ISSN 0871-2573

A SMOOTHING-SPLINES-LIKE ESTIMATOR FOR NONPARAMETRIC REGRESSION WITH A LINEAR GENERALIZED LEAST SQUARES INTERPRETATION

por Maximiano Reis Pinheiro
Universidade Nova de Lisboa
and
Instituto Nacional de Estatistica

Working Paper nº 170

Faculdade de Economia
Universidade Nova de Lisboa
Travessa Estêvão Pinto
Campolide
1000 Lisboa

Maio de 1991

A SMOOTHING-SPLINES-LIKE ESTIMATOR FOR NONPARAMETRIC REGRESSION WITH A LINEAR GENERALIZED LEAST SQUARES INTERPRETATION[1]

by

Maximiano Reis PINHEIRO

Universidade Nova de Lisboa

and

Instituto Nacional de Estatistica

---

[1]This article has grown, in part, from my doctoral dissertation submitted to the University of Geneva, Switzerland. Research support was partially provided by the Gulbenkian Foundation, Portugal. I am greatly indebted to Prof. Pietro Balestra for his valuable guidance. At different stages of my work I have also benefited from helpful suggestions by E. Ronchetti, A. Holly, H. White, J. Krishnakumar and two anonymous referees. Prof. G. Wahba kindly suggested to me how to access certain software I needed for simulations. The usual disclaimer applies.

ABSTRACT

In this paper a new estimator for nonparametric regression is suggested. It is a smoothing-splines-like estimator obtained by minimizing a two term criterion function. The first term of this criterion function is a sum of squares of residuals and the second term is a penalty function defined as a weighted sum of squared errors of local first order Taylor approximations. Under usual regularity conditions, almost sure uniform consistency is proved for both the function and its first derivatives.

Instead of solving the optimization problem to explicitly obtain an element of the function space considered, a procedure is suggested to only estimate the values of the function and the first derivatives at the observation points. This is especially convenient in order to simplify the calculations and to keep the simplicity of the economic interpretation of the usual linear econometric specifications. This practical procedure also allows another interesting interpretation of the estimators as the generalized least squares estimators of a two regime expanded linear regression with a singular two components error covariance structure. The link with conventional linear regression theory is developed by showing that the estimators are unbiased if the true function is linear.

Some Monte Carlo experiments are reported and the relative performance of the estimators is discussed. Finally, an application is presented, using aggregate data from OECD on primary energy requirements per unit of GDP and GDP per capita.

1.INTRODUCTION

Let $\{(x_1,y_1),\ldots,(x_n,y_n),\ldots,(x_N,y_N)\}$ be a sample of N observations of a $\mathbb{R}^k \times \mathbb{R}$ - valued vector of variables and assume the model

$$y_n = f^0(x_n) + \varepsilon_n \quad (n = 1,\ldots,N) \tag{1.1}$$

where $f^0(\cdot)$ is an unknown and unspecified (in general nonlinear) function, x is a vector of random or fixed designed regressors and the ε_n $(n=1,\ldots,N)$ are zero mean residuals uncorrelated with the regressors.

There are several nonparametric regression estimators for model (1.1) which reduce the a *priori* restrictions on the specification of $f^0(\cdot)$ to continuity and smoothness constraints. The best known types of nonparametric regression estimators are smoothing splines, k-nearest neighbor (k-NN) estimators and kernel estimators, with several subtypes. For two recent good expositions of the subject, see EUBANK (1988) and HÄRDLE (1990).

The basic univariate smoothing splines estimator is defined as the solution of the following problem (eg., WAHBA (1975)): In the Sobolev space $W_{m,2}$ of real-valued functions on [a,b] with absolutely continuous (m-1)-st derivative and square integrable m-th derivative, find the function $f(\cdot)$ that minimizes

$$N^{-1} \sum_{n=1}^{N} [y_n - f(x_n)]^2 + h \int_a^b [f^{(m)}(z)]^2 dz \tag{1.2}$$

where h is a positive bandwidth to be chosen by some risk minimization criterion (for example, by generalized cross-validation, e.g. CRAVEN and WAHBA (1979)). The first term of the criterion function (1.2) measures the infidelity to the data, the second term is a penalty function that measures the "roughness" of the solution (the bandwidth h controls the trade-off between these two terms).

In this article we suggest a new estimator for nonparametric regression. It is a smoothing-splines-like estimator also obtained by

minimizing in f a two term criterion function:

$$Q_{N,h}(f) = N^{-1}\sum_{n=1}^{N}[y_n - f(x_n)]^2 + N^{-1}\sum_{s=1}^{N}N_s^{-1}\sum_{n=1}^{N}w_{sn}(r_{sn}/h)^2 \qquad (1.3)$$

where:

1) $r_{sn} = f(x_n) - f(x_s) - \Delta x'_{sn}\frac{\partial f}{\partial x}(x_s)$ is a first order error of a local Taylor approximation;

2) $\Delta x_{sn} = x_n - x_s$;

3) $w_{sn} = w(\|\Delta x_{sn}\|/h)$, where $\|\Delta x_{sn}\|$ is the euclidean norm of Δx_{sn} and $w(\cdot)$ is a non-negative and symmetric weight function with support $[-1;1]$ and such that $\int_{-1}^{1}w(z)dz = 1$;

4) For a fixed s, N_s is the number of strictly positive weights w_{sn};

5) h is a bandwidth ($h > 0$).

As in the smoothing splines problem, the first term of (1.3) is the sum of squares of the residuals. The second term is a penalty function that can be viewed as a discreticized version of the second term of (1.2) with m=2. We will see below that this discrete version of the penalty function is especially convenient for practical calculation and allows interesting relationships with the conventional parametric linear regression[2].

The proof of the consistency of the suggested estimators is based on a result of GALLANT (1985) and GALLANT and NYCHKA (1987), which is a development of the method of "sieves" proposed by GRENANDER (1981). When we

---

[2]Let us notice that, in a way which is similar to the minimization of (1.2), the minimization of (1.3) under the constraint of linearity is equivalent to the minimization of only the first term under the same constraint. In fact, if we impose linearity, $r_{sn} = 0$ for all s and n (and $f^{(2)}(x) = 0$ for all x). In other words, minimizing (1.3) or (1.2) under the constraint of linearity is the same as applying ordinary least squares to the linear specification $y_t = [1\ x'_t]\alpha + u_t$.

need to prove the consistency of an estimator defined as the solution that minimizes a criterion function over an infinite dimensional space, techniques for proving consistency in finite dimensional parameter spaces typically fail. GRENANDER suggested performing the optimization of the criterion function within a finite dimension subset of the function space, and then allowing the subset to "grow" with the sample size in order to progressively "fill" the function space. The sequence of subsets of the parameter space is called a "sieve". In section 2 we prove consistency of the estimators for any sieve that satisfies some general conditions.

In sections 3 and 4 we formulate a practical procedure for minimizing the criterion function (1.3) in the finite sample case when we are only interested in estimating the values of the function and its first derivatives at the observation points. This is especially convenient: (i) to keep a simple and easy economic interpretation of the results[3]; (ii) to explore the links between conventional linear regression and our estimator. In fact, in section 3 we show that the suggested estimators may be interpreted as generalized least squares estimators of a two regime expanded linear regression with singular two components error covariance structure. In section 4 we obtain the explicit expression of the estimators, noting that they can be evaluated in two steps. In a first step, estimates of the values of the function are obtained by smoothing the observations of the explained variable. In the second step, estimates of the values of the first derivative are obtained by local weighted linear regressions with the

---

[3]Remark, for example, that if all the variables are expressed in natural logarithms, the estimates of the derivatives at the observation points will be local elasticities (without the need of any transformation or calculation).

estimates of the first step as explained variable. In section 4, we also show, by the exploration of the properties of the smoothing matrix of the first step, that the suggested estimators are unbiased if the true function is linear.

In section 5, we analyze the case of repeated observations for the regressors, which was ruled out of sections 3 and 4 to simplify the exposition.

In section 6, the results of some Monte Carlo experiments are reported and the relative performance of our estimators to other nonparametric regression estimators is discussed. Finally, in section 7, we present an application of the proposed estimators to analyze the existence of a relationship between GDP per capita and primary energy requirements per unity of GDP. For this purpose we use aggregate data for the 24 members of OECD.

2. CONSISTENCY

In the beginning of the paper we assumed that the data $\{(x_n,y_n)\}_{n=1}^N$ is generated according to model (1.1):

$$y_n = f^0(x_n) + \varepsilon_n \quad (n = 1,\ldots,N)$$

Let us also assume, from this point forward, that:

Assumption (A1): The empirical distribution function μ_N of $\{x_n\}_{n=1}^N$, defined as $\mu_N(x) = N^{-1}\times$(number of $x_n \leq x$, all coordinates), converges to a probability distribution $\mu(x)$ at every continuity point of this last function $\mu(x)$. We shall denote by $\mathcal{X}$ the support of $\mu(x)$.

Assumption (A2): $\mathcal{X}$ is the closure of a bounded and convex subset of $\mathbb{R}^K$.

Assumption (A3): For some bound B, $0 < B < \infty$, the "true function" f^0 is a point in the Banach space $C^2(\mathcal{X})$ with norm

$$\|f^0\|_{II} = \max_{|\lambda|\leq 2} \sup_{x\in\mathcal{X}} |D^\lambda f^0(x)| < B$$

where $\lambda = (\lambda_1,\lambda_2,\ldots,\lambda_K)$, λ_i $(i=1,\ldots,K)$ are non-negative integers, $|\lambda| = \sum_{i=1}^{K}\lambda_i$ and

$$D^{\lambda}f^{0}(x) = \frac{\partial^{|\lambda|}}{\partial x_1^{\lambda_1}\partial x_2^{\lambda_2}\ldots\partial x_K^{\lambda_K}} f^{0}(x)$$

The vector space $C^2(\bar{\mathcal{X}})$ consists of all those functions $f \in C^2(\mathcal{X})$ [4] for which $D^{\lambda}f$ is bounded and uniformly continuous on $\mathcal{X}$ for $|\lambda| \leq 2$.

Assumption (A4): The residuals and the regressors are independent.

