A Global Encompassing Criterion for Nonparametric Regression Models

Christophe Bontemps Jean-Pierre Florens*
Gremaq Gremaq and IDEI

University of Social Sciences, Toulouse, France

January 10th, 95

Abstract

The concept of encompassing aims at validating or invalidating a tentative econometric model by testing its capacity to account for results obtained under alternative specifications.

1 Introduction

One of the most important scientific activity has been, and still is, the comparison of theories and/or models. The idea that a new model, or a new theory, must induce a progress in the knowledge of a phenomena is often emphasized, nevertheless, it seems to be equally important to check that this new model or theory, is able to explain what older, "*established*" models already explained.

Therefore, encompassing formalizes a research strategy which has been applied in many areas of science. This principle has been introduced in econometrics by Hendry, Mizon and Richard (see Mizon [33], Mizon and Richard [34], and Hendry and Richard [30]) and has been developed recently through the work of Gourieroux, Montfort and Trognon (see Gourieroux and Montfort [17] and [18], or Gourieroux, Montfort and Trognon [20]), Lu and Mizon [31], or Govaerts, Hendry and Richard [21] (see Bontemps [5] for

*We are grateful to Jean-François Richard for his comments on previous work.

a survey). Whithin a Bayesian context, Florens, Hendry and Richard [15], have recently provided a theoretical framework for interpreting encompassing as a notion of "sufficiency" among models. Each of these papers dealing with encompassing do present an application to the choice of regressors, applying parametric encompassing in either a "static" linear, a dynamic, or a Bayesian framework.

At a related increasing level, attention has been paid over recent years to the development of robust techniques of inference, with special emphasis on non-parametric methods in order to reduce the impact of errors of specification (see Collomb [9] and [10], Bierens [4], or the numerous publications of Härdle [24], Härdle and Marron, [27] Härdle and Mammen [26], etc...).

The contribution of this paper is that of integrating recent developments in these two areas of research. Our motivation is to describe and analyze the encompassing principle in regression models where we let the models be free of any functional shape.

The concept of encompassing used here is in line with that of parametric encompassing as defined by Mizon and Richard [34]. Formal definitions are offered below, but a brief presentation clarifies the encompassing procedure. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ denote two regression models of a variable Y with respective regressors X and Z, and let $\widehat{f_n}$ and $\widehat{g_n}$ be two estimators of the regression functions associated to these models. One then define the pseudo-true value $G(f)$ of $\widehat{g_n}$ as the plim under $\mathcal{M}_1$ of the latter. The basis of the comparison is the encompassing difference between $\widehat{g_n}$ and an estimator of its pseudo-true value $G(\widehat{f_n})$, which is here a function of z. Encompassing of $\mathcal{M}_2$ by $\mathcal{M}_1$ is realized if that difference (the lack of encompassing) is not significant relative to its asymptotic sampling distribution.

The paper is organized as follows : In the following section we introduce the notations and assumptions, the estimators associated to the models and the pseudo-true value are defined in section 3. The nonparametric encompassing statistics are then defined in section 4 where we derive a basic and a global statistic. In the same section we analyze the asymptotic behavior of these statistics, leading to a Bootstrap approach in section 5. The simulation results of this Bootstrap procedure are given in the final section.

2 Notations and models

Let the full set of variables required by, and available to, an investigator in order to estimate, test and analyze competing regression models be denoted by $S = (Y, X, Z)$. And let $(S_i)_{i=1,\dots,n}$ be n observations where $Y_i \in \Re$, $X_i \in \Re^p$, and $Z_i \in \Re^q$. The variables X_i and Z_i represent the conditioning variables associated with the models $\mathcal{M}_1$ and $\mathcal{M}_2$.

Formaly $(S_i)_{i=1,\ldots,n}$ constitute a centered, square integrable process defined on a probability space (Ω, A, P_0). The probability P_0 is unknown and we shall restrict our attention to functions defined from it.

We assume that the process $(S_i)_{i=1,\ldots,n}$ is *i.i.d.*, hence its distribution is fully characterized by a single observation which is itself described by its density[1] $\varphi(\cdot)$ with respect to the Lebesgue measure in $\Re^{p+q+1}$.

The components of (X_i, Z_i) are assumed to be linearly independent from each other. The latter assumption can be relaxed, allowing for common components in X_i and Z_i or, more generally, for lack of linear independence between components of X_i and Z_i. In such cases, the density φ would be taken with respect to the Lebesgue measure restricted to the appropriate subspace of $\Re^{p+q+1}$.

We use the following notations to represent the regression functions[2] :

$$\begin{aligned} f(\cdot) &= E[Y \mid X = \cdot] \\ g(\cdot) &= E[Y \mid Z = \cdot] \end{aligned}$$

The process $(S_i)_{i=1,\ldots,n}$ being square integrable, the functions f and g are themselves square integrable on (Ω, A, P_0).

In the rest of the paper we will assume the following regularity condition :

Hypothesis 2.1 *There exist a continuous version of the regression functions f and g as well as continuous versions of the joint, marginal and conditional densities (represented here by the same function φ).*

The "*encompassing*" model $\mathcal{M}_1$, to be validated, is based on the exclusion of the variable Z from the regression this exclusion is assumed through the following $\mathcal{H}_1$ hypothesis :

$$\mathcal{H}_1 \quad : \quad E[Y \mid X, Z] = E[Y \mid X]$$

which is a mean-conditional hypothesis.

We also need in our global statistic the following condition :

$$\mathcal{H}_2 \quad : \quad E\left[Y^2 \mid X, Z\right] = E\left[Y^2 \mid X\right]$$

so that the pair $(\mathcal{H}_1, \mathcal{H}_2)$ implies the equality of the conditional variances.

[1]For ease of notation, $\varphi(\cdot)$ will be generically used to represent the joint density of S_i, as well as its marginal or conditionnal densities, all ambiguities being resolved by the list of arguments.

[2]Expectation relative to P_0 are generically represented by the letter E.

The model *"to be encompassed"* $\mathcal{M}_2$ is based on the regression with Z_i as sole regressors. From the perspective of the "owner" of $\mathcal{M}_1$, this model is not of special interest, and is purely instrumental in the construction of encompassing tests aimed at validating $\mathcal{M}_1$.

In our context, both models will be associated with a nonparametric estimator of the regression function involved in its definition. A variety of nonparametric estimators of regression functions are now available (see Härdle[24]), but we will focus our attention on kernel estimators.

