INVESTIGACIONES ECONOMICAS, vol. XXI(2), 1997, 209-252

NONPARAMETRIC CURVE ESTIMATION: AN OVERVIEW

RICARDO CAO
Universidad de La Coruña
MIGUEL A. DELGADO
Universidad Carlos III de Madrid
WENCESLAO GONZALEZ-MANTEIGA
Universidad de Santiago de Compostela

This paper is an overview of nonparametric curve estimation, specially oriented to applications in econometrics. The discussion is mainly centered on the estimation of densities and regression curves, but other relevant functions (derivatives of regression curves, hazard functions, conditional densities and quantile functions) are also considered. We briefly discuss the basic statistical ideas behind these techniques and, in particular, the important practical problem of the choice of smoothing parameter. Illustrative examples, based on data fron the Spanish Family Expenditure Survey are given. (JEL C14*)*

1. Introduction

Nonparametric curve estimation has been one of the most active fields in statistics during the last decades. A proof of this is the long list of texts and survey papers on this topic.[1]

Research funded by Spanish "Dirección General de Investigación Científica y Técnica" (DGICYT) research grant number: PB94-0494 and "Dirección General de Enseñanza Superior" (DGES) research grant numbers PB95-0826 and PB95-0292. We thank Daniel Miles and Coral del Río for helping with the graphs in Section 6. We also thank the helpful comments of the editor and three referees.

[1] Among them, we point out the books by Tapia and Thompson (1978), Prakasa Rao (1983), Devroye and Györfi (1985), Silverman (1986), Devroye (1987), Eubank (1988), Härdle (1990), Rosenblatt (1991), Scott (1992), Tarter and Lock

In econometrics, the assumption of statistical adecuacy or correct specification has been a continuous concern. Haavelmo asks in his influential 1944 paper: "What is the use of testing, say, the significance of regression coefficients when may be the whole assumption of the linear regression equation is wrong?", Haavelmo (1944, p. 66).

In applied economics is important to isolate hypothesis of interest. Many times the functional form of curves related to certain characteristics of the data generating model are not of specific interest, but misspecification of the maintained hypotheses about these functional forms can imply invalid tests of the hypotheses of interest to the economist. Nonparametric estimation of curves permits to relax the assumptions on the underlying model, allowing the isolation of relevant hypotheses. It also helps to check the goodness of fit of parametric specifications. Hence, it is not surprising the enormous success of the application of nonparametric estimation techniques in econometrics.

Applications of nonparametric techniques in applied economic research are available in generous supply. Hildenbrand and Hildenbrand (1986) applied nonparametric density and regression estimation to study the income distribution based on the UK Family Expenditure Survey (FES). Bierens and Pott-Buter (1991) and Delgado and Miles (1997) applied nonparametric estimation of regression curves to the specification of Engel curves. Diebold and Nason (1990) have investigated the presence of exploitable nonlinearities in forecasting asset prices using nonparametric regression. Pagan and Schwert (1990) studied the specification volatility models for financial data comparing variance estimates. Nonparametric estimation has also been employed in formal specification testing. Rosenblatt (1975a) and Robinson (1991) proposed independence tests based on nonparametric density estimates. Specification tests of regression models have been proposed by Eubank and Spiegelman (1990), Kozek (1991), Härdle and Mammen (1993), Delgado and Stengos (1994) and Zheng (1996), among others. Lavergne and Vuong (1996) and Fan

(1993), Green and Silverman (1994), Wand and Jones (1995), and Simonoff (1996) and the papers by Rosenblatt (1971), Tarter and Kronmal (1976), Fryer (1977), Leonard (1978), Bean and Tsokos (1980), Collomb (1981, 1985), Wegman (1982), Izenman (1991) and Delgado and Robinson (1992).

and Li (1996) have proposed tests for selecting significant explanatory variables without specifying the underlying regression function.

In this paper we give an expository presentation of nonparametric techniques for estimating different types of functions related to probability distributions, with the aim of providing a useful tool for the applied researcher in econometrics. The discussion is mainly focused on the estimation of density and regression curves. In Section 2 we introduce the concepts of bias and variance, which are very important for understanding the performance of nonparametric estimators in practice, and in particular for the choice of the amount of smoothing. In Section 3, we also present the nonparametric estimation of other functions, relevant in econometrics, like derivatives of regression curves, hazard functions, conditional densities and conditional quantiles. Some of the most relevant available techniques for the choice of the smoothing parameter are presented in Section 4, while Section 5 contains some examples to illustrate the usefulness of nonparametric estimation in economic practice, estimating different probability curves using the Spanish Family Expenditure Survey (FES). Finally, we discuss the main benefits and drawbacks of the nonparametric methods in Section 6.

