. when } T\to\infty,$$

from the choice of γ and from assumption A.2.

□

Lemma 3 : $|\tilde{L}_T^+(\theta)-E\tilde{L}_T^+(\theta)|\to 0$ a.s, when $T\to\infty$.

Proof of lemma 3 :

We use the following inequality (see Bosq (1988)):

If (Z_t) is a real geometrically mixing process, such that $EZ_t=0$, $\sup_{1\leq t\leq T}|Z_t|\leq N_T$, then :

$\forall\beta'\epsilon]0,1[,\exists T_o(\beta'):\forall T\geq T_o(\beta'),\forall\varepsilon>0:$

$$P\left[\frac{1}{T}\left|\sum_{t=1}^{T}Z_t\right|>\varepsilon\right]\leq 3\,exp\left(-\frac{\varepsilon}{14}\frac{T^{(1-\beta')/2}}{N_T}\right).$$

We have :

$$\tilde{L}_T^+(\theta)-E\tilde{L}_T^+(\theta)=\frac{1}{T}\sum_{t=1}^{T}Z_t,$$

where

$$\begin{aligned}|Z_t| &= \left|\frac{1}{h_T}K\left(\frac{S_t-s}{h_T}\right)\psi_t\mathbb{1}_{|\psi_t|<M_T}-E\left[\frac{1}{h_T}K\left(\frac{S_t-s}{h_T}\right)\psi_t\mathbb{1}_{|\psi_t|<M_T}\right]\right|\\ &\leq \frac{2}{h_T}M_T\,sup_u\,|K(u)|\,.\end{aligned}$$

Therefore using the inequality, we get :

$$P[| \tilde{L}_T^+(\theta) - E\tilde{L}_T^+(\theta) |> \varepsilon] \leq 3\, exp\left(-\frac{\varepsilon}{14}\, \frac{h_T T^{(1-\beta')/2}}{2M_T sup_u |K(u)|}\right).$$

As soon as β', γ are chosen such that : $\frac{1-\beta'}{2} - \gamma \geq \frac{1-\beta}{2}$, we deduce from the condition $lim_{T\to\infty} \frac{T^{\frac{1-\beta}{2}} h_T}{Log\, T} = +\infty$, that:

$$\sum_{t=1}^{T} P[| \tilde{L}_T^+(\theta) - E\tilde{L}_T^+(\theta) |> \varepsilon] < +\infty,$$

which implies the a.s. convergence.

□

D - Equicontinuity of $\tilde{L}_T(\theta) - E\tilde{L}_T(\theta)$

Lemma 4 : $\forall\, \theta_1, \theta_2 \epsilon \Theta$:

$$B_T(\theta_1, \theta_2) = | \tilde{L}_T(\theta_1) - E\tilde{L}_T(\theta_1) - \tilde{L}_T(\theta_2) + E\tilde{L}_T(\theta_2) | = 0\left(\frac{\| \theta_1 - \theta_2 \|}{h_T}\right) \quad a.s.$$

Proof of lemma 4 :

We have :

$$B_T(\theta_1, \theta_2) \leq | \tilde{L}_T(\theta_1) - \tilde{L}_T(\theta_2) | + | E\tilde{L}_T(\theta_1) + E\tilde{L}_T(\theta_2) |,$$

where:

$$\begin{aligned} | \tilde{L}_T(\theta_1) - \tilde{L}_T(\theta_2) | &\leq \frac{1}{T} \sum_{t=1}^{T} \frac{1}{h_T} | \psi_t(\theta_1) - \psi_t(\theta_2) | \left| K\left(\frac{S_t - s}{h_T}\right)\right| \\ &\leq \frac{dim\,\theta}{T} \sum_{t=1}^{T} \frac{1}{h_T} \sup_{\theta\epsilon\Theta} \| \frac{\partial \psi_t}{\partial \theta} \| \sup_u |K(u)| \, \| \theta_1 - \theta_2 \| . \end{aligned}$$

Because of assumption A.3, we deduce that :

$$B_T(\theta_1, \theta_2) = 0\left(\frac{\| \theta_1 - \theta_2 \|}{h_T}\right).$$

□

We may now introduce a finite covering of the compact set Θ by balls Θ_ℓ, $\ell = 1, ..., N$ of size smaller than $\frac{1}{T}$: $\forall \theta_1, \theta_2 \epsilon \Theta_\ell$: $\| \theta_1 - \theta_2 \| < \frac{1}{T}$. This covering may be chosen such that : $N < \{[T\, diam\, \Theta] + 1\}^{dim \theta}$.