We use the Nadaraya [35] and Watson [41] estimator for f and g given by :

$$\widehat{f}_n(x) = \frac{\frac{1}{n.h_n^p}\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h_n}\right)}{\frac{1}{n.h_n^p}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}$$

with the condition $\widehat{f}_n(x) = 0$ if the denominator $\frac{1}{n.h_n^p}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)$(an estimator of the marginal density $\varphi(x)$) is zero.

In this expressions K denote a Parzen-Rosenblatt kernel i.e. an application from $\Re^p$ to $\Re$ which is integrable with respect to the Lebesgue measure, of integral one and which satisfies the limit condition :

$$\lim_{\|x\| \to \infty} \|x\|^p \cdot K(x) = 0$$

where $\|\cdot\|$ denotes the Euclidian norm.

The estimator of g is defined accordingly :

$$\widehat{g}_n(z) = \frac{\frac{1}{n.k_n^q}\sum_{i=1}^{n} Y_i K\left(\frac{Z_i - z}{k_n}\right)}{\frac{1}{n.k_n^q}\sum_{i=1}^{n} K\left(\frac{Z_i - z}{k_n}\right)}$$

In order to alleviate notation, we shall use a common notation K for the kernels involved in these estimators, tough they may obviously differ from each other, in particular for considerations of dimension.

The window-width h_n and k_n are also different and their convergence rates must be adjusted to the regressors dimensions. We shall assume the traditional convergence conditions for these sequences :

Hypothesis 2.2 *(Window-width minimal conditions)*

The window-width h_n and k_n satisfies :

$$\lim_{n \to \infty} h_n = 0 \qquad \textit{and} \qquad \lim_{n \to \infty} n \cdot h_n^p = \infty$$

$$\lim_{n\to\infty} k_n = 0 \qquad and \qquad \lim_{n\to\infty} n \cdot k_n^q = \infty$$

These conditions insure us of the consistency of $\widehat{f_n}(x)$ and $\widehat{g_n}(z)$ in their respective models.

In order to simplify this study, we shall assume an homoscedasticity condition.

Hypothesis 2.3 *(Homoscedasticity)*
Under $\mathcal{M}_1$, $Var[Y \mid X, Z] = \sigma^2$

The previous assumptions, *i.i.d.*, square integrability, continuous version (Hyp.2.1), window-width conditions (Hyp.2.2), and homoscedasticity (Hyp. 2.3) are maintained for the rest of the paper, we only list additional conditions.

3 Pseudo-true value

Encompassing being a notion linking "estimated models", that is models associated to estimators (see Florens, Hendry and Richard [15], or Bontemps [5], for a general discussion of encompassing), we have to derive the convergence properties of these estimators. In line with the objectives of our encompassing analysis we shall derive their limits under the mean-conditional independence assumption associated with $\mathcal{M}_1$.

Theorem 3.1 *Under $\mathcal{H}_1$, we have :*

$$\begin{array}{llll} i) & \widehat{f_n}(x) \overset{n\to\infty}{\longrightarrow} f(x) & & \forall x \\ ii) & \widehat{g_n}(z) \overset{n\to\infty}{\longrightarrow} E\left[f(x) \mid Z = z\right] & & \forall z \end{array}$$

Proof :

The proof follows from the properties of kernel estimators (see Bosq and Lecoutre[7]), under the minimal conditions assumed, a kernel estimator of a conditional expectation tends in probability toward the latter in every point, so :

$$\begin{array}{llll} i) & \widehat{f_n}(x) & \overset{n\to\infty}{\longrightarrow} E\left[Y \mid X = x\right] & \forall x \\ ii) & \widehat{g_n}(z) & \overset{n\to\infty}{\longrightarrow} E\left[Y \mid Z = z\right] & \forall z \end{array}$$

Under $\mathcal{H}_1$, we have :

$$\begin{aligned} E[Y \mid Z=z] &= E[E[Y \mid X=x, Z=z] \mid Z=z] \\ &= E[E[Y \mid X=x,] \mid Z=z] \\ &= E[f(x) \mid Z=z] \end{aligned}$$

□

This property lead us to the definition of the pseudo-true value associated to $\widehat{g_n}(z)$ on $\mathcal{H}_1$. It is defined in the same spirit than the classical "parametric" pseudo-true value associated to an estimator of a parameter of interest in $\mathcal{M}_2$. According to Hendry and Richard[30], the pseudo-true value of $\widehat{g_n}$ is given by the plim under $\mathcal{M}_1$ of the latter.

Definition 3.1 *The pseudo-true value G associated to $\widehat{g_n}(z)$ on $\mathcal{H}_1$ is*

$$G(f)(z) = E[f(x) \mid Z=z]$$

This pseudo-true value is a reinterpretation of $\widehat{g_n}(z)$ under the belief that $\mathcal{M}_1$ is the "true" model. It is estimated in the same way than $\widehat{g_n}(z)$ by $\widehat{G}(f)(z)$:

$$\widehat{G}(f)(z) = \frac{\frac{1}{n.k_n^q} \sum_{i=1}^{n} f(X_i) \cdot K\left(\frac{Z_i - z}{k_n}\right)}{\frac{1}{n.k_n^q} \sum_{i=1}^{n} K\left(\frac{Z_i - z}{k_n}\right)}$$

4 Nonparametric encompassing

4.1 Basic statistic

As usual in a nonparametric framework, we have to assume some regularity assumption and to impose some arbitrary rates of smoothness for the functions involved in our statistics.

Hypothesis 4.1 *(Regularity)*

- *The densities and conditional means derived from (X_i, Y_i, Z_i) are d-times continuously differentiable and bounded.*
- *The marginal densities $\varphi(\cdot)$ are positive on their support, the latter is assume to be compact.*

Hypothesis 4.2 *(Kernels orders)*

The Parzen-Rosenblatt kernels involved in $\widehat{f_n}$ and $\widehat{g_n}$ must satisfy the following conditions :

$$\int_{\Re^p} \prod_{i=1}^{p} x_i^{a_i} K(x_1, x_2, \ldots, x_p) dx_1 \cdots dx_p = \begin{cases} 1 & if \quad a_i = 0 \ , \ \forall\, i = 1, \ldots, p \\ 0 & if \quad 0 < \sum_{i=1}^{p} a_i < d \end{cases}$$

$$and \qquad \int_{\Re^p} |x_i|^m \, |K(x_1, x_2, \ldots, x_p)| \, dx_1 \cdots dx_p < \infty \quad , \ \forall x \in \Re^p$$

In order to have positive kernels, we impose d=2.