Assumption (A5): The residuals $\{\varepsilon_n\}_{n=1}^{N}$ can be factorized in the following way:

$$\varepsilon_n = l(x_n)e_n \quad (n=1,\ldots,N)$$

where:

(i) $l(\cdot)$ is a (possibly unspecified) uniformly bounded function ($|l(x)| < L < \infty$ for all $x \in \mathcal{X}$) and square integrable in $\mathcal{X}$:

$$\int_{\mathcal{X}} l^2(x)d\mu(x) = \theta^2 < \infty$$

(ii) the $\{e_n\}$ are identical and independently distributed variables with common distribution function P(e) having zero mean and finite variance σ^2.

This last assumption reduces ε_n to i.i.d.$(0,\sigma^2)$ residuals when $l(x) \equiv 1$ for all x. But it also allows some heteroskedastic situations, when it is presumed that the variance of the residual ε_n depends on x_n. This departure from the i.i.d. case has practical interest and it is costless in terms of the Uniform Strong Law of Large Numbers involved in the proof of the

---

[4] As usually considered, $C^2(\mathcal{X})$ is the space consisting of all functions f which, together with all their partial derivatives $D^{\lambda}f$ of orders $|\lambda| \leq 2$, are continuous on $\mathcal{X}$.

consistency of the estimators[5].

To establish the uniform strong consistency of the estimators obtained from (1.3) under (A1) to (A5), we shall relate our formulation to the assumptions of the following trivial adaptation of a result due to GALLANT (1985, sec.6) (see also GALLANT and NYCHKA (1987) and GALLANT and WHITE (1989)):

Theorem 2.0: Suppose that $\hat{f}_N$ is obtained by minimizing a sample objective function $Q_{N,h}(f)$ over $\mathcal{F}_N$, where $\mathcal{F}_N$ is a subset of a function space $\mathcal{F}$ on which is defined a norm $\|f\|$. Consider the following conditions:

(a) Compactness: The closure $\overline{\mathcal{F}}$ of $\mathcal{F}$ With respect to $\|f\|$ is compact in the relative topology generated by $\|f\|$.

(b) Uniform convergence: There exist a function $f^0 \in \mathcal{F}$ and a function $\bar{Q}(f,f^0)$ continuous in f with respect to $\|f\|$ such that

$$\lim_{N\to\infty} \sup_{f\in\overline{\mathcal{F}}} |Q_{N,h_N}(f) - \bar{Q}(f,f^0)| = 0 \text{ almost surely}$$

provided that $h_N \to 0$ almost surely.

(c) Identification: $\|f^* - f^0\| = 0$ for any element $f^* \in \mathcal{F}$ such that

$$\bar{Q}(f^*,f^0) \le \bar{Q}(f^0,f^0)$$

(d) Denseness: $\mathcal{F}_N \subset \mathcal{F}_{N+1}$ for all N and $\bigcup_{N=1}^{\infty} \mathcal{F}_N$ is a dense subset of $\overline{\mathcal{F}}$ with respect to $\|f\|$.

If conditions (a)–(d) hold, then

$$\lim_{N\to\infty} \|\hat{f}_N - f^0\| = 0 \text{ almost surely}$$

provided that $h_N \to 0$ almost surely.

---

[5]It would not be the case if it was supposed $\{(x_n,\varepsilon_n)\}$ to follow some general kind of heterogeneous mixing process (for example, GALLANT and WHITE (1987)). This is virtually possible but it would considerably complicate the analysis.

Our strategy will be to verify, one by one, the different conditions of the previous theorem. In our case, according to Assumption (A3), we have:

$$\mathcal{F} = \{f \in C^2(\bar{\mathcal{X}}) \mid \|f\|_{II} < B\} \tag{2.1}$$

Note that by construction $\mathcal{F}$ is bounded in $C^2(\bar{\mathcal{X}})$.

Lemma 2.1: Let $\mathcal{F}$ be defined as in (2.1) and assume (A.1)-(A.2)-(A.3). The closure $\bar{\mathcal{F}}$ of $\mathcal{F}$ with respect to the norm

$$\|f\|_I = \max_{|\lambda| \leq 1} \sup_{x \in \mathcal{X}} |D^\lambda f(x)|$$

is compact in the relative topology on $\bar{\mathcal{F}}$ generated by the norm $\|\cdot\|_I$.

Lemma 2.2: Let $Q_{N,h}(f)$ be defined as in (1.3), and assume (A1)-(A3)-(A4)--(A5). Then, provided that $h_N \to 0$ almost surely,

$$\lim_{N\to\infty} \sup_{f \in \mathcal{F}} |Q_{N,h_N}(f) - \bar{Q}(f,f^0)| = 0 \text{ almost surely}$$

for:

$$\bar{Q}(f,f^0) = \theta^2\sigma^2 + \int_{\mathcal{X}} [f(x) - f^0(x)]^2 d\mu(x) \tag{2.2}$$

Lemmas 2.1 and 2.2 guarantee respectively the compactness condition (a) and the uniform convergence condition (b) of theorem 2.0. Note that the identification condition (c), with $\|f\| = \|f\|_I$, is also satisfied by (2.2) since the elements of $\mathcal{F}$ are continuous functions. The proofs of both lemmas follow in part the arguments developed by GALLANT and WHITE (1989), and are presented in Appendices A and B.

In conclusion, for any sieve $\{\mathcal{F}_N\}$ satisfying condition (d) of theorem 2.0 (with $\|f\| = \|f\|_I$), we will have

$$\lim_{\substack{N\to\infty \\ h_N \to 0 \text{ a.s.}}} \|\hat{f}_N - f^0\|_I = 0 \text{ almost surely} \tag{2.3}$$

or, equivalently,

$$\lim_{\substack{N\to\infty \\ h \to 0 \text{ a.s.}}} \max_{|\lambda| \leq 1} \sup_{x \in \mathcal{X}} |D^\lambda[\hat{f}_N(x) - f^0(x)]| = 0 \text{ almost surely} \tag{2.4}$$

Some possible sieves may be found in the recent literature on neural networks and related topics. For example, HORNIK, STINCHCOMBE and WHITE

(1990) set mild conditions on the "activation function" G such that

$$\Sigma_\infty(G) \equiv U_{p=1}^{\infty} \Sigma_p(G)$$

with

$$\Sigma_p(G) = \{g : \mathcal{X} \to \mathbb{R} \mid g(x) = \sum_{j=1}^{p} \beta_j G(\tilde{x}'\gamma_j),\ \tilde{x} = [1\ x']',\ \beta_j \in \mathbb{R},\ \gamma_j \in \mathbb{R}^{K+1}\}$$

is dense for some relevant function spaces with respect to some uniform norms. However, for what follows it is preferable to consider a sieve of modified polynomials of ascending order, because they are linear in coefficients.

Let us define

$$P_p(x) = \sum_{|\lambda| \le p} \beta_\lambda x_1^{\lambda_1} x_2^{\lambda_2} \quad x_K^{\lambda_K} \tag{2.5}$$

where $\lambda = (\lambda_1,\ldots,\lambda_K)$ and $|\lambda| = \sum_{l=1}^{K} \lambda_l$, with λ_l non-negative integers such that $\lambda_l \ge \lambda_{l+1}$. Lemma 2.3 is based on a result of GALLANT and NYCHKA (1987, lemma A.5) and guarantees the verification of condition (d) of theorem 2.0.

Lemma 2.3: Let $\mathcal{P}_\infty(\mathcal{X}) = U_{p=0}^{\infty} \mathcal{P}_p(\mathcal{X})$, with

$$\mathcal{P}_p(\mathcal{X}) = \{f:\mathcal{X} \to \mathbb{R} \mid f(x) = P_p(x)\phi(x)\}$$

where $P_p(x)$ is the polynomial (2.5) and ϕ is a strictly positive probability density function that has a moment generating function (like the K-variate standard normal density). $\mathcal{P}_\infty(\mathcal{X}) \cap \mathcal{F}$ is dense in $\mathcal{F}$ relative to the "consistency norm" $\|.\|_I$. (Proof: Appendix C).

3 INTERPRETATION AS A GLS-LIKE ESTIMATOR

The results of the previous section guarantee the consistency of the estimator obtained by minimizing (1.3) over $\mathcal{P}_p(\mathcal{X})$ when $p_N \to \infty$ and $h_N \to 0$ almost surely. In the finite sample case with N observations, for each value p_N we will obtain an estimator $\hat{f}(x)$ belonging to $\mathcal{P}_{p_N}(\mathcal{X})$. It will be natural to choose p_N such that there are no implicit *a priori* constraints on the values of the estimated function and its first derivatives at the

observation points. If we do so, it will be equivalent, for the purpose of estimation of these values at the observation points, instead of minimizing (1.3) over $\mathcal{P}_{P_N}(\mathcal{X})$, to minimize the unrestricted criterion function

$$Q_{N,h}(f,g) = N^{-1}\sum_{n=1}^{N}(y_n - f_n)^2 + N^{-1}\sum_{s=1}^{N}N_s^{-1}\sum_{n=1}^{N}w_{sn}(r_{sn}/h)^2 \qquad (3.1)$$

(with $r_{sn} = f_n - f_s - \Delta x'_{st}g_s$) in order to the two vectors:

$$f = [f_1 \ldots f_n \ldots f_N]' \equiv [f(x_1)\ldots f(x_n)\ldots f(x_N)]' \quad (Nx1)$$

$$g = [g'_1 \ldots g'_n \ldots g'_N]' \equiv [\frac{\partial}{\partial x'}f(x_1)\ldots\frac{\partial}{\partial x'}f(x_n)\ldots\frac{\partial}{\partial x'}f(x_N)]' \quad (KNx1)$$

Obviously, the unrestricted minimization of (1.3) does not satisfy automatically the bound B on the norm $\|\hat{f}\|_{II}$ of the corresponding polynomial in $\mathcal{P}_{P_N}(\mathcal{X})$ (as imposed by Assumption (A3)). However, our experience indicates that if the bound is large, there will be no need for its explicit consideration and in practice (3.1) may be minimized without any restriction.

As we noticed in the introduction, the estimation of the two vectors of values of the function and its first derivatives at the observation points is especially convenient to explore the links between our formulation and the conventional linear regression theory. Let:

1) $y = [y_1 \ldots y_n \ldots y_N]'$ (Nx1) and $X = [x_1 \ldots x_n \ldots x_N]'$ (NxK)

2) $\varepsilon = [\varepsilon_1 \ldots \varepsilon_n \ldots \varepsilon_N]'$ (Nx1)

3) $W_s = \mathrm{diag}_n(w_{sn})$ (NxN) and $W = \mathrm{diag}_s(W_s)$ (N^2xN^2);

4) $W_s^* = h^{-2}N_s^{-1}W_s$ (NxN) and $W^* = \mathrm{diag}_s(W_s^*)$ (N^2xN^2);

5) $r_s = [r_{s1} \ldots r_{sn} \ldots r_{sN}]'$ (Nx1) and $r = [r'_1 \ldots r'_n \ldots r'_N]'$ (N^2x1);

6) $D_s = I_N - \iota b'_s$, with $\iota = [1 \ldots 1]'$ (Nx1) and b_s the s-th column of the identity matrix I_N.