We have :

$$
\begin{aligned}
&\sup_{\theta \epsilon \Theta} | \tilde{L}_T(\theta) - E\tilde{L}_T(\theta) | \\
&\quad = \quad Max_{\ell=1,\dots,N}\, sup_{\theta \epsilon \Theta_\ell} \, | \tilde{L}_T(\theta) - E\tilde{L}_T(\theta) | \\
&\quad \leq \quad Max_{\ell=1,\dots,N}\, sup_{\theta \epsilon \Theta_\ell} \, | \tilde{L}_T(\theta) - E\tilde{L}_T(\theta) - \tilde{L}_T(\theta_\ell) - E\tilde{L}_T(\theta_\ell) | \\
&\quad \qquad + Max_{\ell=1,\dots,N}\, sup_{\theta \epsilon \Theta_\ell} \, | \tilde{L}_T(\theta_\ell) - E\tilde{L}_T(\theta_\ell) |,
\end{aligned}
$$

where $\theta_\ell \epsilon \Theta_\ell$, $\ell = 1, ..., N$.

Therefore from lemma 4 :

$$
\begin{aligned}
&sup_{\theta \in \Theta} \, | \tilde{L}_T(\theta) - E\tilde{L}_T(\theta) | \\
&\leq \quad Max_{\ell=1,\dots,N}\, sup_{\theta \epsilon \Theta_\ell} \, 0 \left(\frac{\| \theta - \theta_\ell \|}{h_T} \right) + Max_{\ell=1,\dots,N} \, | \tilde{L}_T(\theta_\ell) - E\tilde{L}_T(\theta_\ell) | \\
&\leq \quad 0 \left(\frac{1}{Th_T} \right) + Max_{\ell=1,\dots,N} \, | \tilde{L}_T(\theta_\ell) - E\tilde{L}_T(\theta_\ell) | \, .
\end{aligned}
$$

Since $Th_T \rightarrow +\infty$, the uniform and pointwise convergence to zero of $\tilde{L}_T(.) - E\tilde{L}_T(.)$ are equivalent.

E - Uniform almost sure convergence of $L_T(\theta) - \tilde{L}_T(\theta)$ to zero

Lemma 5 : $\lim\limits_{T \rightarrow \infty} \sup\limits_{\theta \epsilon \Theta} | L_T(\theta) - \tilde{L}_T(\theta) | = 0$ a.s.

Proof of lemma 5 :

We have :

$$
\begin{aligned}
&\sup_{\theta \in \Theta} | L_T(\theta) - \tilde{L}_T(\theta) | \\
&= \sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} \frac{1}{h_T} K \left(\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T} \right) \psi(Y_t, \hat{\alpha}_T; s; \theta) \right. \\
&\qquad\qquad \left. - K \left(\frac{S(Y_t, \alpha_\infty) - s}{h_T} \right) \psi(Y_t, \alpha_\infty; s; \theta) \right|
\end{aligned}
$$

$$\leq \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_T}\sup_{\theta\in\Theta}\left|K\left(\frac{S(Y_t,\hat{\alpha}_T)-s}{h_T}\right)\psi(Y_t,\hat{\alpha}_T;s;\theta)\right.$$

$$\left. -K\left(\frac{S(Y_t,\alpha_\infty)-s}{h_T}\right)\psi(Y_t,\alpha_\infty;s;\theta)\right|$$

$$\leq \frac{1}{Th_T}\sum_{t=1}^{T}\left|\frac{dK}{du}\left(\frac{S_t-s}{h_T}\right)\right|\sup_{\theta\in\Theta}\mid\psi(Y_t,\alpha_\infty;s;\theta)\mid\parallel\frac{\partial S}{\partial\alpha'}(Y_t,\alpha_\infty)\parallel$$

$$\times\frac{\parallel\hat{\alpha}_T-\alpha_\infty\parallel}{h_T}$$

$$+\frac{1}{Th_T}\sum_{t=1}^{T}\left|K\left(\frac{S_t-s}{h_T}\right)\right|\sup_{\theta\in\Theta}\parallel\frac{\partial\psi}{\partial\alpha'}(Y_t,\alpha_\infty;s;\theta)\parallel\parallel\hat{\alpha}_T-\alpha_\infty\parallel$$

$$+\frac{1}{Th_T^3}\sum_{t=1}^{T} sup_\Theta\, M(Y_t,s;\theta)\parallel\hat{\alpha}_T-\alpha_\infty\parallel^2,$$

using assumption A.4 i). From assumption A.4 ii) and iii), the different terms of the R.H.S. are of respective orders:

$$O\left(\frac{\parallel\hat{\alpha}_T-\alpha_\infty\parallel}{h_T}\right), O(\parallel\hat{\alpha}_T-\alpha_\infty\parallel), O\left(\frac{\parallel\hat{\alpha}_T-\alpha_\infty\parallel^2}{h_T^3}\right).$$

Finally the result follows from condition ii) concerning the speed of convergence of $\hat{\alpha}_T$.