Let h_n and k_n be the window-width associated to the estimators $\widehat{f_n}(x)$ and $\widehat{g_n}(z)$ respectively, these window-widths must verify :

Hypothesis 4.3

$$\lim_{n\to\infty} n \cdot h_n^{p+2d} = 0 \qquad and \qquad \lim_{n\to\infty} n \cdot k_n^{q+2d} = 0$$

*These latter conditions are instruments to "kill the bias" remaining asymptotically from the estimation of the regression functions and are standard in nonparametric estimation(see Bierens [**?**]).*

To these three hypothesis which insure us of the asymptotic normality of the estimators we must add a condition on the relation between the rates of convergence of the two smoothing parameters used :

Hypothesis 4.4 *The window-widths h_n and k_n must moreover satisfy :*

$$\log(n) \cdot \frac{k_n^q}{h_n^p} \overset{n\to\infty}{\longrightarrow} 0$$

In the univariate case (p=q=1), the latter conditions means heuristically that the window-width k_n must converge to zero "faster" than h_n.

The encompassing statistic in this nonparametric context, is based on the encompassing difference, which is here the function $\delta_{f,g}(z)$ built on the difference between the estimator associated to $\mathcal{M}_2$, $\widehat{g_n}$ and an estimator of its pseudo-true value, $\widehat{G}(\widehat{f_n})$, i.e. :

$$\delta_{f,g}(z) = \widehat{g_n}(z) - \widehat{G}(\widehat{f_n})(z)$$

Under the previous hypothesis, Bontemps, Florens and Richard [6], have derived the asymptotic behavior of this statistic, once normalized.

Theorem 4.1 *Under $\mathcal{H}_1$, and under the hypothesis 4.1 and 4.2, and if the window-with satisfy the hypothesis 4.3 and 4.4, we get :*

$$\sqrt{n \cdot k_n^q} \cdot \delta_{f,g}(z) \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^2 \int K^2}{\varphi(z)}\right)$$

A proof of this result is given in the mathematical appendix.

The asymptotic behavior of this statistic, is in line with the (parametric) normality obtained in the classic asymptotic study of the encompassing statistic in parametric cases (see Hendry and Richard [30]). Nevertheless, the local (because functional) characteristic of this statistic may be disappointing, or ambiguous to use in practice. These considerations motivate the construction of a global encompassing criterion, similar the quadratic parametric criterion, provided by Mizon [33], or Mizon and Richard [34], and which may be more useful.

This statistic may rest on different types of criteria :

- An empirical quadratic criterion :

$$\Xi = \alpha_1(n) \cdot \sum_{l=1}^{L} \left(\delta_{f,g}(Z_l)\right)^2 \varpi\left(Z_l\right)$$

  Where $(Z_l)_{l=1}^{L}$ are L arbitrary values of Z, and ϖ is a weight function.

- An integral type criterion :

$$\Psi = \alpha_2(n) \cdot \int \left(\delta_{f,g}(z)\right)^2 \varpi\left(z\right) dz$$

  Where $\varpi\left(\cdot\right)$ is a weight function.

Or any criterion giving a global vision of the encompassing difference, using any type of distance or norm.

The convergence rates $\alpha_1(n)$ and $\alpha_2(n)$ are then chosen in order to derive an asymptotic convergence to a distribution upon which the test is based. In a parametric context, the parametric "n" rate of convergence is used to derive a χ^2 distribution from which are derived Wald Encompassing Tests (see Hendry and Richard [30], or Lu and Mizon [31]. Obviously in this nonparametric framework, these rates may involve the window-width in their definition.

4.2 A global Criterion

After some work, the simulations on the first criterion gave very disappointing results, so we present here a more interesting result concerning the "integral type" criterion which is :

$$\Psi = n \cdot \sqrt{k_n} \cdot \int \left(\delta_{f,g}(z)\right)^2 \varpi(z)\, dz$$

The asymptotic behavior of this criterion has been established and is presented in this section. We need to introduce some additional hypothesis to derive the next result. For simplicity we place this work in the univariate case (p=q=1)

The existence and regularity of the following conditional moment are assumed

Hypothesis 4.5 *Let f_4 denote the following function*

$$f_4(x) = E\left[(Y - f(x))^4 \mid X = x\right]$$

we assume that

$$\int f_4(x)\varphi(x)dx < \infty$$

As in the previous section we need to impose a condition on the relation between the rates of convergence of the two smoothing parameters h_n and k_n.

Hypothesis 4.6 *The window-width h_n and k_n must verify :*

$$\log(n) \cdot \frac{\sqrt{k_n}}{h_n} \overset{n\to\infty}{\longrightarrow} 0$$

Theorem 4.2 *Under $(\mathcal{H}_1, \mathcal{H}_4)$, and under the hypothesis 4.1, 4.2 and 4.5, if the window-widths h_n and k_n, satisfy the hypothesis 4.3 and 4.6, we have :*

$$\begin{aligned} n \cdot \sqrt{k} \int \left(\delta_{f,g}(z)\right)^2 \varpi(z)dz = \frac{1}{\sqrt{k}}\, \sigma^2 \cdot \int K^2(u)\nu_n(u)du \\ + n\sqrt{k} \cdot I_{n,2} \\ + O_p(k) + O_p\left(\frac{1}{\sqrt{n \cdot k}}\right) \end{aligned} \quad (1)$$

In this expression, we have :

- The term $n\sqrt{k}\ I_{n,2}$ converges to an asymptotic centered normal distribution :

  $$n \cdot \sqrt{k} \cdot I_{n,2} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \alpha_1\right)$$

  where :

  $$\alpha_1 = 2 \cdot \sigma^4 \left[\int \varphi^2(x)\nu(x)dx\right] \cdot \left[\int \left[\int K(u)K(u+v)du\right]^2 dv\right]$$

- The weight function $\varpi(z)$ is chosen for simplicity :

  $$\varpi(z) = \widehat{\varphi_n}^2(z) \cdot \nu_n(z)$$

  where $\nu_n(z)$ is a random measurable function, and where $\widehat{\varphi_n}(z)$ is the Nadaraya-Watson estimator of the marginal density φ. Furthermore, we suppose that there exist a deterministic function ν such that :

  $$\sup_{x \in \Re} \left| \frac{\nu_n(z)}{\nu(z)} - 1 \right| \xrightarrow{P} 0$$

as $n \to \infty$

Remark :