2. Nonparametric estimation of a density or a regression curve

One of the most important curves associated with a random variable is the cumulative distribution function. The distribution function, F, of a population, X, is, for a given x_0, the probability of a very elementary event:

$$F(x_0) = P(X \leq x_0) = E\left[1\left(X \leq x_0\right)\right].$$

Given a random sample $X_1, X_2, \ldots, X_n$, the last expected value is typically estimated by using the pertaining sample mean, resulting in the well known empirical distribution function

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\left(X_i \leq x\right).$$

The empirical distribution function, F_n, is the most popular nonparametric estimator of the cumulative distribution function. Among the properties of this estimator we point out that it is unbiased and also the nonparametric maximum likelihood estimator of the underlying distribution function.

The density function of X, $f(x) = dF(x)/dx$, can be represented as

$$\begin{aligned} f(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h} \qquad [1] \\ &= \lim_{h \to 0} \frac{1}{2h} \int_{x-h}^{x+h} dF(u) \\ &= \lim_{h \to 0} \frac{E\left[1\left(|X-x| \leq h\right)\right]}{2h}. \end{aligned}$$

Now, the last ratio can be estimated replacing the unknown expectation by its empirical analogue, suggesting the naive estimator

$$\hat{f}(x) = \frac{1}{2h} \int_{x-h}^{x+h} dF_n(u) = \frac{1}{n} \sum_{i=1}^{n} \frac{1\left(|X_i - x| \leq h\right)}{2h}. \qquad [2]$$

This estimator was proposed first by Rosenblatt (1956). For a fixed sample size, it does not make sense to define the estimator by the limit, when $h \to 0$, of the previous quantity. In fact this limit is zero since, for h small enough, the numerator in [2] equals zero. However, as $n \to \infty$, one may think of the value h (often called the smoothing parameter or bandwidth) as a sequence h_n tending to zero.

The estimator in [2] is the relative frequency, per unit of length, of the observations within the interval $[x-h, x+h]$. The simple formulae for the mean and the variance of the naive estimator follow from its structure as a sum of iid Bernoulli random variables:

$$\begin{aligned} E\left(\hat{f}(x) - f(x)\right) &= \frac{F(x+h) - F(x-h)}{2h} - f(x), \\ Var\left(\hat{f}(x)\right) &= \frac{\left(F(x+h) - F(x-h)\right) \cdot \left(1 - F(x+h) + F(x-h)\right)}{4nh^2}. \end{aligned}$$

From these expressions one can clearly see that, apart from some particular values of x, the estimator is biased and, for small values of h, the bias and the variance are approximately

$$\begin{aligned} E\left(\hat{f}(x) - f(x)\right) &\simeq \frac{f'''(x)h^2}{6}, \\ Var\left(\hat{f}(x)\right) &\simeq \frac{f(x)}{2nh}. \end{aligned}$$

Thus, it seems reasonable to require, for consistency, that $h \to 0$ and $nh \to \infty$ as $n \to \infty$. The behaviour of these two terms is opposite: while the bias increases with the smoothing parameter, the variance decreases as the bandwidth gets large. Hence, it is intuitive that the choice of the smoothing parameter is very important in practice, since it regulates the balance between the bias (systematic error) and the variance (stochastic error) of the estimator.

Let us now concentrate on the regression setup. Now we are given a two-dimensional random sample $X_1, X_2, \ldots, X_n$. We assume that the pairs $X_i = (Z_i, Y_i)$ are independently distributed copies of (Z, Y). The random variable Z will be called the explanatory variable (or independent variable) and we will refer to Y as the response variable (or the dependent variable). Our interest here is to estimate the regression function

$$m(z) = E(Y|Z = z),$$

whenever it exists.

If the marginal distribution of Z is discrete and has a small number of different values, the estimation of this function only makes sense at these points and it is easy to argue conditionally. Since $m(z)$ is a conditional mean, we just use the data (Z_i, Y_i) for which $Z_i = z$ and average them. In other words, this means using the estimator

$$\hat{m}(z) = \frac{\sum_{i=1}^n 1\left(Z_i = z\right) Y_i}{\sum_{i=1}^n 1\left(Z_i = z\right)}.$$

This is an unbiased estimator that converges to the true value $m(z)$, since

$$\lim_{n\to\infty} \frac{Var(\hat{m}(z))}{\frac{\sigma^2(z)}{np_Z(z)}} = 1,$$

where $\sigma^2(z)$ is the conditional variance of Y given $Z = z$ (assumed to exist) and $p_Z(.)$ is the probability function of Z.