Note that when we pre-multiply the idempotent (non-symmetric) matrix D_s by X or f we obtain, respectively,

$$D_s X = [\Delta x_{s1} \ldots \Delta x_{sn} \ldots \Delta x_{sN}]'$$

and

$$D_s f = [(f_1 - f_s) \ldots 0 \ldots (f_N - f_s)]'$$

Then:

$$r_s = D_s f - D_s X g_s \quad (s=1,\ldots N) \quad \text{and} \quad r = Df - \mathbb{D}(I_N \otimes X)g$$

where $D = [D_1' \ldots D_s' \ldots D_N']'$ $(N^2 \times N)$ and $\mathbb{D} = \text{diag}_s(D_s)$ $(N^2 \times N^2)$. Using this matrix notation, we can write criterion (1.3) as follows:

$$Q_{N,h}(f,g) = N^{-1}(y - f)'(y - f) + N^{-1} r' W^* r \tag{3.2}$$

Equivalently (Appendices E and F):

$$Q_{N,h}(f,g) = N^{-1}[J(y - f) + \bar{I}r]'(\bar{W}^+ + JJ')^-[J(y - f) + \bar{I}r] \tag{3.3a}$$

or

$$Q_{N,h}(f,g) = N^{-1}(Jy - Z\beta)'\Sigma^-(Jy - Z\beta) \tag{3.3b}$$

where:

1) $J = \iota \otimes I_N$ $(N^2 \times \bar{T})$, $Z = [(J - \bar{I}D) \quad \bar{I}\mathbb{D}(I \otimes X)]$ $(N^2 \times N(K+1))$, $\beta = [f' \; g']'$ $(N(K+1) \times 1)$ and $\Sigma = \bar{W}^+ + JJ'$;

2) $\bar{W} = \text{diag}_s(\bar{W}_s)$, with $\bar{W}_s$ identical to W_s^* except that its element (s,s) is null;

3) $\bar{W}^+$ is the Moore–Penrose generalized inverse of $\bar{W}$: $\bar{W}^+ = \text{diag}_s(\bar{W}_s^+)$, with $\bar{W}_s^+ = h^2 N_s \, \text{diag}(w_{s1}^+, \ldots, w_{s,s-1}^+, 0, w_{s,s+1}^+, \ldots, w_{sN}^+)$ and, for $n \neq s$, $w_{sn}^+ = 1/w_{sn}$ if $w_{sn} > 0$, $w_{sn}^+ = 0$ otherwise;

4) $\bar{I} = \bar{W}^+\bar{W} = \text{diag}_s(\bar{I}_s)$ $(N^2 \times N^2)$ with $\bar{I}_s = \text{diag}_t(i_{sn})$ and $i_{sn} = 1$ if $w_{sn}^+ > 0$, $i_{sn} = 0$ otherwise.

In Appendix E we show that (3.3a) and (3.3b) are numerically invariant with the choice of the g-inverse.

Criterion form (3.3b) allows an interesting econometric interpretation. In fact, we have:

$$Jy - Z\beta = Jy - (J - \bar{I}D)f - \bar{I}\mathbb{D}(I \otimes X)g = [y - (I - \bar{I}_s D_s)f - \bar{I}_s D_s X g_s]_{s=1,\ldots,N}$$

For a fixed s, the n-th element of $[y - (I - \bar{I}D_s)f - \bar{I}_s D_s X g_s]$ is:

$$\begin{cases} \bar{y}_n - f_s - \Delta x'_{sn} g_s & \text{if } n \neq s \text{ and } w_{sn} = w(\|\Delta x_{sn}\|/h) > 0 \\ y_n - f_n & \text{otherwise} \end{cases}$$

In Appendix G, we show that for every vector v such that $v'\Sigma v = 0$, we have $Z'v = 0$. In consequence, the estimators that we suggest are the generalized least squares estimators (for a fixed bandwidth h) of the two regime expanded linear regression

$$Jy = Z\beta + u \Leftrightarrow y_n = \begin{cases} f_s + \Delta x'_{sn} g_s + u_{sn} & \text{if } n \neq s \text{ and } w_{sn} > 0 \\ f_n + u_{sn} & \text{otherwise} \end{cases} \qquad (s,n=1,\ldots,N)$$

with the singular error covariance matrix (defined up to a multiplicative scalar) $\Sigma = \bar{W}^+ + JJ'$. This covariance matrix corresponds to a two uncorrelated components error structure

$$u_{sn} = u^{(1)}_{sn} + u^{(2)}_{n}$$

with $E(u^{(1)}_{sn}) = E(u^{(2)}_{n}) = 0$ $\forall s,n$ and (up to a multiplicative scalar)

$$V(u^{(1)}) = \bar{W}^+ , \quad V(u^{(2)}) = JJ' (= V(J\varepsilon))$$

The non-zero diagonal elements of $\bar{W}^+$ may be interpreted as pseudo-variances of the Taylor (unknown but non-stochastic) residuals r_{sn}.

4. THE EXPRESSION FOR THE ESTIMATORS OF THE VALUES OF THE FUNCTION AND ITS FIRST DERIVATIVES

The first order conditions of the problem of minimizing (3.2) with respect to vectors f and g are the following:

$$\frac{\delta Q}{\delta f} = -2N^{-1}\{(\bar{y} - f) - D'W^*[Df - \mathbb{D}(I\otimes X)g]\} = 0 \qquad (4.1a)$$

$$\frac{\delta Q}{\delta g} = -2N^{-1}(I\otimes X')\mathbb{D}'W^*[Df - \mathbb{D}(I\otimes X)g] = 0 \qquad (4.1b)$$

Note that criterion (3.2) is quadratic in f and g, with hessian matrix:

$$H = \frac{\delta^2 Q}{\delta\beta\delta\beta'} = 2N^{-1}\begin{bmatrix} (I^{-1} + D'W^*D) & -D'W^*\mathbb{D}(I\otimes X) \\ -(I\otimes X)\mathbb{D}'W^*D & (I\otimes X')\mathbb{D}'W^*\mathbb{D}(I\otimes X) \end{bmatrix} =$$

$$= 2N^{-1}\left\{\begin{bmatrix} I^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} D' \\ -(I\otimes X')\mathbb{D}' \end{bmatrix} W^*[D \; -\mathbb{D}(I\otimes X)]\right\}$$

which is obviously non-negative definite. Consequently, the solutions of the first order conditions system (4.1) are the global optimal solutions. Assume that

$$(I\otimes X')\mathbb{D}'W^{*}\mathbb{D}(I\otimes X) = \mathrm{diag}_{s}(h^{-2}N_{s}^{-1}(D_{s}X)'W_{s}(D_{s}X)) \qquad (4.2)$$

is non-singular. In particular, it will be necessary that $N_{s} \geq K$ and that all matrices $(D_{s}X)$ $(s=1,\ldots,N)$ have maximal rank K. The solution of (4.1) is unique and given by:

$$\hat{f} = Ay \qquad (4.3)$$

$$\hat{g}_{s} = [(D_{s}X)'W_{s}(D_{s}X)]^{-1}(D_{s}X)'W_{s}D_{s}\hat{f} \quad (s=1,\ldots,N) \qquad (4.4)$$

where

$$A = \{I_{N} + D'[W^{*} - W^{*}\mathbb{D}(I\otimes X)[(I\otimes X')\mathbb{D}'W^{*}\mathbb{D}(I\otimes X)]^{-1}(I\otimes X')\mathbb{D}'W^{*}]D\}^{-1} =$$

$$= \{I_{N} + h^{-2}\sum_{s=1}^{N}N_{s}^{-1}D_{s}'[W_{s} - W_{s}(D_{s}X)'[(D_{s}X)'W_{s}(D_{s}W)]^{-1}(D_{s}X)'W_{s}]D_{s}\}^{-1} \qquad (4.5)$$

Before analyzing the properties of this smoothing matrix A, we may note an interesting feature of the estimators (4.3)–(4.4). Vectors $\hat{g}_{s}$ $(s=1,\ldots,N)$ of the estimated reaction coefficients are calculated only after having first evaluated $\hat{f}$. That is, the reaction coefficients at the different observation points are evaluated by local weighted linear regressions, but only after having smoothed y_{n}. Compared with the usual weighted local regression (e.g., MÜLLER (1987)), which uses the observations of the explained variable directly, estimators (4.3)–(4.4) present an important logical advantage.

Let $h^{-2}C^{*} = W^{*} - W^{*}\mathbb{D}(I\otimes X)[(I\otimes X')\mathbb{D}'W^{*}\mathbb{D}(I\otimes X)]^{-1}(I\otimes X')\mathbb{D}'W^{*} = h^{-2}\mathrm{diag}_{s}(C_{s}^{*})$, with $C_{s}^{*} = N_{s}^{-1}\{W_{s} - W_{s}(D_{s}X)[(D_{s}X)'W_{s}(D_{s}X)]^{-1}(D_{s}X)'W_{s}\}$, and let $B = D'C^{*}D =$ $= \sum_{s=1}^{N}D_{s}'C_{s}^{*}D_{s}$. We may write: $A^{-1} = I_{N} + h^{-2}B$. Continue to assume that (4.2) is non-singular. The following propositions may be easily established (Appendix H):

1) $0 < \lambda_{i} \leq 1$ $(i=1,\ldots,N)$, where the λ_{i} $(i=1,..,N)$ are the eigenvalues of A.

2) Assume that matrix $R = [\iota_N \; X]$ has maximal rank (K+1). Multiplicity of unity eigenvalue of A is at least (K+1). In fact, the (K+1) columns of R are independent eigenvectors of A associated to eigenvalue 1.