This result is similar to a result given in Härdle and Mammen [**?**], in the context of the comparison of a parametric and a nonparametric curve estimate, based on the following criterion :

$$T_n = n\sqrt{h^p} \int \left(\widehat{f}(x) - F(f_{\widehat{\theta}})(x)\right)^2 \varpi(x)dx$$

where $F(f_{\widehat{\theta}})(x)$ is a kernel estimator of the regression of $f_\theta(x)$ (the parametric function) on Z_i's for $\theta = \widehat{\theta}$:

$$F(f_{\widehat{\theta}})(x) = \frac{\sum_i f_{\widehat{\theta}}(X_i) \cdot K\left(\frac{X_i - x}{h}\right)}{\sum_i K\left(\frac{X_i - x}{h}\right)}$$

The remaining bias $\frac{1}{\sqrt{k}}\ \sigma^2 \cdot \int K^2(u)\nu_n(u)du$ in the asymptotic decomposition of our global statistic may be estimated easily in order to obtain an unbiased statistic, but it seems to be more useful to consider alternative methods. We will study in the next section Bootstrap methods as an alternative to asymptotic.

5 Bootstrap

.

6 Mathematical Appendix

0.1 Proof of theorem 4.2

Let $\psi = \int \left(\delta_{f,g}(z)\right)^2 \varpi(z)dz$ be the object of our asymptotic study.

Using our definition for the weight function $\varpi(z) = \widehat{\varphi}^2(z) \cdot \nu_n(z)$ we get :

$$\psi = \frac{1}{n^2k^2} \int \left(\sum_i \left(Y_i - \widehat{f}(X_i)\right) K\left(\frac{Z_i - z}{k}\right)\right)^2 \nu_n(z)dz$$

which may be decomposed as :

$$\begin{aligned} \psi &= \tfrac{1}{n^2k^2} \int \left(\sum_i \left(Y_i - f(X_i)\right) K\left(\tfrac{Z_i - z}{k}\right) + \sum_j \left(f(X_j) - \widehat{f}(X_j)\right) K\left(\tfrac{Z_j - z}{k}\right)\right)^2 \nu_n(z)dz \\ &= \tfrac{1}{n^2k^2} \int \left(Q_1(z) + Q_2(z)\right)^2 \nu_n(z)dz \end{aligned}$$

with :

$$\begin{aligned} & Q_1(z) &= \sum_i \left(Y_i - f(X_i)\right) K\left(\tfrac{Z_i - z}{k}\right) \\ and & & \\ & Q_2(z) &= \sum_j \left(f(X_j) - \widehat{f}(X_j)\right) K\left(\tfrac{Z_j - z}{k}\right) \end{aligned}$$

From this expression we get **three** terms corresponding to the integral of Q_1 and Q_2 squared and to the cross product respectively :

$$\begin{aligned} F_1 &= \tfrac{1}{n^2k^2} \int \left[\sum_i \left(Y_i - f(X_i)\right) K\left(\tfrac{Z_i - z}{k}\right)\right]^2 \nu_n(z)dz \\ F_2 &= \tfrac{1}{n^2k^2} \int \left[\sum_j \left(f(X_j) - \widehat{f}(X_j)\right) K\left(\tfrac{Z_j - z}{k}\right)\right]^2 \nu_n(z)dz \\ and & \\ F_3 &= \tfrac{1}{n^2k^2} \int \left[\sum_i \left(Y_i - f(X_i)\right) K\left(\tfrac{Z_i - z}{k}\right)\right] \cdot \left[\sum_j \left(f(X_j) - \widehat{f}(X_j)\right) K\left(\tfrac{Z_j - z}{k}\right)\right] \nu_n(z)dz \end{aligned}$$

From a global point of view, we have **:**

$$\psi = F_1 + F_2 + 2 \cdot F_3$$

and we'll prove that

- F_1 gave us the asymptotic normality, and bias (as in Hall [23]) :

$$\begin{aligned} F_1 &= \tfrac{1}{nk}\,\sigma^2 \cdot \textstyle\int K^2(u)\nu_n(u)du \\ &+I_{n,2} \\ &+O_p\left(\tfrac{1}{\sqrt{n^3 \cdot k^2}}\right) \end{aligned}$$

with

$$\left(n \cdot \sqrt{k}\right) \cdot I_{n,2} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4} \cdot \alpha_1\right)$$

- $n\sqrt{k} \cdot F_2 \to 0$ in probability. This is due to our bandwidths hypothesis 4.3 and 4.6, and because :

$$F_2 \leq O_p\left(Max\left(\frac{\log(n)}{n \cdot h}\ ,\ h^{2 \cdot d}\right)\right) \cdot \int \widehat{\varphi}^2(z)\nu_n(z)dz$$

- $n\sqrt{k} \cdot F_3 \longrightarrow 0$ in probability. This convergence is quite slow because :

$$n\sqrt{k} \cdot F_3 = O_p\left(k\right)$$

0.1.1 Study of $\mathbf{F}_1$

The expression of $\mathbf{F}_1$ is :

$$F_1 = \frac{1}{n^2 k^2} \int \left[\sum_i \left(Y_i - f(X_i)\right) K\left(\frac{Z_i - z}{k}\right)\right]^2 \nu_n(z)dz$$

We can separate this study in two parts, using the following decomposition :

$$\begin{aligned} \mathbf{F}_1 &= \frac{1}{n^2 k^2} \int \sum_i \left(Y_i - f(X_i)\right)^2 K^2\left(\frac{Z_i - z}{k}\right) \nu_n(z)dz \\ &\quad +2 \cdot \frac{1}{n^2 k^2} \int \sum_{i,j} \sum_{i<j} \left(Y_j - f(X_j)\right) K\left(\frac{Z_j - z}{k}\right) \left(Y_i - f(X_i)\right) K\left(\frac{Z_i - z}{k}\right) \nu_n(z)dz \\ &= I_{n,1} + 2 \cdot I_{n,2} \end{aligned}$$

These terms $I_{n,1}$ and $I_{n,2}$ have a very different asymptotic behavior, we show that :

i)

$$I_{n,1} = \frac{1}{nk} \cdot \sigma^2 \left(\int K^2(u)\nu_n(u)du \right) + O_p\left(\frac{1}{\sqrt{n^3 \cdot k^2}} \right) \tag{2}$$

while

ii)

$$\left(n \cdot \sqrt{k} \right) \cdot I_{n,2} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4} \cdot \alpha_1 \right) \tag{3}$$