The situation is somehow more complicated when the distribution of Z is absolutely continuous. In this case for all the possible values z, except a finite number of them, there is no a single observation of Y_i where $Z_i = z$. Furthermore, even for these values for which we have observations, the probability of more than one observation at $Z_i = z$ is zero.

In the continuous case, we can think of $m(z)$ as some limit quantity. Since

$$m(z) = E(Y|Z=z) = \int_{\mathbb{R}} yf(y|z)\,dy = \frac{r_Z(z)}{f_Z(z)}, \qquad [3]$$

where f_Z is the density of Z and

$$\begin{aligned} r_Z(z) &= \int_{\mathbb{R}} y f_{Z,Y}(z,y)\,dy \\ &= \lim_{h\to 0} \frac{1}{2h} \int_{z-h}^{z+h} \int_{\mathbb{R}} y F_{Z,Y}(du,y) \\ &= \lim_{h\to 0} \frac{E\left[Y 1\left(|Z-z| \le h\right)\right]}{2h}, \end{aligned}$$

where $f_{Z,Y}(.,.)$ and $F_{Z,Y}(.,.)$ are the bivariate density function and distribution function of (Z,Y) respectively.

This suggests to estimate $r_Z(z)$, as in [2], by means of the two-dimensional empirical distribution function, $\hat{F}_{Z,Y}$,

$$\hat{r}_Z(z) = \frac{1}{2h} \int_{z-h}^{z+h} \int_{\mathbb{R}} y\hat{F}_{Z,Y}(du,dy) = \frac{1}{n}\sum_{i=1}^{n} Y_i \frac{1\left(|Z_i - z| \le h\right)}{2h}, \qquad [4]$$

with $h \to 0$ as $n \to \infty$. Thus, $m(z)$ is estimated by

$$\hat{m}(z) = \frac{\sum_{i=1}^{n} Y_i 1\left(|Z_i - z| \le h\right)}{\sum_{i=1}^{n} 1\left(|Z_i - z| \le h\right)}. \qquad [5]$$

This estimator is just a local average of the Y_i, and it was first suggested by Nadaraya (1964) and Watson (1964). It also can be written as a weighted average

$$\hat{m}(z) = \sum_{i=1}^{n} Y_i W_i(z) \quad \text{with} \quad W_i(z) = \frac{1\left(|Z_i - z| \le h\right)}{\sum_{i=1}^{n} 1\left(|Z_i - z| \le h\right)}. \qquad [6]$$

The observations Y_i with corresponding Z_i value far enough from z are not considered in the average. The weights in [6] can be generalized to any probabilistic weights that dampen the $Y_i's$ with corresponding Z_i value far from z. Consistency properties for a very general kind of weights has been studied by Stone (1977).

Most of the nonparametric methods in curve estimation may be thought in the way presented above. Although for most of the curves is not possible to find unbiased estimators of them, typically one may find different ways of approximating the underlying curve with the property that it is possible to find unbiased estimators for this smoothed function. Generally, the error in the approximation is controlled by a ‘smoothing parameter’. The stochastic error in the estimation also depends on this parameter in an inverse way. This is why the choice of this smoothing parameter is so important in balancing the approximation error (or bias part) and the stochastic error (or variance part). Undoubtedly any practical choice has to be a compromise between both.

The following Subsections will be devoted to the presentation of some of the popular methods in nonparametric density and regression estimation in the light of the approach given above.

2.1. Kernels

The naive estimator in [4] can also written as

$$\hat{r}_Z(z) = \frac{1}{h}\int_{\mathbb{R}} K\left(\frac{u-z}{h}\right)\int_{\mathbb{R}} y\hat{F}_{Z,Y}\left(du, dy\right) = \frac{1}{nh}\sum_{i=1}^{n} Y_i K\left(\frac{Z_i - z}{h}\right), \quad [7]$$

with $K(u) = 1\left(|u| \leq 1\right)/2$. The function K is called the kernel. In fact the kernel we are using is the uniform density function. We can use more general kernel functions giving more weight at zero, like the triangular or the Gaussian densities. These ideas were first presented by Rosenblatt (1956) and Parzen (1962).