Appendix 2
Asymptotic Normality of the kernel M-Estimator

The proof of the asymptotic normality of the kernel M-estimator follows the usual steps. The main point of this proof deals with the fact that the observations are not assumed to be i.i.d., but may have some correlations. To take into account this feature, we use some theorems proved in Tenreiro (1993), which are presented in part A of this appendix. The proof is based on the expansion of the first order condition (2.2), given in this appendix for d=1:

$$(2.2) \quad \sum_{t=1}^{T} \frac{1}{\sqrt{T}\sqrt{h_T}} K\left(\frac{S(Y_t, \hat{\alpha}_T) - s}{h_T}\right) \frac{\partial \psi}{\partial \theta_i}(Y_t, \hat{\alpha}_T; s; \hat{\theta}_T(s)) = 0, \forall i.$$

We first expand the first order condition around α_∞. Using an assumption similar to (A.4) with ψ replaced by $\frac{\partial \psi}{\partial \theta_i}$, (2.2) is equivalent to:

$$\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial \psi}{\partial \theta_i}(Y_t, \alpha_\infty; s; \hat{\theta}_T(s))$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} \frac{dK}{du}\left(\frac{S_t - s}{h_T}\right) \frac{\partial \psi}{\partial \theta_i}(Y_t, \alpha_\infty; s; \hat{\theta}_T(s)) \frac{\partial S}{\partial \alpha'}(Y_t; \alpha_\infty) \frac{\hat{\alpha}_T - \alpha_\infty}{h_T}$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial^2 \psi}{\partial \theta_i \partial \alpha'}(Y_t, \alpha_\infty; s; \hat{\theta}_T(s))(\hat{\alpha}_T - \alpha_\infty)$$

$$+0_P\left(\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} \frac{\| \hat{\alpha}_T - \alpha_\infty \|^2}{h_T^2} N(Y_t; s; \hat{\theta}_T(s))\right) = 0,$$

where N is for the bounding function, and where the last term is of order: $0_p\left(\sqrt{Th_T} \| \hat{\alpha}_T - \alpha_\infty \|\right)$, if we assume: $\hat{\alpha}_T - \alpha_\infty = O(h_T^3)$.

A second expansion around $\theta_\infty(s)$ gives:

$$\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial \psi}{\partial \theta_i}(Y_t, \alpha_\infty; s; \theta_\infty(s))$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial^2 \psi}{\partial \theta_i \partial \theta'}(Y_t, \alpha_\infty; s; \theta_\infty(s))(\hat{\theta}_T(s) - \theta_\infty(s))$$

$$+\sum_{t=1}^{T} \frac{1}{2\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) (\hat{\theta}_T(s) - \theta_\infty(s))' \frac{\partial^3 \psi}{\partial\theta_i \partial\theta\partial\theta'}(Y_t, \alpha_\infty; s; \tilde{\theta}_T)$$

$$\times(\hat{\theta}_T(s) - \theta_\infty(s))$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} \frac{dK}{du}\left(\frac{S_t - s}{h_T}\right) \frac{\partial\psi}{\partial\theta_i}(Y_t, \alpha_\infty; s; \theta_\infty(s)) \frac{\partial S}{\partial\alpha'}(Y_t; \alpha_\infty) \frac{\hat{\alpha}_T - \alpha_\infty}{h_T}$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} \frac{dK}{du}\left(\frac{S_t - s}{h_T}\right) \frac{\partial^2\psi}{\partial\theta_i \partial\theta'}(Y_t, \alpha_\infty; s; \tilde{\tilde{\theta}}_T)(\hat{\theta}_T(s) - \theta_\infty(s))$$

$$\times \frac{\partial S}{\partial\alpha'}(Y_t; \alpha_\infty) \frac{\hat{\alpha}_T - \alpha_\infty}{h_T}$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial^2\psi}{\partial\theta_i \partial\alpha'}(Y_t, \alpha_\infty; s; \hat{\theta}_T(s))(\hat{\alpha}_T - \alpha_\infty)$$

$$+0_p\left(\sqrt{Th_T} \parallel \hat{\alpha}_T - \alpha_\infty \parallel\right) = 0.$$

where $\tilde{\theta}_T$ and $\tilde{\tilde{\theta}}_T$ belong to the segment $]\hat{\theta}_T(s), \theta_\infty(s)[$.

As usual the idea of the proof consists in checking that the different terms of the expansion are negligible except the two first ones. To prove that they are negligible, different arguments are used for the fourth term of the expansion [a central limit theorem] and for the other ones [a law of large number].

A - <u>Two lemmas</u> [see Tenreiro (1993)]

These results are valid for processes (S_t, Y_t) satisfying the strict stationarity and the mixing condition described in Appendix 1, and for a continuous kernel with bounded support.