From the previous propositions we obtain directly that (see Appendix I): $\sum_{n=1}^{N} a_{sn} = 1$, $\sum_{n=1}^{N} a_{sn} \Delta x_{sn} = 0$ and $0 < a_{ss} \leq 1$ (s=1,...,N). Figure 1 presents an example of the weights a_{sn} induced by the minimization of criterion (3.2) (example developed with a quadratic weight function, K = 1, N = 30, $1/6 \leq x_n \leq 5$, $x_n - x_{n-1} = 1/N \; \forall n$ and h = 0.4). Note that some weights are negative. By (4.3),

$$\hat{f}_s = \sum_{n=1}^{N} a_{sn} y_n \quad \text{and} \quad E(\hat{f}_s | h) = \sum_{t=1}^{N} a_{sn} f_n^0 \quad (s=1,\ldots,N)$$

and $\hat{f}_s$ and $E(\hat{f}_s | h)$ are weighted means (with some negative weights) respectively of the observed y_n and of the true f_n^0. In general, $\hat{f}$ will be biased, as will $\hat{g}_s$ (s=1,...,N). However, if the true function is linear, that is, if $f^0 = \iota\alpha_0 + X\alpha_1 = R\alpha$, we shall have

$$E(\hat{f}) = E_h[E(\hat{f} | h)] = E_h(AR\alpha) = E_h(R\alpha) = R\alpha = f^0$$

and

$$E(\hat{g}_s) = E_h\left[E\left([(D_s X)' W_s (D_s X)]^{-1}(D_s X)' W_s D_s \hat{f} | h\right)\right] =$$

$$= E_h\left[[(D_s X)' W_s (D_s X)]^{-1}(D_s X)' W_s (D_s' X)\alpha_1\right] = \alpha_1 = g_s^0 \quad (s=1,\ldots,N)$$

5. THE CASE OF REPEATED OBSERVATIONS FOR THE REGRESSORS

In sections 3 and 4 we considered the case of non-repeated observations for the regressors, i.e.

$$x_n \neq x_{n'} \text{ for } n \neq n'$$

When there are repeated observations we cannot directly extend the previous results. They do not take account of the restrictions imposed by repetitions. In fact,

$$x_n = x_{n'} \Rightarrow f_n = f_{n'} \text{ and } g_n = g_{n'} \tag{5.1}$$

and these natural constraints will not be satisfied in general by (4.3) and (4.4). One possibility will be to explicitly minimize the criterion function $Q_{N,h}$ under these constraints[6]. Rather than doing this, another (more convenient and equivalent) possibility will be to substitute the constraints into the criterion function and to work with a concentrated unconstrained criterion.

Let $y_{(n)}$ represent, not the n-th scalar observation for the endogenous variable, but the $\bar{N}_n \times 1$-vector

$$y_{(n)} = [y_{(n)1}\ y_{(n)2}\ \dots\ y_{(n)\bar{N}_n}]'$$

consisting of those observations of the endogenous variable which correspond to the same x_n. Let $\bar{N}$ be the number of different observations of the explained variable corresponding to the same x_n and let $N = = \sum_{n=1}^{\bar{N}} \bar{N}_n$ be the total number of observations (as before). We can introduce the constraints in (3.1) and then manipulate both terms to achieve a convenient form for this criterion function. Denoting $\bar{y}_n = \bar{N}_n^{-1} \sum_{l=1}^{\bar{N}n} y_{(n)l}$, the two terms of (3.1) become:

$$N^{-1} \sum_{n=1}^{\bar{N}} \sum_{l=1}^{\bar{N}_2} (y_{(n)l} - f_n)^2 = N^{-1} \sum_{n=1}^{\bar{N}} \sum_{l=1}^{\bar{N}_n} [y_{(n)l} - \bar{y}_n) + (\bar{y}_n - f_n)]^2 =$$

$$= N^{-1} \sum_{n=1}^{\bar{N}} \bar{N}_n (\bar{y}_n - f_n)^2 + N^{-1} \sum_{n=1}^{\bar{N}} \sum_{l=1}^{\bar{N}} (y_{(n)l} - \bar{y}_n)^2$$

(5.2a)

and

---

[6]Note that this procedure will not change the proof of consistency of the estimators, due to the fact that the true function (and its derivatives) will necessarily satisfy these constraints:

$$x_n = x_{n'} \rightarrow f^0(x_n) = f^0(x_{n'}) \text{ and } \frac{\partial}{\partial x} f^0(x_n) = \frac{\partial}{\partial x} f^0(x_{n'})$$

$$N^{-1} \sum_{s=1}^{\bar{N}} \sum_{i=1}^{\bar{N}} N_s^{-1} \sum_{n=1}^{\bar{N}} \sum_{j=1}^{\bar{N}_n} w_{sn}(r_{sn}/h)^2 =$$

$$= N^{-1} \sum_{s=1}^{\bar{N}} \bar{N}_s N_s^{-1} \sum_{n=1}^{\bar{N}} \bar{N}_n w_{sn} (r_{sn}/h)^2 \qquad (5.2b)$$

As the second term of (5.2a) does not depend on f and g, the minimization of (3.1) under the constraints (5.1) will be equivalent to the minimization of the unconstrained concentrated criterion

$$Q^c_{N,h}(f,g) = N^{-1} \sum_{n=1}^{\bar{N}} \bar{N}_n (\bar{y}_n - f_n)^2 + N^{-1} \sum_{s=1}^{\bar{N}} \bar{N}_n w_{sn}(r_{sn}/h)^2 \qquad (3.1')$$

Using (3.1') instead of (3.1), the analysis of sections 3 and 4 may be easily transposed for the case of repeated observations for the regressors. In Appendix D we present a list of modified expressions which replace the expressions of sections 3 and 4 in the case of repeated observations[7].

6. A COMPARISON WITH OTHER NONPARAMETRIC ESTIMATORS

We developed some Monte Carlo experiments to compare estimators (4.3)--(4.4) with the other estimators suggested in the literature. To simplify, we considered only one explanatory variable, uniformly spaced in interval [1/6;5], with the sample size fixed at thirty. The corresponding observations of the explained variable were obtained using the cubic function

$$f^0(x) = 12.5x - 7.5x^2 + x^3 \qquad (6.1)$$

to which we added pseudo-random residuals. We generated one hundred sets of

---

[7] In Appendices D through G, the proofs of the results of sections 3 and 4 are developed under the general case of repeated observations for the regressors. This confirms that all the results will remain valid in this general case.

thirty values from a normal distribution of zero mean and unit variance[8]. To obtain the residuals, we added to the values of $f^0(x)$, we multiplied these one hundred sets of values successively by 0.25, 0.50 and 1.00. In this way we generated three hundred samples of thirty observations each.

The cubic function (6.1) presents an important variation near the boundaries of interval [1/6;5]. In nonparametric regression, this is synonymous with important bias boundary effects, *i.e.*, tendency to strongly biased estimations in boundary regions. This characteristic will allow us to evaluate the relative performance of the different combinations.

Tables I through III summarize the results of the experiments. For the values of the function, the squared bias, average variance and risk presented on these tables were calculated by the following expressions:

$$\text{average squared bias (ASB)} = (1/30)\sum_{n=1}^{30}(\bar{f}_n - f_n^0)^2$$

$$\text{average variance (AV)} = (1/30)\sum_{n=1}^{30}(1/100)\sum_{i=1}^{100}(\hat{f}_{n(i)} - \bar{f}_n)^2$$

$$\text{risk} = \text{ASB} + \text{AV} = (1/30)\sum_{n=1}^{30}(1/100)\sum_{i=1}^{100}(\hat{f}_{n(i)} - f_n^0)^2$$

where $\bar{f}_n = (1/100)\sum_i \hat{f}_{n(i)}$ and $\hat{f}_{n(i)}$ is the estimation of f_n^0 obtained when using sample i. For the estimation of the derivatives, the concepts are perfectly analogous.

In Table I-A we compare estimator (4.3)[9] with three subtypes of the

---

[8]We used *Matlab* random numbers generator, described in FORSYTHE *et al* (1977).

[9]Using the quadratic weight function $w(z) = 0.75(1 - z^2)$ if $|z| < 1$, 0 otherwise. Simulations which we performed in PINHEIRO (1989) point out that the quality of the fit is not very sensitive to the choice of the weight function.

Nadaraya-Watson (NW) estimator: NW with the Epanechnikov kernel, NW with the optimal kernel of order 6[10] and NW with quadratic kernel and boundary bias correction method suggested by RICE (1984a). For all these estimators, Rice's T criterion (RICE(1984b)) was used to choose the bandwidth h[11]. Results presented in Table I-A clearly point to the superiority of estimator (4.3), especially with regard to the values of the average squared bias.

In Table I-B we compare estimates obtained using (4.3) with those obtained by cubic spline smoothing. For both estimators we used generalized cross validation criterion to choose the bandwidth. The results are much more balanced than in Table I-A: the smoothing spline estimates present a slightly greater risk but a smaller bias than the estimates from (4.3). The exception being the estimates for $\sigma = 1$. For this case the situation is reversed.

In Table II we compare the estimates of the first derivative using (4.4) with those given by two kinds of estimators proposed on the literature. The first is simply:

$$\hat{g}(x) = \frac{\hat{f}(x^{*}) - \hat{f}(x^{**})}{x^{*} - x^{**}} \qquad (6.2)$$

where: 1) $\hat{f}(x)$ is an estimator of $f^{0}(x)$ (in our case, the NW estimator with quadratic kernel or with optimal kernel of order 6); 2) $x^{*} = x + h_{opt}$ if $(x + h_{opt}) \leq \max(x_n)$, $\max(x_n)$ otherwise; 3) $x^{**} = x - h_{opt}$ if

---

[10] $w(z) = = (35/256)(-99z^{6} + 189z^{4} + 105z^{2} + 15)$ if $|z| < 1$, $w(z) = 0$ otherwise (e.g. GASSER *et al* (1985)).

[11] Simulations reported by RICE (1984b) and HÄRDLE *et al* (1988) seem to indicate that Rice's criterion is the best risk minimization criterion in the context of kernel nonparametric regression.

$(x - h_{opt}) \geq \min(x_n)$, $\min(x_n)$ otherwise; 4) h_{opt} is the optimal value of h associated with $\hat{f}(x)$. The asymptotic behavior of this estimator is studied by ULLAH (1989). Another possible estimator for the first derivative is:

$$\hat{g}(x) = h^{-2}\sum_{n=1}^{N}\left[\int_{s_{n-1}}^{s_n} w_1^*((x-s)/h)ds\right]y_n \qquad (6.3)$$

where $x_n \leq x_{n+1}$ $(n=1,\ldots,N-1)$, $s_n = (x_n + x_{n+1})/2$, $(n=1,\ldots,N-1)$, $s_N = x_N$. We used the optimal values of h given by the NW estimator with kernel of order 6 and the weight function $w_1^*(z) = (105/32)(-9z^5 + 14z^3 - 5z)$ if $|z| < 1$, 0 otherwise. GASSER *et al* (1985) show that this function is optimal in the class of first derivative kernels of order 5.