- **Study of** $I_{n,1}$ **:**

We shall note E' the expectation conditional to X_i and Z_i :
First we show that :

$$(nk)^4 \cdot E'\left[\left(I_{n,1} - E'\left[I_{n,1} \right] \right)^2 \right] = \sum_i E'\left[\left(Y_i - f(X_i) \right)^2 - \sigma^2 \right]^2 \cdot \left[\int K^2\left(\frac{Z_i - z}{k} \right) \nu_n(z)dz \right]^2$$

we note that :

$$\int K^2\left(\frac{Z_i - z}{k} \right) \nu_n(z)dz \leq k \int K^2\left(u \right) \nu(u)du$$

Under 4.5 and the weak low of large numbers :

$$\begin{aligned} (nk)^4 \cdot E'\left[\left(I_{n,1} - E'\left[I_{n,1} \right] \right)^2 \right] &= O_p\left((nk)^2 \right) \sum_i f_4\left(X_i \right) \\ &= O_p\left(nk^2 \right) \end{aligned}$$

It now follows via the Chebyshev's inequality that if $(\lambda_n)_{n \in N}$ is any sequence of constants diverging to ∞,

$$\Pr\left\{ \left| I_{n,1} - E'\left[I_{n,1} \right] \right| > \lambda_n \cdot \sqrt{\frac{nk^2}{(nk)^4}} \mid X_i, Z_i \right\} \xrightarrow{n \to \infty} 0$$

so

$$I_{n,1} = E'\left[I_{n,1} \right] + O_p\left(\sqrt{\frac{nk^2}{(nk)^4}} \right)$$

which proves $i)$

- **Study of** $I_{n,2}$

This term gives the asymptotic normality in (1)

$$(nk)^2 I_{n,2} = \frac{1}{n^2k^2} \sum_{i,j} \sum_{i<j} (Y_j - f(X_j))(Y_i - f(X_i)) \int K\left(\frac{Z_j - z}{k}\right) K\left(\frac{Z_i - z}{k}\right) \nu_n(z)dz$$

We may rewrite this term as :

$$(nk)^2 I_{n,2} = \sum_{i,j} \sum_{i<j} (Y_j - f(X_j))(Y_i - f(X_i)) W_{n,i,j}$$

or

$$(nk)^2 I_{n,2} = \sum_{i=2}^{n} Y_{n,i} \quad (4)$$

where

$$Y_{n,i} = (Y_i - f(X_i)) \sum_{j=1}^{i-1} (Y_j - f(X_j)) W_{n,i,j} \quad 2 \le i \le n$$

Let $\mathcal{F}_{n,i}$ denotes the σ-field generated by $S_1, \cdots, S_n$ (where $S_i = (X_i, Y_i, Z_i)$) then

$$E[Y_{n,i} \mid \mathcal{F}_{n,i}] = 0 \quad a.s. \quad \forall i$$

therefore the sequence

$$\left\{ \left(M_{n,i} = \sum_{j=2}^{i} Y_{n,j} \ , \mathcal{F}_{n,i} \right) \quad , 2 \le i \le n \right\}$$

is a martingale triangular array (see Hall[23]).The conditional variance of $M_{n,n}$ is given by :

$$V_n = \sum_{i=2}^{n} E\left[Y_{n,i}^2 \mid \mathcal{F}_{n,i-1}\right] = \sum_{i=2}^{n} \sigma^2 \left[\sum_{j=1}^{i-1} (Y_j - f(X_j)) W_{n,i,j}\right]^2$$

Which may be cut into the sum of the squared terms and the double crossed-product :

$$\begin{aligned} V_n &= n.\sigma^2 \sum_{j=1}^{i-1} (Y_j - f(X_j))^2 W_{n,i,j}^2 \\ &+2n\sigma^2 \sum_{1 \le j \le l \le i-1} (Y_j - f(X_j))(Y_l - f(X_l)) W_{n,i,j} W_{n,i,l} \\ &= V_{n,1} + V_{n,2} \end{aligned}$$

Hall ([23] lemma 1 and 2), on the basis of a central limit theorem due to Brown [8], give us the latter result :

$$\frac{1}{n \cdot k^{3/2}} \cdot M_{n,n} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4}\alpha_1\right)$$

where α_1 is defined by :

$$\frac{1}{n^2 \cdot k^3} \cdot V_{n,1} \longrightarrow \frac{1}{4}\alpha_1$$

and

$$\alpha_1 = 2 \cdot \sigma^4 \left[\int \varphi^2(x)\nu(x)dx\right] \cdot \left[\int \left[\int K(u)K(u+v)du\right]^2 dv\right]$$

reporting this result in (4), we get :

$$\frac{1}{n \cdot k^{3/2}} \cdot M_{n,n} = \left(n \cdot \sqrt{k}\right) \cdot I_{n,2} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4} \cdot \alpha_1\right)$$

which proves (3) and therefore ii).

so :

$$\begin{aligned} F_1 &= \tfrac{1}{nk}\,\sigma^2 \cdot \textstyle\int K^2(u)\nu_n(u)du \\ &+2 \cdot I_{n,2} \\ &+O_p\left(\tfrac{1}{\sqrt{n^3 \cdot k^2}}\right) \end{aligned}$$

$\square_1$

0.1.2 Study of $\mathbf{F}_2$

We have to check that $n \cdot \sqrt{k} \cdot F_2$ disappears asymptotically, where F_2 is :

$$F_2 = \frac{1}{n^2 k^2} \int \left[\sum_j \left(f(X_j) - \widehat{f}(X_j)\right) K\left(\frac{Z_j - z}{k}\right)\right]^2 \nu_n(z)dz$$

The study of this term is simplified by the use of the hypothesis 4.6, because

$$F_2 \leq \left(\sup_{X_j} \left|f(X_j) - \widehat{f}(X_j)\right|\right)^2 \int \left[\frac{1}{nk} \cdot \sum_j K\left(\frac{Z_j - z}{k}\right)\right]^2 \nu_n(z)dz$$

and

- $\left(\sup_{X_j}\left|f(X_j)-\widehat{f}(X_j)\right|\right)^2 \longrightarrow 0$
- $\int\left[\frac{1}{nk}\cdot\sum_j K\left(\frac{Z_j-z}{k}\right)\right]^2\nu_n(z)dz \longrightarrow \int\varphi^2(z)\cdot\nu_n(z)dz$