Consistency in probability : lemma 1

We have : $\frac{1}{Th_T} \sum_{t=1}^{T} g(Y_t) K\left(\frac{S_t - s}{h_T}\right) \rightarrow E[g(Y)/S = s)] f_S(s)$, in probability under the following assumptions :

i) $E \parallel g(Y_t) \parallel^{2+\delta} < +\infty$, for a positive δ;

ii) $E[g(Y_t)/S_t = .] f_S(.)$ is continuous in s;

iii) $E[\| g(Y_t) \|^{2+\delta} / S_t = .] f_S(.)$ is bounded in a neighbourhood of s;

iv) $E[| g_i(Y_t) \| g_i(Y_o) | / S_t = ., S_o = .] f_{S_t, S_o}(.,.)$ are uniformly bounded in a neighbourhood of (s,s);

v) $h_T \to 0$, $Th_T \to +\infty$, if $T \to +\infty$.

<u>**Asymptotic normality**</u> **: lemma 2**

We have :

$$\sqrt{Th_T}\left\{\frac{1}{Th_T}\sum_{t=1}^{T} g(Y_t)K\left(\frac{S_t - s}{h_T}\right) - E\left[\frac{1}{Th_T}\sum_{t=1}^{T} g(Y_t)K\left(\frac{S_t - s}{h_T}\right)\right]\right\}$$

$$\xrightarrow{d} N(0, E[g(Y_t)g'(Y_t)/S_t = s] f_S(s) \int K^2(u)du),$$

under the following assumptions :

i) $E \| g(Y_t) \|^{3+\delta} < +\infty$, for a positive or nul δ;

ii) $E[g(Y_t)g'(Y_t)/S_t = .] f_S(.)$ is continuous in s;

iii) $E[\| g(Y_t) \|^{3+\delta} / S_t = .] f_S(.)$ is bounded in a neighbourhood of s.

iv) The following functions are uniformly bounded in a neighbourhood of (s,s) or of (s,s,s) respectively :

$E[| g_i(Y_t) g_j(Y_o) | / S_t = ., S_o = .] f_{S_t, S_o}(.,.)$,

$E[| g_i(Y_t) |^2 | g_j(Y_o) | / S_t = ., S_o = .] f_{S_t, S_o}(.,.)$,

$E[| g_i(Y_t) \| g_j(Y_o) |^2 / S_t = ., S_o = .] f_{S_t, S_o}(.,.)$,

$E[| g_i(Y_\tau) \| g_j(Y_t) \| g_k(Y_o) | / S_t = ., S_\tau = ., S_o = .] f_{S_\tau, S_t S_o}(.,.,.)$,

v) $h_T \to o, Th_T \to +\infty$, if $T \to +\infty$.

In the following sections we denote by P and $\mathcal{N}$ the set of conditions on function g for the convergence in probability (lemma 1) and for the asymptotic normality (lemma 2) respectively.

Finally we have to consider the term of bias :

$$\sqrt{Th_T}\, E\left[\frac{1}{Th_T}\sum_{t=1}^{T} g(Y_t)K\left(\frac{S_t-s}{h_T}\right)\right]$$

$$= \frac{\sqrt{T}}{\sqrt{h_T}}\, E\left[g(Y_t)K\left(\frac{S_t-s}{h_T}\right)\right].$$

If the kernel is of order one, i.e. such that :

$$\int u\, K(u)du = o,$$

and if : $\tilde{g}(u) = E[g(Y_t)/S_t = u]$ takes the value zero for u=s, and is twice continuously differentiable at u=s, we get :

$$= \frac{\sqrt{T}}{\sqrt{h_T}}\, E\left[g(Y_t)K\left(\frac{S_t-s}{h_T}\right)\right].$$

$$= \sqrt{Th_T}\, E\left[E(g(Y_t)/S_t)\frac{1}{h_T}\, K\left(\frac{S_t-s}{h_T}\right)\right]$$

$$= \sqrt{Th_T}\, \int \tilde{g}(u)\frac{1}{h_T}K\left(\frac{u-s}{h_T}\right) f_S(u)du,$$

$$= \sqrt{Th_T}\, \int \tilde{g}(s+h_T v)K(v)f_S(s+h_T v)dv.$$

$$\sim \sqrt{Th_T}\, \int \left(\frac{\partial \tilde{g}}{\partial u}(s)h_T v + \frac{1}{2}\,\frac{\partial^2 \tilde{g}}{\partial u^2}(s)h_T^2 v^2\right) K(v)f_S(s+h_T v)dv$$

$$\sim 0[\sqrt{Th_T^5}]$$

Therefore under the previous conditions and if moreover $Th_T^5 = o(1)$, we also get the asymptotic normality :

$$\frac{1}{\sqrt{Th_T}}\sum_{t=1}^{T} g(Y_t)K\left(\frac{S_t-s}{h_T}\right)$$

$$\xrightarrow{d} N(0, E[g(Y_t)g'(Y_t)/S_t = s]f_S(s)\, \int K^2(u)du).$$

B - The set of assumptions

This set contains all the assumptions of appendix 1, plus some other ones :

additional conditions on the speed of convergence of $\hat{\alpha}_T$

$$\hat{\alpha}_T - \alpha_\infty = O(h_T^3) \text{ and } \hat{\alpha}_T - \alpha_\infty = o\left(\frac{1}{\sqrt{Th_T}}\right).$$

condition on the bandwidth

$$h_T = o(T^{-1/(d+4)})$$

condition on the kernel :