Results presented in Table II point to a remarkable superiority of (4.4) compared to other estimators. The smoothing spline estimates are the single exception. As for the case of the estimation of the function values, these last estimates present a smaller average squared bias and a greater variance than those obtained from the estimator (4.4). Adding up bias and variance, estimates using smoothing splines once again present a greater risk than the estimates using (4.4).

Average simulated bias, variance and risk do not allow the detection of eventual local differentiated behaviors, namely in boundary regions. These boundaries are critical for nonparametric regression as seen above. For the same estimators which were considered in Tables I and II , we evaluated average risk for the ten observations in each sample closer to the boundary. Results are presented in Tables III-A and III-B. First, if we compare Tables I and II with Table III, we may note that the risk in the boundary region is clearly greater than in the sample average. But the more important conclusion is that the estimators we suggest and the smoothing splines estimators are those that perform better in the boundary region, even better than NW-2BC, which includes an explicit correction of boundary effects.

7.AN APPLICATION

In the previous section we pointed out that estimators (4.3)-(4.4) present relatively good performance in boundary regions and, consequently, in cases of small samples. In this section we apply these estimators to explore a possible relationship between the primary energy requirements per unit of GDP in OECD countries and the respective GDP per capita. We make use of twenty-four observations of both variables (corresponding to the twenty-four member countries). In Appendix J we present the data set and the results. Figure 2 gives a graphic illustration of the results.

This application serves as a warning against the problem of ill conditionning of the data. As the generality of nonlinear and nonparametric methods, criterion function (1.3) is scale sensitive. Asymptotically, scale sensitivity does not affect consistency, but for small samples it may be determinant for the goodness of the fit. This happens because the small number of observations imposes some constraints in the admissible range of the bandwidth. If the scale of explanatory variables is very different from the scale of the explained variable the optimal value of h in terms of risk minimization will be far from the admissible range. To rule out this problem it will be convenient to multiply each explanatory variable by a scale factor, thereby uniformizing the standard deviations of the explained and explanatory variables.

8. CONCLUDING REMARKS

The approach we developped in this paper presents some interesting features. First, it clarifies the relationship between classic linear econometric methods and smoothing splines estimators, undoubtly one of the most popular nonparametric regression estimators.

But we are also convinced that (4.3)-(4.4) have their own importance as estimators. In fact, simulations which we have carried out seem to indicate a good relative performance of our estimators, even comparing with smoothing splines which were the better alternative. This good performance was also verified near the boundaries, and it is undeniable that a good behavior in boundary region may be essential in many problems applied to economics, characterized by small samples.

Furthermore, the unbiasness property in the case of linearity of the true function is promising. It opens the way to the developpement of a test of linearity of the function, against a general alternative of non-linearity of unspecified form.

APPENDIX A

Proof of Lemma 2.1

Since $\mathcal{X}$ is bounded and convex, theorem 1.31(3) of ADAMS (1975, p.11) guarantees that the following imbedding exists and is compact:

$$C^2(\bar{\mathcal{X}}) \to C^1(\bar{\mathcal{X}}) \quad [12] \qquad \text{(a.1)}$$

By definition, an imbedding is compact if the imbedding identity operator $\mathfrak{I}$ is compact, which implies that the operator $\mathfrak{I}$ from $C^2(\bar{\mathcal{X}})$ into $C^1(\bar{\mathcal{X}})$ is compact. Also by definition, this means that $A = \mathfrak{I}(A)$ is precompact in $C^1(\bar{\mathcal{X}})$ whenever A is bounded in $C^2(\bar{\mathcal{X}})$. As $\mathcal{F}$ is bounded in $C^2(\bar{\mathcal{X}})$, $\mathcal{F}$ and $\bar{\mathcal{F}}$ are, respectively, precompact and compact in $C^1(\bar{\mathcal{X}})$.

---

[12] As remarked by ADAMS (1975), 1.31, p.12), the compactness of the imbedding (a.1) may be obtained under less restrictive assumptions than the convexity of $\bar{\mathcal{X}}$. But the convexity of $\bar{\mathcal{X}}$ is a convenient and natural hypothesis for our purposes.

APPENDIX B

Proof of Lemma 2.2

The proof will be developed in two parts. In a first part we will prove that the second term of (1.3) converges to zero almost surely, provided that $h \to 0$ almost surely. In the second part, we will prove that the first term of (1.3) converges to $\bar{Q}(f,f^0)$.

(i) By the mean-value theorem, we have

$$f(x_n) = f(x_s) + \Delta x'_{sn} \frac{\partial f}{\partial x}(\bar{x}_{sn})$$

where $\bar{x}_{sn} = x_s + \upsilon \Delta x_{sn}$ for some $0 \le \upsilon \le 1$. Hence:

$$w(\|\Delta x_{sn}\|/h_n)\left[\frac{f(x_n) - f(x_s) - \Delta x'_{sn} \frac{\partial f}{\partial x}(x_s)}{h_N}\right]^2 =$$

$$= w(\|\Delta x_{sn}\|/h_N)\left\{\left(\frac{\Delta x_{sn}}{h_N}\right)'\left[\frac{\partial f}{\partial x}(\bar{x}_{sn}) - \frac{\partial f}{\partial x}(x_s)\right]\right\}^2$$

(by the Cauchy-Schwarz inequality)

$$\le w\left(\frac{\|\Delta x_{sn}\|}{h_N}\right)\left(\frac{\|\Delta x_{sn}\|}{h_N}\right)^2 \left\|\frac{\partial f}{\partial x}(\bar{x}_{sn}) - \frac{\partial f}{\partial x}(x_s)\right\|^2$$

$$\le w\left(\frac{\|\Delta x_{sn}\|}{h_N}\right)\left\|\frac{\partial f}{\partial x}(\bar{x}_{sn}) - \frac{\partial f}{\partial x}(x_s)\right\|^2$$

(with $\bar{w} = \sup(w(u) \mid u \in [-1,1])$)

$$\le \bar{w} \sup_{\|x - x_s\| \le h_N} \left\|\frac{\partial f}{\partial x}(x) - \frac{\partial f}{\partial x}(x_s)\right\|^2$$

$$\le K\bar{w} \max_{j=1,\ldots,K} \sup_{\|x - x_s\| \in h_N} \left[\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_s)\right]^2 \xrightarrow[h_N \to 0 \text{ a.s.}]{} 0 \text{ a.s.}$$

because $f \in C^2(\bar{X})$. Consequently, we will also have

$$0 \le N^{-1}\sum_s N_s^{-1}\sum_n w\left(\frac{\|\Delta x_{sn}\|}{h_N}\right)\left[\frac{f(x_n) - f(x_s) - \Delta x'_{sn} \frac{\partial f}{\partial x}(x_s)}{h_n}\right] \xrightarrow[h_N \to 0 \text{a.s.}]{} 0 \text{ a.s.}$$

and (1.3) will converge almost surely to the limit of its first term.

(ii) It remains to prove that the first term of (1.3) converges uniformly and

almost surely to

$$\bar{Q}(f,f^0) = \theta^2\sigma^2 + \int_{\mathcal{X}} [f(x) - f^0(x)]^2 d\mu(x)$$

For this part of the proof we follow directly GALLANT and WHITE (1989, p.4.2-4.3). Let us begin by writing:

$$N^{-1}\sum_{s=1}^{N}(y_s - f(x_s))^2 = N^{-1}\sum_{s=1}^{N}\{l(x_s)e_s + [f^0(x_s) - f(x_s)]\}^2 \qquad (b.1)$$

If we denote by E the support of the distribution P(e), by assumptions (A1) and (A5) we have:

$$\int_E \int_{\mathcal{X}} \{l(x)e + [f^0(x) - f(x)]\}^2 dP(e)\, d\mu(x) =$$

$$= \int_{\mathcal{X}} l^2(x)d\mu(x) \int_E e^2 dP(e) + \int_{\mathcal{X}} [f^0(x) - f(x)]^2 d\mu(x) +$$

$$2\int_E e dP(e) \int_{\mathcal{X}} l(x)[f^0(x) - f(x)]d\mu(x)$$

$$= \theta^2\sigma^2 + \int_{\mathcal{X}} [f^0(x) - f(x)]^2 d\mu(x) \qquad (b.2)$$

The Uniform Strong Law of Large Numbers, as stated by GALLANT (1987, P.159), guarantees the almost sure uniform convergence of (b.1) to (b.2) if there exists an integrable function d(x,e) which dominates $\{l(x)e + [f^0(x) - f(x)]\}^2$. The following function satisfies this domination requirement:

$$d(e,x) = (L|e| + 1)^2[2 \sup_{f\in\mathcal{F}} |f(x)| + 1]^2$$

In fact,

(a) $$\int_E \int_{\mathcal{X}} d(e,x)dP(e)d\mu(x) = \int_E (L|e| + 1)^2 dP(e) \int_{\mathcal{X}} [2 \sup_{f\in\mathcal{F}} |f(x)| + 1]^2 d\mu(x)$$

Since e^2 is integrable and

$$2 \sup_{f\in\mathcal{F}} |f(x)| + 1 \le 2 \sup_{f\in\mathcal{F}} \|f\|_{II} + 1 \le 2B + 1 < \infty$$

d(e,x) is an integrable function.