From Györfi, Härdle, Sarda and Vieu [22], (or Bierens [4]), we have, under the hypothesis 4.2 and 4.3 :

$$\left(\sup_{X_j}\left|f(X_j)-\widehat{f}(X_j)\right|\right) = O_p\left(Max\left(\frac{\sqrt{\log(n)}}{\sqrt{n\cdot h}}\ ,\ h^d\right)\right)$$

so that

$$\left(\sup_{X_j}\left|f(X_j)-\widehat{f}(X_j)\right|\right)^2 = O_p\left(Max\left(\frac{\log(n)}{n\cdot h}\ ,\ h^{2\cdot d}\right)\right)$$

then, with $d=2$:

$$\left(n\cdot\sqrt{k}\right)\cdot F_2 \leq O_p\left(Max\left(\frac{n\cdot\sqrt{k}\cdot\log(n)}{n\cdot h}\ , n\cdot\sqrt{k}\ h^4\right)\right)\cdot\int\widehat{\varphi}^2(z)\nu_n(z)dz$$

Under 4.6, we check that :

$$\frac{n\cdot\sqrt{k}\cdot\log(n)}{n\cdot h} \overset{n\to\infty}{\longrightarrow} 0$$

and

$$n\cdot\sqrt{k}\ h^4 \leq n\cdot h^5 \overset{n\to\infty}{\longrightarrow} 0$$

So that the term $\left(n\cdot\sqrt{k}\right)\cdot F_2$ tend to 0 in probability. $\square_2$

0.1.3 Study of $\mathbf{F_3}$

$$\begin{aligned}
F_3 &= \frac{1}{n^2k^2}\int\left[\sum_i\left(Y_i-f(X_i)\right)K\left(\frac{Z_i-z}{k}\right)\right]\left[\sum_j\left(f(X_j)-\widehat{f}(X_j)\right)K\left(\frac{Z_j-z}{k}\right)\right]\nu_n(z)dz \\
&= \frac{1}{n^2k}\sum_i\sum_j\left(Y_i-f(X_i)\right)\left(f(X_j)-\widehat{f}(X_j)\right)\int\frac{1}{k}K\left(\frac{Z_i-z}{k}\right)K\left(\frac{Z_j-z}{k}\right)\nu_n(z)dz
\end{aligned}$$

The term involving the kernel product may be viewed as

$$\int\frac{1}{k}\cdot\Upsilon(z)K\left(\frac{Z_j-z}{k}\right)dz$$

with $\Upsilon(z) = K\left(\frac{Z_i - z}{k}\right)\nu_n(z)$ and is asymptotically equivalent to :

$$K\left(\frac{Z_j - Z_i}{k}\right)\nu(Z_i)$$

So, we shall concentrate our attention on the study of

$$\widetilde{F_3} = \frac{1}{n^2 k}\sum_i \sum_j (Y_i - f(X_i))\left(f(X_j) - \widehat{f}(X_j)\right) K\left(\frac{Z_j - Z_i}{k}\right)\nu(Z_i)$$

For $i = 1, \cdots, n$, let $\varepsilon_i = Y_i - f(X_i)$. In order to prove that $n \cdot \sqrt{k}\,\widetilde{F_3} \to 0$ in probability, we are going to study the two first moments of $\widetilde{F_3}$.

Obviously

$$E\left[\widetilde{F_3}\right] = 0$$

Let us study $\widetilde{F_3}^2$

$$\widetilde{F_3}^2 = \frac{1}{n^4 k^2}\left(\sum_i \sum_j \varepsilon_i \cdot \left(f(X_j) - \widehat{f}(X_j)\right) \cdot K\left(\frac{Z_j - Z_i}{k}\right)\nu(Z_i)\right)^2$$

Before developing this sum, we may introduce the expression $f(X_j) - \widehat{f}(X_j)$ in order to obtain the triple sum (squared) :

$$\widetilde{F_3}^2 = \frac{1}{n^4 k^2}\left(\sum_i \sum_j \frac{1}{nh}\sum_l \varepsilon_i \cdot \varepsilon_l \cdot K\left(\frac{X_l - Xj}{h}\right) K\left(\frac{Z_j - Z_i}{k}\right)\frac{\nu(Z_i)}{\widehat{\varphi}(X_j)}\right)^2$$

If we develop this sum we have :

$$\begin{aligned}\widetilde{F_3}^2 &= \tfrac{1}{n^4 k^2}\tfrac{1}{n^2 h^2}\sum_{i,j,l,i',j'l'} \varepsilon_i \varepsilon_{i'} \varepsilon_l \varepsilon_{l'} \cdot K\left(\tfrac{X_l - Xj}{h}\right) K\left(\tfrac{X_{l'} - Xj'}{h}\right) \\ &\times K\left(\tfrac{Z_j - Z_i}{k}\right) K\left(\tfrac{Z_{j'} - Z_{i'}}{k}\right) \cdot \tfrac{\nu(Z_i)}{\widehat{\varphi}(X_j)} \cdot \tfrac{\nu(Z_{i'})}{\widehat{\varphi}(X_{j'})}\end{aligned}$$

The interest of this decomposition is not to confuse the reader, but to show that the expectation of the terms involved in this sum is **zero**, except for a few values of $(i, j, l, i', j'l')$. In other words

$$\begin{aligned}& E\left[\varepsilon_i \varepsilon_{i'} \varepsilon_l \varepsilon_{l'} \cdot K\left(\frac{X_l - Xj}{h}\right) K\left(\frac{X_{l'} - Xj'}{h}\right) K\left(\frac{Z_j - Z_i}{k}\right) K\left(\frac{Z_{j'} - Z_{i'}}{k}\right) \cdot \frac{\nu(Z_i)}{\widehat{\varphi}(X_j)}\frac{\nu(Z_{i'})}{\widehat{\varphi}(X_{j'})}\right] \\ = \;& 0\end{aligned}$$

Except

- for the $i = i' = l = l'$ case where we have the sum over **three** letters , i, j et j' of this expectation (denoted below as "the A case")

- for the three cases where we get "couples", that is : $C_1 : (i = i' \;\; et \;\; l = l')$; $C_2 : (i = l \;\; et \;\; i' = l')$ and $C_3 : (i = l' \;\; et \;\; i' = l)$.