The kernel is of order one :

$$\int u_j K(u)du = 0, \forall j.$$

additional technical conditions

(A.5) i) There exists a neighbourhood $V(\theta_\infty(s))$ of $\theta_\infty(s)$ such that: In a neighbourhood of α_∞ and for any i, we have :

$$\begin{aligned}
&\left| K\left(\frac{S(Y,\alpha)-s}{h_T}\right)\frac{\partial\psi}{\partial\theta_i}(Y,\alpha;s;\theta) - K\left(\frac{S(Y,\alpha_\infty)-s}{h_T}\right)\frac{\partial\psi}{\partial\theta_i}(Y,\alpha_\infty;s;\theta)\right.\\
&\quad - \frac{dK}{du}\left(\frac{S(Y,\alpha_\infty)-s}{h_T}\right)\frac{\partial\psi}{\partial\theta_i}(Y,\alpha_\infty;s;\theta)\frac{\partial S}{\partial\alpha'}(Y,\alpha_\infty)\frac{\alpha-\alpha_\infty}{h_T}\\
&\quad \left. -K\left(\frac{S(Y,\alpha_\infty)-s}{h_T}\right)\frac{\partial^2\psi}{\partial\theta_i\partial\alpha'}(Y,\alpha_\infty;s;\theta)(\alpha-\alpha_\infty)\right|\\
&\quad \leq \frac{\|\alpha-\alpha_\infty\|^2}{h_T^2}N(Y;s;\theta), \forall Y, \theta\in V(\theta_\infty(s)).
\end{aligned}$$

ii) $\quad E\left[\sup_{\theta\epsilon V(\theta_\infty)} N(Y_o;s;\theta)\right] < \infty.$

iii) The following functions are bounded in a neighbourhood of s:

$$E\left[\sup_{\theta \epsilon V(\theta_\infty)} |\frac{\partial \psi}{\partial \theta_i}(Y_o, \alpha_\infty; s; \theta)| \parallel \frac{\partial S}{\partial \alpha'}(Y_o, \alpha_\infty) \parallel / S_o = .\right] f_S(.),$$

$$E\left[\sup_{\theta \epsilon V(\theta_\infty)} \parallel \frac{\partial^2 \psi}{\partial \theta_i \partial \alpha'}(Y_o, \alpha_\infty; s; \theta) \parallel / S_o = .\right] f_S(.).$$

(A.6) The following functions are bounded in a neighbourhood of s :

$$E\left[\sup_{\theta \epsilon V(\theta_\infty)} \parallel \frac{\partial^3 \psi}{\partial \theta_i \partial \theta \partial \theta'}(Y_o, \alpha_\infty; s; \theta) \parallel / S_o = .\right] f_S(.), \forall i.$$

(A.7) $\frac{\partial \psi}{\partial \theta}(., \alpha_\infty; s; \theta_\infty)$ satisfies $\mathcal{N}$.

(A.8) $\frac{\partial^2 \psi}{\partial \theta \partial \theta'}(., \alpha_\infty; s; \theta_\infty)$ satisfies $\mathcal{P}$.

(A.9) $\frac{\partial \psi}{\partial \theta}(., \alpha_\infty; s; \theta_\infty)\frac{\partial S}{\partial \alpha'}(., \alpha_\infty)$ satisfies $\mathcal{N}$.

(A.10) $E\left[\frac{\partial \psi}{\partial \theta}(Y_o, \alpha_\infty; s; \theta_\infty)\frac{\partial S}{\partial \alpha'}(Y_o, \alpha_\infty)/S_o = .\right] f_S(.)$

is continuously differentiable in a neighbourhood of s.

C - The expansion

From assumptions A.5 ii), iii) and A.6, (2.2) is equivalent to:

$$\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial \psi}{\partial \theta}(Y_t, \alpha_\infty; s; \theta_\infty(s))$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial^2 \psi}{\partial \theta \partial \theta'}(Y_t, \alpha_\infty; s; \theta_\infty(s))(\hat{\theta}_T(s) - \theta_\infty(s))$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} \frac{dK}{du}\left(\frac{S_t - s}{h_T}\right) \frac{\partial \psi}{\partial \theta}(Y_t, \alpha_\infty; s; \theta_\infty(s))\frac{\partial S}{\partial \alpha'}(Y_t; \alpha_\infty)\frac{\hat{\alpha}_T - \alpha_\infty}{h_T}$$

$$+0_p\left(\sqrt{Th_T} \parallel \hat{\alpha}_T - \alpha_\infty \parallel\right) + o_p\left(\sqrt{Th_T} \parallel \hat{\theta}_T(s) - \theta_\infty(s) \parallel\right) = 0,$$

and also to:

$$\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial \psi}{\partial \theta}(Y_t, \alpha_\infty; s; \theta_\infty(s))$$

$$+\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial^2 \psi}{\partial \theta \partial \theta'}(Y_t, \alpha_\infty; s; \theta_\infty(s))(\hat{\theta}_T(s) - \theta_\infty(s))$$

$$+\sqrt{Th_T}\left\{\sum_{t=1}^{T} \frac{1}{Th_T} \frac{dK}{du}\left(\frac{S_t - s}{h_T}\right) \frac{\partial \psi}{\partial \theta}(Y_t, \alpha_\infty; s; \theta_\infty(s)) \frac{\partial S}{\partial \alpha'}(Y_t; \alpha_\infty)\right.$$

$$\left. -E\left[\sum_{t=1}^{T} \frac{1}{Th_T} \frac{dK}{du}\left(\frac{S_t - s}{h_T}\right) \frac{\partial \psi}{\partial \theta}(Y_t, \alpha_\infty; s; \theta_\infty(s)) \frac{\partial S}{\partial \alpha'}(Y_t; \alpha_\infty)\right]\right\}$$

$$\times \frac{\hat{\alpha}_T - \alpha_\infty}{h_T}$$

$$+\sqrt{Th_T} E\left[\frac{1}{h_T} \frac{dK}{du}\left(\frac{S_o - s}{h_T}\right) \frac{\partial \psi}{\partial \theta}(Y_o, \alpha_\infty; s; \theta_\infty(s)) \frac{\partial S}{\partial \alpha'}(Y_o; \alpha_\infty)\right]$$

$$\times \frac{\hat{\alpha}_T - \alpha_\infty}{h_T}$$

$$+0_p(1) + o_p\left(\sqrt{Th_T} \parallel \hat{\theta}_T(s) - \theta_\infty(s) \parallel\right) = 0,$$

using the assumption for no effect of the first step estimation and the condition on the speed of convergence of $\hat{\alpha}_T$.

D - Analysis of the different remaining terms of the expansion

1) First term

We have : $E\left[\frac{\partial \psi}{\partial \theta}(Y_o, \alpha_\infty; s; \theta_\infty(s))/S_o = s\right] = 0$, because of the definition of θ_∞. From lemma 2, assumption A.7, conditions on the bandwidth and on the kernel we deduce the asymptotic normality :

$$\sum_{t=1}^{T} \frac{1}{\sqrt{Th_T}} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial \psi}{\partial \theta}(Y_t, \alpha_\infty; s; \theta_\infty(s)) \xrightarrow{d} N[0, \Omega(s)],$$

with $\Omega(s)$ given in (2.7).

2) Second term

From lemma 1 and assumption A.8 :

$$\sum_{t=1}^{T} \frac{1}{Th_T} K\left(\frac{S_t - s}{h_T}\right) \frac{\partial^2 \psi}{\partial\theta\partial\theta'}(Y_t, \alpha_\infty; s; \theta_\infty(s)),$$

tends in probability to $-J(s)f_S(s)$.

3) Third term

From lemma 2 and assumption A.9, the factor appearing before $\frac{\hat{\alpha}_T - \alpha_\infty}{h_T}$ is asymptotically normal. Therefore this term is negligible as soon as $\hat{\alpha}_T - \alpha_\infty = o(h_T)$, a condition which is satisfied.

4) Fourth term

We get :

$$\sqrt{Th_T}\, E\left[\frac{1}{h_T}\frac{dK}{du}\left(\frac{S-s}{h_T}\right)\frac{\partial\psi}{\partial\theta}(Y, \alpha_\infty; s; \theta_\infty(s))\frac{\partial S}{\partial\alpha'}(Y; \alpha_\infty\right]$$

$$= \sqrt{Th_T}\, E\left[\frac{1}{h_T}\frac{dK}{du}\left(\frac{S-s}{h_T}\right)\tilde{g}(S_o)\right],$$

where $\tilde{g}(S) = E\left[\frac{\partial\psi}{\partial\theta}(Y, \alpha_\infty; s; \theta_\infty(s))\frac{\partial S}{\partial\alpha'}(Y; \alpha_\infty)/S_o\right]$.

This term is also equal to :

$$\sqrt{Th_T} \int \frac{1}{h_T}\frac{dK}{du}\left(\frac{u-s}{h_T}\right)\tilde{g}(u)f_S(u)du$$

$$\sqrt{Th_T} \int \frac{dK}{du}(v)\tilde{g}(s + h_T v)f_S(s + h_T v)dv$$

$$= 0(\sqrt{Th_T}\, h_T),, \text{ since } \int \frac{dK}{du}dv = 0.$$

Therefore the fourth term is of order $0\left(\sqrt{Th_T}\, h_T \frac{\hat{\alpha}_T - \alpha_\infty}{h_T}\right)$ and then negligible since $\hat{\alpha}_T - \alpha_\infty = o\left(\frac{1}{\sqrt{Th_T}}\right)$.