(b) $$d(e,x) = (L|e| + 1)^2[2 \sup_{f\in\mathcal{F}}|f(x)| + 1]^2$$

$$= \{L|e| + 2 \sup_{f\in\mathcal{F}}|f(x)| + 1 + 2L|e| \sup_{f\in\mathcal{F}}|f(x)|\}^2$$

$$\ge \{l(x)e + [f^0(x) - f(x)]\}^2$$

APPENDIX C

Proof of Lemma 2.3

(i) Let $\|.\|_{m,q,S}$ denote the Sobolev norm

$$\|f\|_{m,q,S} = \left\{ \sum_{|\lambda| \leq m} \int_S |D^\lambda f(x)|^q \, dx \right\}^{1/q}$$

Denote by $W_{m,q,S}$ the normed, linear space of real value functions f with $\|f\|_{m,q,S}$ finite. By Lemma A.5 of GALLANT and NYCHKA (1987, p.389),

$$\mathcal{P}_\infty(\mathbb{R}^K) = \bigcup_{p=0}^{\infty} \mathcal{P}_p(\mathbb{R}^K)$$

with $\mathcal{P}_p(\mathbb{R}^K) = \{f:\mathbb{R}^k \mathbb{R} | f(x) = P_p(x)\phi(x)\}$, is $\|.\|_{m,2,\mathbb{R}^K}$-dense in $C_0^\infty(\mathbb{R}^K)$ for any integer $m \geq 1$, where $C_0^\infty(\mathbb{R}^K)$ is the collection of infinitely many times continuously differentiable functions with compact support. Therefore the restriction to $\mathcal{X}$ of $\mathcal{P}_\infty(\mathbb{R}^K)$, $\mathcal{P}_\infty(\mathcal{X})$, is $\|.\|_{m,2,\mathcal{X}}$ – dense in $C_0^\infty(\mathbb{R}^K)|_{\mathcal{X}}$, where $C_0^\infty(\mathbb{R}^K)|_{\mathcal{X}}$ is the restriction to $\mathcal{X}$ of $C_0^\infty(\mathbb{R}^K)$.

(ii) As $\mathcal{X}$ is convex and bounded, it has both the segment property (SP) and the strong local Lipschitz property (SLLP) (ADAMS, 1975, P.66-67). The SP implies that $C_0^\infty(\mathbb{R}^K)|_{\mathcal{X}}$ is $\|.\|_{m,2,\mathcal{X}}$ – dense in $W_{m,2,\mathcal{X}}$ (ADAMS, 1975, th.3.18, p.54). The SLLP implies that the following imbedding exists and is compact (ADAMS, 1975, th.6.2, part III.1, p.144):

$$W_{2+[K/2],2,\mathcal{X}} \rightarrow C^1(\bar{\mathcal{X}})$$

The result follows.

APPENDIX D

Expressions for the case of repeated observations

Let us define:

1) $\bar{y} = [\bar{y}_1 \ldots \bar{y}_{\bar{N}}]'$ $(\bar{N}\times 1)$ and $X = [\bar{x}_1 \ldots \bar{x}_{\bar{N}}]'$ $(\bar{N}\times K)$;

2) $\Omega = \text{diag}_n(\bar{N}_n^{-1})$ $(\bar{N}\times\bar{N})$;

and also let us re-define:

3) $W_s = \text{diag}_n(\bar{N}_n w_{sn})$ $(\bar{N}\times\bar{N})$ and $W = \text{diag}_s(W_s)$ $(\bar{N}^2\times\bar{N}^2)$;

4) $W_s^* = h^{-2}\bar{N}_s N_s^{-1} W_s$ $(\bar{N}\times\bar{N})$ and $W^* = \text{diag}_s(W_s^*)$ $(\bar{N}^2\times\bar{N}^2)$;

5) $r_s = [r_{s1} \ldots r_{st} \ldots r_{s\bar{N}}]'$ $(\bar{N}\times 1)$ and $r = [r_1' \ldots r_s' \ldots r_{\bar{N}}']'$ $(\bar{N}^2\times 1)$;

6) $D_s = I_{\bar N} - \iota e_s'$, with $\iota = [1\ldots1]'$ $(\bar N\times1)$ and e_s the sth column of the identity matrix $I_{\bar N}$.

7) $J = \iota\otimes I_{\bar N}$ $(\bar N^2\times\bar N)$, $Z = [(J-\bar I\mathbb{D}) \quad \bar I\mathbb{D}(I\otimes X)]$ $(\bar N^2\times\bar N(K+1))$, $\beta = [f'\ g']'$ $(\bar N(K+1)\times1)$ and $\Sigma = \bar W^{+} + J\Omega J'$;

8) $\bar W = \mathrm{diag}_s(\bar W_s)$, with $\bar W_s$ identical to W_s^{*} except that its element (s,s) is null;

9) $\bar W^{+}$ is the Moore–Penrose generalized inverse of $\bar W$: $\bar W^{+} = \mathrm{diag}_s(\bar W_s^{+})$, with $\bar W_s^{+} = h^2N_s\,\mathrm{diag}(w_{s1}^{+},\ldots,w_{s,s-1}^{+},0,w_{s,s+1}^{+},\ldots,w_{s\bar N}^{+})$ and, for $n\neq s$, $w_{sn}^{+} = 1/w_{sn}$ if $w_{sn} > 0$, $w_{sn}^{+} = 0$ otherwise;

10) $\bar I = \bar W^{+}\bar W = \mathrm{diag}_s(\bar I_s)$ $(\bar N^2\times\bar N^2)$ with $\bar I_s = \mathrm{diag}_t(i_{sn})$ and $i_{sn} = 1$ if $w_{sn}^{+} > 0$, $i_{sn} = 0$ otherwise.

Directly from (3.1′) we obtain:

$$Q_{N,h}(f,g) = N^{-1}(\bar y - f)'\Omega^{-1}(\bar y - f) + N^{-1}r'W^{*}r \tag{3.2′}$$

and the expressions of sections 3 and 4 will be modified as follows:

$$Q_{N,h}(f,g) = N^{-1}[J(\bar y - f) + \bar I r]'(\bar W^{+} + J\Omega J')^{-}[J(\bar y - f) + \bar I r] \tag{3.3a′}$$

$$Q_{N,h}(f,g) = N^{-1}(J\bar y - Z\beta)'\Sigma^{-}(J\bar y - Z\beta) \tag{3.3b′}$$

$$\frac{\delta Q}{\delta f} = -2N^{-1}\{\Omega^{-1}(\bar y - f) - D'W^{*}[Df - \mathbb{D}(I\otimes X)g]\} = 0 \tag{4.1a′}$$

$$\frac{\delta Q}{\delta g} = -2N^{-1}(I\otimes X')\mathbb{D}'W^{*}[Df - \mathbb{D}(I\otimes X)g] = 0 \tag{4.1b′}$$

$$H = \frac{\delta^2 Q}{\delta\beta\delta\beta'} = 2N^{-1}\begin{bmatrix}(\Omega^{-1} + D'W^{*}D) & -D'W^{*}\mathbb{D}(I\otimes X)\\ -(I\otimes X)\mathbb{D}'W^{*}D & (I\otimes X')\mathbb{D}'W^{*}\mathbb{D}(I\otimes X)\end{bmatrix} =$$

$$= 2N^{-1}\left\{\begin{bmatrix}\Omega^{-1} & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix}D'\\ -(I\otimes X')\mathbb{D}'\end{bmatrix}W^{*}[D\ -\mathbb{D}(I\otimes X)]\right\}$$

$$(I\otimes X')\mathbb{D}'W^{*}\mathbb{D}(I\otimes X) = \mathrm{diag}_s\{h^{-2}\bar N_s N_s^{-1}(D_sX)'W_s(D_sX)\} \tag{4.2′}$$

$$\hat f = A\bar y \tag{4.3′}$$

$$\hat g_s = [(D_sX)'W_s(D_sX)]^{-1}(D_sX)'W_sD_s\hat f \qquad (s=1,\ldots,\bar N) \tag{4.4′}$$

$$A = \{I_{\bar N} + \Omega D'[W^{*} - W^{*}\mathbb{D}(I\otimes X)[(I\otimes X')\mathbb{D}'W^{*}\mathbb{D}(I\otimes X)]^{-1}(I\otimes X')\mathbb{D}'W^{*}]D\}^{-1} =$$

$$= \{I_{\bar N} + h^{-2}\Omega\textstyle\sum_{s=1}^{\bar N}\bar N_s N_s^{-1}D_s'[W_s - W_s(D_sX)'[(D_sX)'W_s(D_sW)]^{-1}(D_sX)'W_s]D_s\}^{-1} \tag{4.5′}$$

$$h^{-2}C^{*} = W^{*} - W^{*}\mathbb{D}(I\otimes X)[(I\otimes X')\mathbb{D}'W^{*}\mathbb{D}(I\otimes X)]^{-1}(I\otimes X')\mathbb{D}'W^{*} = h^{-2}\mathrm{diag}_s(C_s^{*})$$

$$C_s^* = \bar{N}_s N_s^{-1}(W_s - W_s(D_s X)[(D_s X)' W_s (D_s X)]^{-1}(D_s X)' W_s)$$

$$B = D' C^* D = \sum_{s=1}^{\bar{N}} D_s' C_s^* D_s$$

$$A^{-1} = I_{\bar{N}} + h^{-2}\Omega B$$

APPENDIX E[13]

Proof of invariance of (3.3') to the choice of the g-inverse

To show that (3.3') is invariant with respect to the choice of the g-inverse, let $U = [(\bar{W}^+)^{1/2} \quad J\Omega^{1/2}]'$ and $\eta = [r'\bar{W}^{1/2} \quad \bar{\varepsilon}'\Omega^{-1/2}]'$, where $\bar{\varepsilon} = \bar{y} - f$. Note that $\Sigma = \bar{W}^+ + J\Omega J' = U'U$, $(J\bar{u} + \bar{I}r) = U'\eta$ and (from (3.3a')):

$$Q_{N,h}(f,g) = N^{-1}\eta' U(U'U)^{-}U'\eta \qquad (e.1)$$

Invariance of (e.1) follows directly from the invariance of idempotent quadratic forms with generalized inverses (e.g., RAO and MITRA (1971, lemma 2.2.6.d, p.22)).

APPENDIX F

Proof of the equivalence of (3.2') and (3.3')

In Appendix E we proved the invariance of (3.3') with respect to the choice of the g-inverse. To show the equivalence of (3.2') and (3.3'), it remains to prove the equivalence for one particular choice of the g-inverse. Let $F = \text{diag}(b_1,\ldots,b_{\bar{N}})$ $(\bar{N}^2 \times \bar{N})$, where b_n is the n-th column of $I_{\bar{N}}$. We have:

$$F'J = I_{\bar{N}}, \quad F'r = 0 \quad \text{and} \quad \bar{W}F = \bar{W}^+F = 0 \quad (\Rightarrow \bar{W}^+\bar{W}F = \bar{I}F = 0)$$

---

[13] In Appendices E through I, the proofs are developed for the general case of repeated observations for the regressors (as in section 5). For the non-repeated observations case, it suffices to make the following substitutions in the expressions of the Appendices: $\bar{y} \to y$, $\bar{\varepsilon} \to \varepsilon$, $\Omega \to I$ and $\bar{N} \to N$.