- In these cases we have the sum over **four** letters (j, j', i and l in the first case)

Following this decomposition of the indexes, we have :

$$
\begin{aligned}
E\left[\widetilde{F_3}^2\right] &= \tfrac{1}{n^6k^2}\tfrac{1}{h^2}\sum_{i,j,j'} E[A] \\
&+ \tfrac{1}{n^6k^2}\tfrac{1}{h^2}\sum_{j,j',i,l} E[C_1] \\
&+ \tfrac{1}{n^6k^2}\tfrac{1}{h^2}\sum_{j,j',i,i'} E[C_2] \\
&\tfrac{1}{n^6k^2}\tfrac{1}{h^2}\sum_{j,j',i,l} E[C_3]
\end{aligned}
$$

Let us study first

$$
\begin{aligned}
\tfrac{1}{n^6k^2}\tfrac{1}{h^2}\sum_{i,j,j'} E[A] &= \tfrac{1}{n^4}\left(\textstyle\sum_i f_4(X_i)\right. \\
&\times \sum_{j,j'} \tfrac{1}{nh^2} K\left(\tfrac{X_i - Xj}{h}\right) K\left(\tfrac{X_i - Xj'}{h}\right) \\
&\cdot \tfrac{1}{nk^2} K\left(\tfrac{Z_j - Z_i}{k}\right) K\left(\tfrac{Z_{j'} - Z_i}{k}\right) \cdot \left.\tfrac{\nu(Z_i)}{\widehat{\varphi}(X_j)}\tfrac{\nu(Z_i)}{\widehat{\varphi}(X_{j'})}\right)
\end{aligned}
$$

in reorganizing the terms, we get :

$$
\begin{aligned}
\tfrac{1}{n^6k^2}\tfrac{1}{h^2}\sum_{i,j,j'} E[A] &= \tfrac{1}{n^4}\left(\textstyle\sum_i f_4(X_i)\right. \\
&\times \textstyle\sum_j \tfrac{1}{nh^2} K\left(\tfrac{X_i - Xj}{h}\right) K\left(\tfrac{Z_j - Z_i}{k}\right) / \widehat{\varphi}(X_j) \\
&\cdot \textstyle\sum_{j'} \tfrac{1}{nk^2} K\left(\tfrac{X_i - Xj'}{h}\right) K\left(\tfrac{Z_{j'} - Z_i}{k}\right) / \widehat{\varphi}(X_{j'}) \cdot \left.\nu^2(Z_i)\right)
\end{aligned}
$$

the term

$$
\sum_j \frac{1}{nh^2} K\left(\frac{X_i - Xj}{h}\right) K\left(\frac{Z_j - Z_i}{k}\right) / \widehat{\varphi}(X_j) \longrightarrow \frac{\varphi(X_i, Z_i)}{\varphi(X_i)} = \varphi(Z_i \mid X_i)
$$

is assumed to be bounded sa under 4.5, and the low of the large numbers

$$\frac{1}{n^6k^2}\frac{1}{h^2}\sum_{i,j,j'}E\left[A\right]=O_p(1)\cdot\frac{1}{n^4}\sum_i f_4(X_i)\nu^2(Z_i)=O_p\left(\frac{1}{n^3}\right)$$

then, the term $\frac{1}{n^6k^2}\frac{1}{h^2}\ \sum_{j,j',i,l}\ E\left[C_1\right]$,

$$\begin{aligned}\tfrac{1}{n^6k^2}\tfrac{1}{h^2}\sum_{i,j,j'}E\left[C_1\right] &= \tfrac{1}{n^4}\Bigg(\sum_{i,l}\sigma^2\sigma^2\\ &\times\sum_{j,j'}\tfrac{1}{nh^2}K\left(\tfrac{X_i-Xj}{h}\right)K\left(\tfrac{X_i-Xj'}{h}\right)\\ &\times\tfrac{1}{nk^2}K\left(\tfrac{Z_j-Z_i}{k}\right)K\left(\tfrac{Z_{j'}-Z_i}{k}\right)\cdot\tfrac{\nu(Z_i)}{\widehat{\varphi}(X_j)}\tfrac{\nu(Z_i)}{\widehat{\varphi}(X_{j'})}\Bigg)\end{aligned}$$

as in the previous expression, we may group the terms in order to have a similar expression,

$$\begin{aligned}\tfrac{1}{n^6k^2}\tfrac{1}{h^2}\sum_{i,j,j'}E\left[C_1\right] &= \tfrac{1}{n^4}\Bigg(\sum_{i,l}\sigma^4\\ &\times\textstyle\sum_j\tfrac{1}{nh^2}K\left(\tfrac{X_i-Xj}{h}\right)K\left(\tfrac{Z_j-Z_i}{k}\right)/\widehat{\varphi}(X_j)\\ &\times\left(\textstyle\sum_{j'}\tfrac{1}{nk^2}K\left(\tfrac{X_i-Xj'}{h}\right)K\left(\tfrac{Z_{j'}-Z_i}{k}\right)/\widehat{\varphi}(X_{j'})\right)\cdot\nu^2(Z_i)\Bigg)\end{aligned}$$

.

this time, we get

$$\frac{1}{n^6k^2}\frac{1}{h^2}\sum_{i,j,j'}E\left[C_1\right]=O_p(1)\cdot\frac{1}{n^4}\sigma^4\sum_{i,l}\nu^2(Z_i)=O_p\left(\frac{1}{n^2}\right)$$

At the end, we finally have :

$$E\left[\widetilde{F_3}^2\right]=O_p\left(\frac{1}{n^3}+\frac{1}{n^2}\right)=O_p\left(\frac{1}{n^2}\right)$$

and then

$$E\left[\left(n\sqrt{k}.F_3\right)^2\right]=O_p\left(k\right)$$

which is the announced result for F_3 $\square_3$

We get the final result by adding the three terms F_1, F_2 and F_3 . $\square$

References

[1] Aït-Sahalia, Y. (92) : "The Delta and Bootstrap Methods for Nonlinear Functionals of Nonparametric Kernel Estimators Based on Dependent Multivariate Data", *Mimeo*, Department of Economics, MIT.

[2] Amemiya T. (80) : "Selection of Regressors", *International Economic Review*, 21(2), pp. 331-354.

[3] Atkinson A.C. (70) : "A Method for Discriminating Between Models", *Journal of the Royal Statistical Society*, series B, no 32, pp. 323-344.

[4] Bierens H.J. (87) : " Kernel Estimators of Regression Functions " in *Advances in econometrics,* Fifth World Congres, Vol.1, T.F. Bewley.

[5] Bontemps C. (95) : " Enveloppement dans les Modèles de Regression Paramétriques et Non-parametriques", *Ph.D Dissertation*, University of Social Sciences, Toulouse.