References

Altman, N.S. (1988) : "Kernel Smoothing of Data with Correlated Errors", JASA.

Andrews, D. (1990) : "Generic Uniform Convergence", Cowles Foundation, D.P. 940.

Azzalini, A., Bowman, A. and Härdle, W. (1989) : "On the Use of Nonparametric Regression for Model Checking", Biometrika, 76, 1-12.

Benedetti, J.K. (1977) : "On the Nonparametric Estimation of Regression Functions", JRSS B, 39, 248-53.

Bierens, H.J. (1983) : "Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions, JASA, 77, 699-707.

Bierens, H.J. (1987) : "Kernel Estimators of Regression Functions", Cambridge Univ. Press.

Black, F. (1976) : "Studies of Stock Price Volatility Changes", Proceedings from the American Statistical Association, Business and Economic Statistics Section, 177-181.

Black, F. and Scholes, M. (1973) : "The Pricing of Options and Corporate Liabilities", Journal of Political Economy, 81, 637-659.

Bollerslev, T., Engle, R. and Nelson, D. (1993) : "ARCH Models", forthcoming in the Handbook of Econometrics, Vol. 4.

Bosq, D. (1988) : "Prévision nonparamétrique d'un processus stationnaire non borné", DP. LSTA 81, Univ. Paris VI, Paris.

Bosq, D., and Lecoutre J.P. (1987) : "Théorie de l'Estimation Fonctionnelle", Economica.

Burguete, J., Gallant, R. and Souza, G. (1982) : "On Unification of the Asymptotic Theory of Nonlinear Econometric Methods", Econometric Review, 1, 151-190.

Cheng, K.F. and R.L. Taylor (1980) : "On the Uniform Complete Convergence of Estimates for Multivariate Density Functions and Regression Curves", Ann. Inst. Stat. Math., 32, 187-199.

Cleveland, W. (1979) : "Robust Locally Werghted Regression and Smoothing Scatterplots", JASA, 74, 823-836.

Collomb, G. (1981) : "Estimation Nonparametrique de la Regression : Revue bibliographique", ISR, 49, 75-93.

Collomb, G. (1976) : "Estimation Nonparametrique de la Régression par la Méthode du Noyau", Thèse, Univ. P. Sabatier, Toulouse.

Collomb, G. (1977) : "Quelques Propriétés de la Méthode du Noyau pour l'Estimation Non Paramétrique de la Régression en un Point Fixé", C.R. Acad. Sc. Paris, 285, 289-292.

Collomb, G. and Hardle, W. (1986) : "Strong Uniform Convergence Rates in Robust Nonparametric Time Series Analysis and Prediction Kernel Regression Estimation From Dependent Observations", Stochastic Processes and Their Applications, 23, 77-89.

Devroye, L.P. (1979) : "The Uniform Convergence of the Nadaraya-Watson Regression Function Estimate", Canad. J. Statist. 6, 179-191.

Engle, R. (1993) : "Econometric Models for Financial Data", presented at the Econometric Society European Meeting.

Engle, R. and Ng, V. (1993) : "Measuring and Testing the Impact of News on Volatility", Journal of Finance, forthcoming.

Fan, J. (1992) : "Design Adaptive Nonparametric Regression", JASA, 87, 998-1004.

Fan, J. and I. Gijbels (1992) : "Spatial and Design Adaptation : Variable Order Approximation in Function Estimation", Institute of Statistics, Mimeo Series, 2080, Univ. of North Carolina, Chapel-Hill.

Fan, J., Heckman, N. and M. Wang (1992) : "Local Polynomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions, Univ. of British Columbia, WP 92-028.

Gallant, R.A. (1987) : "Nonlinear Statistical Models", Wiley.

Gallant, R., and Tauchen, G. (1989) : "Semi Non-Parametric Estimation of Conditionally Constrained Heterogeneous Processes : Asset Pricing Applications", Econometrica, 57, 1091-1120.

Gasser, T. and H. Müller (1979) : "Kernel Estimation of Regression Functions", in Lecture Notes in Mathematics : Smoothing Technologies for Curve Estimation, eds Gasser, T. and M. Rosenblatt, Springer, 23-68.

Gasser,T., Muller, H. and Mammitzsch, V. (1985) : "Kernels for Nonparametric Curve Estimation", JRSS B, 47, 238-252.

Gouriéroux, C. and Laurent, J.P. (1994) : "Estimating a Dynamic Hedging", CREST-DP.

Gouriéroux, C. and Monfort, A. (1989) : "Statistique et Modèles Econométriques, Economica.

Gouriéroux, C. and Monfort A., (1992) : "Qualitative Threshold ARCH Models", Journal of Econometrics, 52, 159-199.

Gouriéroux, C. and Monfort, A. (1993) : "Pseudo-Maximum Likelihood Methods", the Handbook of Statistics, Vol. 11, North-Holland, 335-362.