It is easy to verify that:

$$\Sigma^{-} = \bar{W} - FJ'\bar{W} - \bar{W}JF' + F(\Omega^{-1} + J'\bar{W}J)F' \qquad (f.1)$$

is a g-inverse of Σ. Substituting (f.1) in (3.3a'), we obtain:

$$\bar{N}^{-1}(J\bar{\varepsilon} + \bar{I}r)\Sigma^{-}(J\bar{\varepsilon} + \bar{I}r) = \bar{N}^{-1}\bar{\varepsilon}'\Omega^{-1}\bar{\varepsilon} + \bar{N}^{-1}r'\bar{W}r \qquad (f.2)$$

As $r_{ss} = 0$ for every s, we have $r'\bar{W}r = r'W^{*}r$ and (f.2) is identical to (3.2').

APPENDIX G

Proof of $v'\Sigma v = 0 \Rightarrow Z'v = 0$

Note that $Z = [(J-\bar{I}D) \quad \bar{I}D(I\otimes X)] = U'C$, with $C = [B_1'\bar{W}^{1/2} \quad B_2'\Omega^{-1/2}]'$, $B_1 = [-D \quad D(I\otimes X)]$ and $B_2 = [I_{\bar{N}} \quad O]$. Then:

$$v'\Sigma v = 0 \Leftrightarrow (Uv)'(Uv) = 0 \Leftrightarrow Uv = 0 \Rightarrow Z'v = C'Uv = 0$$

APPENDIX H

Proof of the properties of the smoothing matrix A

Let us begin by noting that

$$\det(A^{-1} - \eta I) = 0 \Leftrightarrow \det[\Omega^{1/2}B\Omega^{1/2} - h^2(\eta - 1)I] = 0 \qquad (h.1)$$

As $\Omega^{1/2}B\Omega^{1/2}$ is a symmetric non-negative definite matrix, its eigenvalues are all real and non-negative: $h^2(\eta_i - 1) \geq 0 \Rightarrow \eta_i \geq 1 \Rightarrow 0 < \lambda = \eta_i^{-1} \leq 1$ $(i=1,\ldots,\bar{N})$, which proves proposition 1.

From (h.1) we obtain (with $\lambda_i = \eta_i^{-1}$):

$$\det[\Omega B - h^2\lambda_i^{-1}(1 - \lambda_i)I] = 0 \Leftrightarrow \upsilon_i = h^2\lambda_i^{-1}(1 - \lambda_i) \Leftrightarrow \lambda_i = (1 + h^{-2}\upsilon_i)^{-1}$$

and

$$Az = \lambda z \Leftrightarrow A^{-1}z = \lambda^{-1}z \Leftrightarrow (I + h^{-2}\Omega B)z = \lambda^{-1}z \Leftrightarrow \Omega Bz = z$$

with

$$\upsilon = h^2\lambda_i^{-1}(1 - \lambda_i) \qquad (h.2)$$

To show proposition 2, just note that $Az = z \Leftrightarrow$ (by (h.2)) $\Omega Bz = 0 \Leftrightarrow Bz = 0$ and that

$$B\iota = \sum_{s=1}^{\bar{T}} D'_s C^*_s D_s \iota = 0 \quad \text{(because } D_s \iota = 0, \forall s\text{)}$$

$$BX = \sum_{s=1}^{\bar{T}} D'_s C^*_s D_s X = 0 \quad \text{(because } C^*_s D_s X = 0, \forall s\text{)}$$

APPENDIX I

$$\sum_n a_{sn} = 1, \ \sum_n a_{sn} \Delta x_{sn} = 0 \text{ and } 0 < a_{ss} \leq 1 \ \forall s$$

It follows directly from $A\iota = \iota$ that $\sum_s a_{sn} = 1$ $(s=1,\ldots,\bar{N})$ (proposition 2). Recall that D_s is idempotent. Hence:

$$B(D_s X) = \sum_{s=1}^{\bar{N}} D'_s C^*_s D_s X = 0 \Leftrightarrow A(D_s X) = D_s X$$

Taking the line s of this last equality, we have $\sum_n a_{sn} \Delta x_{sn} = \Delta x_{ss} = 0$. To show that $a_{ss} > 0$ $(s=1,\ldots,\bar{N})$, note that

$$A = \Omega^{1/2} L \Omega^{-1/2} \quad \text{with} \quad L = (I + h^{-2}\Omega^{1/2} B \Omega^{1/2})^{-1} \quad \text{positive definite}$$

We have $a_{ss} = T_s^{-1/2} l_{ss} T_s^{1/2} = l_{ss} > 0$. Finally, to show that $a_{ss} \leq 1$ $(s=1,\ldots,\bar{T})$, we write:

$$a_{ss} = b'_s A b_s = b'_s \Omega^{-1/2} L \Omega^{-1/2} b_s = b'_s L b_s = b'_s R_* \Gamma R'_* b_s$$

where Γ is the diagonal matrix of eigenvalues γ_i (≤ 1) of L and R_* is an orthogonal matrix. Let r_{*s} be the transposed s-th line of R_*:

$$a_{ss} = r'_{*s} \Gamma r_{*s} = \sum_{i=1}^{\bar{N}} \gamma_i r^2_{*si} \leq \sum_{i=1}^{\bar{N}} r^2_{*si} = 1$$

APPENDIX J

Application to OECD data

All data refer to 1988 (Source: "OECD in Figures", 1990 ed., Suppl. to *The OECD Observer*, No 164 June/July 1990).

(1) Gross Domestic Product (at market prices) per capita – Thousands of US dollars (at current prices using current exchange rates)

(2) GDP per capita after scale correction ((1)×std(3)/std(1), where std stands for standard deviation)

(3) Total Primary Energy Requirements (in million tons of oil equivalent) per unit of GDP

(4) Estimated function values for each country ($\hat{f}_n$ by (4.3)) - std in parenthesis (conditional in h)

(5) Estimated derivative values for each country ($\hat{g}_n$ by (4.4)) - std in parenthesis (conditional in h)

	(1)	(2)	(3)	(4)		(5)	
Turkey	1.3030	0.0256	0.7126	0.6224	(0.0768)	-2.7206	(0.7284)
Portugal	4.2650	0.0839	0.3767	0.4638	(0.0486)	-2.6330	(0.7034)
Greece	5.5440	0.1032	0.3895	0.4197	(0.0433)	-1.9415	(0.5867)
Spain	8.7220	0.1716	0.2488	0.2970	(0.0437)	-1.6977	(0.5565)
Ireland	9.1820	0.1807	0.2988	0.2904	(0.0438)	0.0506	(0.3375)
NewZeland	12.5550	0.2471	0.3421	0.2980	(0.0322)	0.0245	(0.2483)
U.Kingdon	14.4130	0.2836	0.2534	0.2985	(0.0275)	-0.0176	(0.2421)
Italy	14.4300	0.2840	0.1830	0.2981	(0.0274)	-0.0116	(0.2420)
Australia	14.9400	0.2940	0.3348	0.2988	(0.0263)	-0.0251	(0.2418)
Belgium	15.1800	0.2987	0.3059	0.2985	(0.0257)	-0.0270	(0.2418)
Netherlands	15.4610	0.3043	0.2827	0.2983	(0.0252)	-0.0290	(0.2419)
Austria	16.7480	0.3296	0.2267	0.2970	(0.0233)	-0.0738	(0.2446)
France	17.0020	0.3346	0.2199	0.2966	(0.0231)	-0.0766	(0.2446)
Luxembourg	17.5920	0.3462	0.5091	0.2973	(0.0228)	-0.0963	(0.2450)
Canada	18.6750	0.3675	0.5149	0.2942	(0.0230)	-0.2088	(0.2506)
U.States	19.5580	0.3449	0.4003	0.2894	(0.0239)	-0.2513	(0.2545)
Germany	19.5810	0.3854	0.2281	0.2884	(0.0240)	-0.2504	(0.2545)
Denmark	20.9120	0.4115	0.7171	0.2804	(0.0266)	-0.3340	(0.2619)
Finland	21.2660	0.4185	0.2809	0.2785	(0.0275)	-0.3438	(0.2635)
Sweden	21.5460	0.4240	0.3091	0.2766	(0.0283)	-0.3494	(0.2647)
Norway	21.6540	0.4262	0.3073	0.2759	(0.0286)	-0.3494	(0.2648)
Japan	23.1900	0.4564	0.1402	0.2633	(0.0338)	-0.3916	(0.2708)
Iceland	23.9360	0.4711	0.2831	0.2574	(0.0365)	-0.4338	(0.2855)
Switzerland	27.5810	0.5428	0.1532	0.1997	(0.0619)	-0.8033	(0.5625)

REFERENCES

ADAMS, R. A. (1975) *Sobolev Spaces*, New York: Academic Press.

CRAVEN, P. and G. WAHBA (1979) "Smoothing Noisy Data with Spline Functions", *Numerische Mathematik*, 31, p.377-403.

EUBANK, R. L. (1988) *Spline Smoothing and Nonparametric Regression*, New York: Marcel Dekker.

FORSYTHE, G. E., M. A. MALCOLM and C. B. MOLER (1977) *Computer Methods for Mathematical Computations*, Englewood Cliffs N. J.: Prentice-Hall.

GALLANT, A. R. (1985) "Identification et Convergence en Régression Semi-Non-

Paramétrique", *Annales de l'INSEE*, 59/60, p.239-264.

GALLANT, A. R. (1987) *Nonlinear Statistical Models*, New York: J. Wiley & Sons.

GALLANT, A. R. and D. W. NYCHKA (1987) "Semi-Nonparametric Maximum Likelihood Estimation", *Econometrica*, 55, p.363-390.

GALLANT, A. R. and H. WHITE (1987) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Oxford: Basil Blackwell.

GALLANT, A. R. and H. WHITE (1989) "On Learning the Derivatives of an Unknown Mapping with Multilayer Feedforward Networks", University of California, San Diego, Department of Economics Discussion Paper.

GASSER, T., H.-G., MÜLLER and V. MAMMITZCH (1985) "Kernels for Nonparametric Curve Estimation", *Journal of the Royal Statistical Society*, Series B, 47, p.238-252.

GRENANDER, U. (1981) *Abstract Inference*, New York: J. Wiley & Sons.

HÄRDLE, W. (1990) *Applied Nonparametric Regression*, Econometric Society Monographs, Cambridge: Cambridge University Press.