[6] Bontemps C., J.P. Florens and J.F. Richard (93) : "Encompassing in Regression Models : Parametric and Nonparametric Procedures", *Mimeo*, Gremaq, University of Social Sciences, Toulouse

[7] Bosq D. et J.F. Lecoutre (87) : " Théorie de l'estimation fonctionnelle" *Economica.*

[8] Brown B. M. (71) : "Martingale Central Limit Theorems", *Annals of Mathematical Statistics,* Vol 42, no 1, pp. 59-66.

[9] Collomb G. (76) : " Estimation Non-paramétrique de la Régression par la Méthode du Noyau", *Thèse*, Université Paul Sabatier, Toulouse.

[10] Collomb G. (77) : " Estimation Non-paramétrique de la Régression par la Méthode du Noyau : Propriété de Convergence Asymptotiquement Normale Indépendante", *Annales Scientifiques de l'Université de Clermont*, Vol. 15, pp.24-26.

[11] Cox D.R.(61) : " Tests of Separate Families of Hypotheses" in *Proceeding of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1,University of California Press, Berkeley, pp. 105-123.

[12] Cox D.R.(62) : "Further Results on Tests of Separate Families of Hypotheses", *Journal of the Royal Statistical Society*, Series B, no. 24, pp. 406-424.

[13] Dhaene G. (93) : “ Encompassing : Formulation, Properties and Testing ” Ph. D Dissertation Université Catholique de Louvain, Louvain-la-Neuve.

[14] Engle R.F. , D.F.Hendry and J.F. Richard (83) : “Exogeneity”, Econometrica, No. 55, pp 277-304.

[15] Florens J.P., D.F. Hendry and J. F. Richard (94) : “ Encompassing and Specificity ” *Mimeo*, Gremaq, Université des Sciences Sociales de Toulouse.

[16] Florens J.P., S. Larribeau et M. Mouchart (92) : “ Bayesian encompassing tests of a unit root hypothese.” *Mimeo*, Gremaq, No 92.27.

[17] Gourieroux C. et A. Monfort (91) : “Testing Non Nested Hypotheses ”, *Mimeo* No 9207, Crest-Cepremap-Insee.

[18] Gourieroux C. et A. Monfort (92) : “Testing Encompassing and Simulating Dynamic Econometric Models ”, *Mimeo* No 9214, Crest-Cepremap-Insee.

[19] Gourieroux C., A. Monfort et E. Renault (92) : “Indirect Inference ” *Mimeo*, Crest-Cepremap, Insee, Gremaq.

[20] Gourieroux C., A. Monfort et A. Trognon (83) : “ Testing Nested or Nonnested Hypotheses” *Journal of Econometrics*, Vol.21. pp 83-115.

[21] Govaerts B., D. Hendry and J.F. Richard (94) : “Encompassing in Stationary Linear Dynamic Models”, *Journals of Econometrics*, Vol. 63, pp. 245-270.

[22] Györfi L., W. Härdle, P. Sarda and P. Vieu (89) : “Nonparametric Curve Estimation from Time Series”, *Springer Verlag.*

[23] Hall P. (84) : “ Integrated Square Error Properties of Kernel Estimators of Regression Functions ”, *Annals of Statistics,* Vol. 12*,* no 1, pp.241-260.

[24] Härdle W. (90) : “ Applied Nonparametric Regression ”. *Econometric society monographs*, Cambridge University Press.

[25] Härdle W., P. Hall and J.S. Marron (92) : “Regression smoothing that are not far from their optimum”, *Journal of the Royal Statistical Society,* Series B, Vol.87, no. 417, pp. 227-233.

[26] Härdle W. and E. Mammen (93) : "Comparing nonparametric versus parametric regression fits", *Annals of Statistics,* Vol. 21, no 4, pp. 1926-1947.

[27] Härdle W. and J.S. Marron (89) : "Optimal bandwidth selection in nonparametric procedure regression function estimation ", *Annals of statistics*, Vol.13, No.4.

[28] Hausman J.A. (78) : " Specification test in Econometrics", *Econometrica*, vol 46, no 6.

[29] Hendry D. (93) : " The Roles of Economic Theory and Econometrics in Time Series Economics" , Mimeo, Nuffield College, Oxford.

[30] Hendry D. and J.F. Richard (89) : " Recent development in the theory of encompassing " in *Contribution to operation research and economics*, MIT Press.

[31] Lu M. and G. E. Mizon (93) : " The Encompassing Principle and Specification Test", *Mimeo,* Economics Departement, Southampton University, UK.

[32] Marron J. S. (88) : "Automatic Smoothing Parameter Selection : A Survey", Empirical Economics, Vol. 13, pp. 187-208.

[33] Mizon G. E. (84): " The Encompassing Approach in Econometrics " In *Econometrics and Quantitative Economics,* edited by D. F. Hendry and K. F. Wallis. Ch. 6 Oxford : Basil & Blackwell

[34] Mizon G. E. and J. F. Richard (86) : " The encompassing principle and its application to testing non-nested hypotheses " *Econometrica*, Vol.54, No 3.

[35] Nadaraya E.A. (64) : "On estimating regression", *Theory of Probability and its Applications*, 9, pp. 141-142.

[36] Pesaran M.H. (74) : "On the General Problem of Model Selection", *Review of Economic Studies*, no 41, pp. 153-171.

[37] Sawa T. (78) : "Information Criteria for Discriminating Among Alternative Regression Models", *Econometrica,* no. 46, pp. 1273-92.

[38] Serfling (80) : "Approximation theorems ", *Wiley series in probability and mathematical statistics,* Wiley & Sons.

[39] Stone, C. J. (82) : "Optimal global rates of convergence for nonparametric regression", *Annals of Statistics,* Vol 10, no 4, pp. 1040-1053.

[40] Vieu P. (93) : "Bandwidth selection for kernel regression", *Computational Statistics and Data Analysis*, à paraître.

[41] Watson, G.S. (64) : "Smooth regression analysis", *Sankhya, Series A*, 26, pp. 359-372.

[42] White H. (80) : "Using least squares to approximate unknown regression function " *International Economic Review*, Vol.21(1).

[43] White H. (82) : "Maximum likelihood estimator of misspecified models " *Econometrica*, Vol.50, pp. 1–26.

[44] White H. (84) : "Asymptotic Theory for Econometricians", *A Series of Monographs and Textbooks*, Academic Press.