Gouriéroux, C., Monfort, A. and Trognon, A. (1984) : "Pseudo Maximum Likelihood Methods : Theory", Econometrica, 52, 681-700.

Gouriéroux, C. and Scaillet, O. (1994) : "Determination of the Term Structure from Bond Data : Correction for Cash-Flow Patterns", CREST-DP.

Greblicki, W. and Krzyzak, A. (1980) : "Asymptotic Properties of Kernel Estimates of Regression Function", Journal of Statistical Planning, 4, 81-90.

Györfi, L., Hardle, W., Sarda, P., and Vieu, P (1989) "Nonparametric Curve Estimation from Time Series", Springer Lecture Notes in Statistics, 60, Springer-Verlag.

Hardle, W. and J.S. Marron (1985) : "Optimal Bandwith Selection in Nonparametric Regression Function Estimation", Annals of Statistics, 13, 1465-1481.

Hardle, W. (1990) : "Applied Nonparametric Regression", Cambridge Univ. Press.

Hardle, W., and Linton, O. (1993) : "Applied Nonparametric Methods", Humboldt-Univ. Zu Berlin, DP 9312.

Huber, P.J. (1981) : "Robust Statistics", Wiley.

Hull, J. and White, A. (1987) : "The Pricing of Options on Assets with Stochastic Volatilities", The Journal of Finance, 3, 281-300.

Jennrich, R. (1969) : "Asymptotic Properties of Nonlinear Least Squares Estimators", Annals of Mathematical Statistics, 40, 633-643.

Mack, Y.P., and B. Silverman (1982) : "Weak and Strong Uniform Consistency of Kernel Regression Estimates", Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 61, 405-415.

Müller, H. (1984) : "Smooth Optimum Kernel Estimators of Densities, Regression Curves, and Modes", The Annals of Statistics, 15, 182-201.

Nadaraya, E.A. (1964) : "On Estimating Regression", Theory Prob. Appl., 10, 186-190.

Nadaraya, E.A. (1973) : "Some Limit Theorems Related to Nonparametric Estimates of Regression Curve", Bull. Acad. Sci. Georgian S.S.R., 71, 57-60 (in Russian).

Prakasa Rao, B.L.S. (1983) : "Non Parametric Functional Estimation" Academic Press.

Rice, J. (1984) : "Bandwith Choice for Nonparametric Regression", The Annals of Statistics, 12, 1215-1230.

Robinson, P. (1984) : "Robust Nonparametric Autoregression", in "Robust and Nonlinear Time Series Analysis", J. Franck et alii eds, Lecture Notes in Statistics, Vol 26, 247-255, Springer-Verlag.

Robinson, P. (1988) : "Semi-Parametric Econometrics : A Survey", Journal of Applied Econometrics, 3, 35-51.

Robinson, P. (1989) : "Nonparametric Estimation of Time Varying Parameters", in P. Hackl, ed., "Statistical Analysis and Forecasting of Economic Structural Change", 253-264, Springer Verlag.

Robinson, P. (....) : "Time Varying Nonlinear Regression", in "Economic Structural Change, Analysis and Forecasting", P. Hackl and A., Westland, eds. 179-190, Springer-Verlag.

Roussas, G. (1969) : "Nonparametric Estimation in Markov Processes", Ann. Inst. Statist. Math., 21, 73-87.

Schuster, E.F. (1972) : "Joint Asymptotic Distribution of the Estimated Regression Function at a Finite Number of Distinct Points", Ann. Math. Statist., 43, 84-88.

Schwert, G. (1989) : "Business Cycles, Financial Crisis and Stock Volatility", Carnegie Rochester Conference Series on Public Policy, 39, 83-126.

Staniswalis, J. (1989) : "The Kernel Estimate of a Regression Function in Likelihood-Based Models", JASA, 84, 276-283.

Stone, C.J. (1977) : "Consistent Nonparametric Regression", Ann. Statist., 5, 595-620.

Stone, C.J. (1980) : "Optimal Rates of Convergence of Nonparametric Estimators", Annals of Statistics, 8, 1348-1360.

Stone, C.J. (1982) : "Optimal Global Rates of Convergence for Nonparametric Regression", Annals of Statistics, 10, 1040-1053.

Tenreiro, C. (1993) : "Estimation non parametrique d'une fonction de moments conditionnels sous des conditions locales", Comptes rendus de l'Académie des Sciences, forthcoming.

Tibshirani, R. and Hastie, T. (1987) : "Local Likelihood Estimation", JASA, 82, 559-567.

Watson, G.S. (1964) : "Smooth Regression Analysis", Sankhya, Series A, 26, 359-372.

White, H. (1981) : "Maximum Likelihood Estimation of Misspecified Models", Econometrica, 50, 1-25.

Wiggins, J. (1987) : "Option Values Under Stochastic Volatility", Journal of Financial Economics, 19, 351-372.