HÄRDLE, W., P. HALL and J. S. MARRON (1988) "How Far Are Automatically Chosen Regression Smoothing Parameters From Their Optimum?" (with discussion), *Journal of the American Statistical Association*, 83, p.86-99.

HORNIK, K., M. STINCHCOMBE and H. WHITE (1990) "Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks", University of California, San Diego, Department of Economics Discussion Paper (to appear in *Neural Networks).*

MÜLLER, H.-G. (1987) "Weighted Local Regression and Kernel Methods for Nonparametric Curve Fitting", *Journal of the American Statistical Association*, 82, p.231-238.

PINHEIRO, M. R. (1989) *Une Approche Econométrique de la Régression*

Non-paramétrique: Estimation et Test de Linéarité, Doctoral Thesis, University of Geneva, Switzerland.

RAO, C. R. and S. K. MITRA (1971) *Generalized Inverses of Matrices and its Applications*, New York: John Wiley & Sons.

RICE, J. (1984a) "Boundary Modification for Kernel Regression", *Communications in Statistics - Theory and Methods*, 13, p.893-900.

RICE, J. (1984b) "Bandwidth Choice for Nonparametric Regression", *Annals of Statistics*, 12, p.1215-1230.

ULLAH, A. (1989) "Nonparametric Estimation and Hypothesis Testing in Econometric Models", *Empirical Economics*, 13, p.223-249.

WAHBA, G. (1975) "Smoothing Noisy Data with Spline Functions", *Numerische Mathematik*, 24, p.383-393.

TABLE I-A

Estimation of Function Values

(4.3): estimator (4.3) (with quadratic weight function)
NW-2: Nadaraya-Watson estimator with quadratic kernel
NW-2BC: Same as NW-2 but with boundary effects correction (Rice's method)
NW-6: Nadaraya-Watson estimator with optimal kernel of order 6
BANDWIDTH CHOICE: RICE'S T CRITERION

σ		(4.3)	NW-2	NW-2BC	NW-6
0.25	average squared bias	0.0089	0.0783	0.0150	0.0681
	average variance	0.0166	0.0179	0.0213	0.0232
	risk	0.0255	0.0963	0.0363	0.0913
0.50	average squared bias	0.0110	0.1192	0.0253	0.1104
	average variance	0.0641	0.0602	0.0726	0.0808
	risk	0.0751	0.1794	0.0979	0.1912
1.00	average squared bias	0.0141	0.2128	0.0679	0.1978
	average variance	0.2565	0.2080	0.2572	0.2474
	risk	0.2705	0.4209	0.3251	0.4452

TABLE I-B

Estimation of Function Values

(4.3): estimator (4.3) (with quadratic weight function)
SSPLINES: Smoothing splines estimator
BANDWIDTH CHOICE: GCV CRITERION

σ		(4.3)	SSPLINES
0.25	average squared bias	0.0025	0.0019
	average variance	0.0212	0.0244
	risk	0.0238	0.0263
0.50	average squared bias	0.0075	0.0065
	average variance	0.0775	0.0852
	risk	0.0850	0.0918
1.00	average squared bias	0.0183	0.0225
	average variance	0.3322	0.3022
	risk	0.3504	0.3247

TABLE II

Estimation of First Derivative Values

(4.4) Rice T: estimator (4.4) (quadratic weight function and Rice's T crit.)
(4.4) GCV: estimator (4.4) (quadratic weight function and GCV crit.)
U-2: estimator (6.2) using the function estimates from NW-2
U-6: estimator (6.2) using the function estimates from NW-6
GM-5: estimator (6.3) with optimal kernel of order 5
SSPLINE: Smoothing splines estimator (with GCV crit.)

σ		(4.4) Rice T	U-2	U-6	GM-5	(4.4) GCV	SSPLINE
0.25	av. squared bias	0.9958	2.9655	8.6805	2.7029	0.4273	0.2768
	average variance	0.1214	0.0910	0.1412	0.1187	0.3068	0.7036
	risk	1.1172	3.0565	8.8217	2.8216	0.7341	0.9803
0.50	av. squared bias	1.1124	4.0303	12.8683	3.2324	0.8702	0.5193
	average variance	0.3734	0.2314	0.4727	0.3211	0.8255	2.2013
	risk	1.4858	4.2617	13.3410	3.5535	1.6957	2.7206
1.00	av. squared bias	1.1738	5.9472	19.9513	4.1587	1.8475	1.0224
	average variance	1.4938	0.6363	0.8552	0.6790	2.1271	7.2191
	risk	2.6676	6.5835	20.7765	4.8378	3.9746	8.2415

TABLE III-A

Estimation of Function Values

in Boundary Regions

(n≤5 or n≥26)

σ		(4.3) Rice T	NW-2	NW-2BC	NW-6	(4.3) GCV	SSPLINE
0.25		0.0456	0.2421	0.0676	0.2289	0.0317	0.0360
0.50	risk	0.1093	0.4021	0.1574	0.4298	0.1061	0.1283
1.00		0.3326	0.8285	0.4862	0.8596	0.4317	0.4586

TABLE III-B

Estimation of First Derivative Values

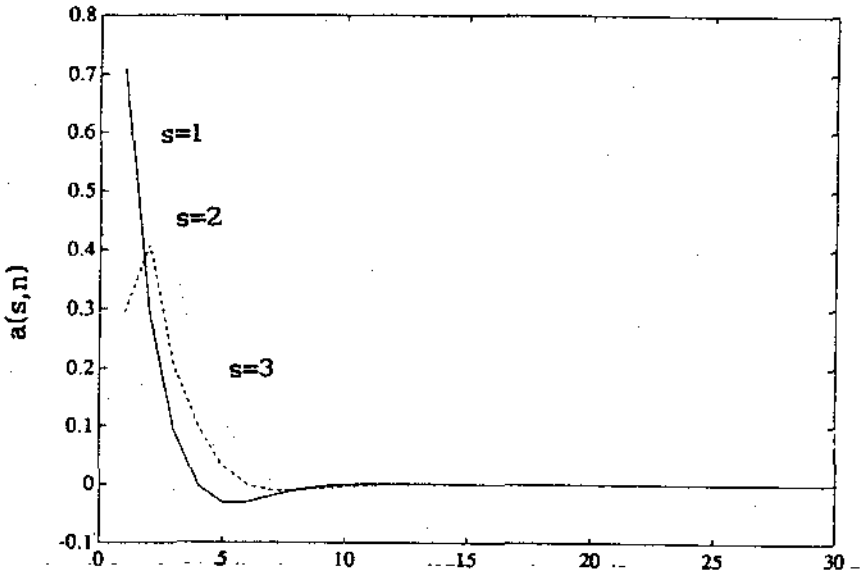
in Boundary Regions

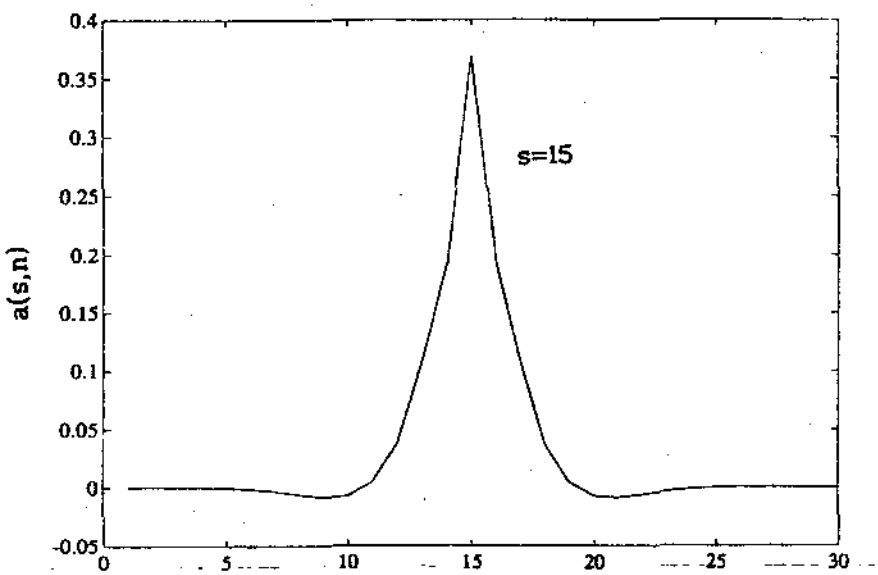
(n≤5 or n≥26)

σ		(4.4) Rice T	U-2	U-6	GM-5	(4.4) GCV	SSPLINE
0.25		3.1554	8.9932	24.1441	8.2504	1.7512	1.8628
0.50	risk	3.9998	12.3271	35.1400	10.0393	3.9998	4.6393
1.00		5.6994	18.6537	52.3589	13.0086	9.1323	12.7684

FIGURE 1

An Example of weights induced by (4.3)





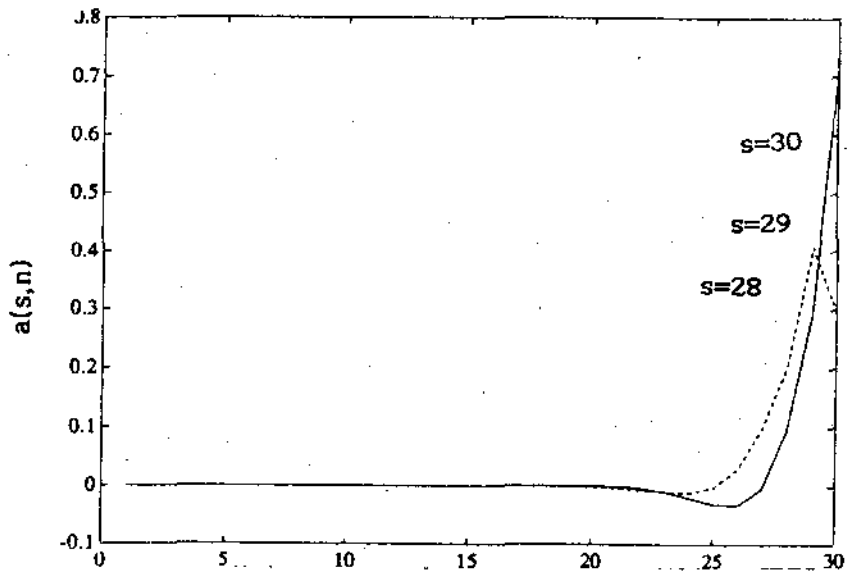


FIGURE 2

Observed and estimated total primary energy requirements

Observed values: joined by a dashed line ---

Estimated values: joined by a full line ——