Kernel functions are generally required to hold that

$$\int_{\mathbb{R}} uK\left(u\right)du = 0, \int_{\mathbb{R}} K\left(u\right)du = 1 \text{ and } \int_{\mathbb{R}} u^2 K\left(u\right)du < \infty. \quad [8]$$

A kernel which is nonnegative and satisfies [8] corresponds to the density of some distribution with zero mean and finite variance. Under these circumstances for h fixed and $r_Z(.)$ smooth enough

$$\begin{aligned} E\left[\hat{r}_Z(z)\right] &= r_Z\left(z\right) + O\left(h^2\right) \quad [9] \\ Var\left[\hat{r}_Z(z)\right] &= O\left(\frac{1}{nh}\right). \end{aligned}$$

Then $\hat{r}_Z(z)$ is asymptotically unbiased if $h \to 0$ as $n \to \infty$ and it will be consistent if also $nh \to \infty$ as $n \to \infty$. The bias term can be made smaller when $r_Z(.)$ has many derivatives by using particular kernel functions.

2.2. Histograms and regressograms

One could think of using a cumulative histogram to produce nonparametric estimators of the underlying distribution function, but this is a little bit artificial. In order to compute a histogram we have to consider contiguous intervals $I_j = (a_j, a_{j+1}],\ j = 0, 1, 2, \ldots, k, k+1$, where $a_0 = -\infty$, $a_{k+1} = \infty$ and a_1 and a_k are typically chosen in such a way that all the sample is contained in the interval $(a_1, a_k]$. If $f_Z(.)$ is continuous at z, the mean value theorem of integral calculus implies that

$$\lim_{\mu(I_j)\to 0} \frac{P(Z \in I_j)}{\mu(I_j)} = f_Z(z) \quad \text{if} \quad z \in I_j,$$

where $\mu(I_j) = a_{j+1} - a_j$. This suggests the histogram estimator

$$\hat{f}_Z^H(z) = \frac{n^{-1}\sum_{i=1}^n 1(Z_i \in I_j)}{\mu(I_j)}.$$

Now

$$E\left(\hat{f}_Z^H(z)\right) = \frac{P(Z \in I_j)}{\mu(I_j)} = \frac{1}{\mu(I_j)} \int_{I_j} f_Z(u)\, du \overset{def}{=} f_Z^H(z).$$

The bias of the histogram is just the approximation error between $f_Z^H(z)$ and $f_Z(z)$. The length of the interval where z falls has to tend to zero, when the sample size tends to infinity. Otherwise the histogram may not be asymptotically unbiased. In order to find precise rates for the bias, some regularity conditions on $f_Z(.)$ have to be assumed. If there exists a neighbourhood of z where $f_Z(.)$ is twice continuously differentiable and $f_Z''(.)$ is bounded in a neighbourhood of z, the mean value theorem leads to

$$\int_{I_j} f_Z(u)\, du = f_Z(z)\mu(I_j) + f_Z'(z)\int_{I_j}(u-z)du +$$

$$
\begin{aligned}
& + \frac{1}{2}\int_{I_j} f''_Z(\lambda u + (1-\lambda)\, z)(z-u)^2 du \\
= \; & f_Z(z)\mu(I_j) + f'_Z(z)\left(\frac{a_j + a_{j+1}}{2} - z\right)\mu(I_j) \\
& + O(\mu(I_j)^3),
\end{aligned}
$$

where $0 < \lambda < 1$. As a consequence, an asymptotic expansion for $f_Z^H(z)$ comes up:

$$f_Z^H(z) = f_Z(z) + f'_Z(z)\left(\tfrac{a_j + a_{j+1}}{2} - z\right) + O(\mu(I_j)^2),$$

so the smallest bias is attained at the midpoint of the interval.

Using the binomial structure of $\hat{f}_Z^H$, its variance can be computed immediately:

$$Var(\hat{f}_Z^H(z)) = \frac{n^{-1}P(Z \in I_j)P(Z \notin I_j)}{\mu(I_j)^2} = f_Z^H(z)\frac{P(Z \notin I_j)}{n\mu(I_j)}.$$

The first thing to note is that, under the necessary assumption that $\mu(I_j)$ tends to zero,

$$Var(\hat{f}_Z^H(z)) = \frac{f_Z(z)}{n\mu(I_j)} + O\left(\frac{1}{n}\right).$$

Now it is clear that the width of the intervals cannot decrease very fast, since $n\mu(I_j)$ has to tend to infinity, as n goes to infinity.

In a regression context, $r_Z(z) = \lim_{n\to\infty} E\left(Y \cdot 1\left(Z \in I_j\right)\right)/\mu\left(I_j\right)$ is estimated by

$$\hat{r}_Z^H\left(z\right) = \frac{1}{\mu\left(I_j\right)} n^{-1}\sum_{i=1}^{n} Y_i 1\left(Z_i \in I_j\right).$$

The regressogram estimator of $m\left(z\right) = r_Z\left(z\right)/f_Z\left(z\right)$ is given by

$$\hat{m}^H\left(z\right) = \frac{\sum_{i=1}^n Y_i 1\left(Z_i \in I_j\right)}{\sum_{i=1}^n 1\left(Z_i \in I_j\right)}.$$

That is, the regressogram is an average of those $Y_i's$ with corresponding Z_i in the interval I_j.

2.3. Frequency polygons

A reasonable way of balancing the bias and the variance in the histogram consists of computing the mean squared error and to choose the width of the interval by minimizing this expression. For simplicity, let us consider the point $z_0 = a_j$ and denote by $\mathrm{MSE}(\hat{f}_Z^H(z_0))$ the mean squared error of the histogram, then

$$\mathrm{MSE}(\hat{f}_Z^H(z_0)) = \frac{1}{4} f'_Z(z_0)^2 \mu(I_j)^2 + \frac{f_Z(z_0)}{n\mu(I_j)} + O(\frac{1}{n}) + O(\mu(I_j)^3).$$

Now, the value of $\mu(I_j)$ that minimizes the dominant part (the first two terms) of $\mathrm{MSE}(\hat{f}_Z^H(z_0))$ is

$$\left(\frac{2 f_Z(z_0)}{n f'_Z(z_0)^2} \right)^{1/3},$$

whenever $f'_Z(z_0) \neq 0$. If the derivative of f_Z vanishes at z_0, the previous analysis should be extended in the sense of going further on in the Taylor expansion for the bias. In such a situation the bias is even smaller at that particular point and the 'optimal' width is different.

The main problem of this approach is that the formula for the interval width minimizing the asymptotic mean squared error (AMSE) depends on some unknown terms, namely, $f_Z(z_0)$ and $f'_Z(z_0)$. The first term is just what we want to estimate. A typical tool in this situation consists of using a preliminary histogram (may be with some 'ad hoc' choice of the interval length) to estimate $f_Z(z_0)$. The problem is even harder when estimating the first derivative. A preliminary histogram should not be used since the estimator of the first derivative would be zero. This can be solved by modifying the definition of the histogram from the beginning. Given a partition of the whole real line, the idea behind a simple histogram is just to fit the data to a density function that is forced to be constant within every interval of that partition. One may be a little bit more flexible by using densities which are not piece-wise horizontal lines but piece-wise straight lines. By doing this, the estimator of the first derivative of the density function at a given point z is just the slope of the straight line corresponding to the interval where z falls.

The modified version of the histogram (let us denote $\hat{f}_Z^{MH}(.)$) can also be thought as the empirical estimator of some smoothed version, $f_Z^{MH}(.)$, of the density function. This new smoothed version of $f_Z(.)$ is defined as a piece-wise linear function:

$$f_Z^{MH}(z) = b_j + c_j z, \text{ whenever } z \in I_j,$$

where the values b_j and c_j are defined as those that minimize (in b and c) the quadratic difference:

$$g(b,c) = \int_{I_j} (f_Z(u) - b - cu)^2 \, du.$$

This leads to the values:

$$b_j = \frac{F_{1,j}M_{1,j} - F_{0,j}M_{2,j}}{M_{1,j}^2 - M_{2,j}M_{0,j}},$$

$$c_j = \frac{F_{0,j}M_{1,j} - F_{1,j}M_{0,j}}{M_{1,j}^2 - M_{2,j}M_{0,j}},$$

where

$$F_{r,j} = \int_{I_j} u^r f_Z(u) \, du, \ r = 0,1, \ j = 1,2,\ldots,k$$

and

$$M_{r,j} = \int_{I_j} u^r \, du, \ r = 0,1,2, \ j = 1,2,\ldots,k$$

To estimate f^{MH} without any bias, we just have to find unbiased estimators, $b_{n,j}$ and $c_{n,j}$, of b_j and c_j. This is very simple. We define

$$\hat{F}_{Z,r,j} = \frac{1}{n}\sum_{i=1}^{n} 1\left(Z_i \in I_j\right) Z_i^r, \ r = 0,1,$$

and plug these estimators in the formulas for b_j and c_j. This results in

$$b_{n,j} = \frac{\hat{F}_{Z,1,j}M_{1,j} - \hat{F}_{Z,0,j}M_{2,j}}{M_{1,j}^2 - M_{2,j}M_{0,j}},$$

$$c_{n,j} = \frac{\hat{F}_{Z,0,j}M_{1,j} - \hat{F}_{Z,1,j}M_{0,j}}{M_{1,j}^2 - M_{2,j}M_{0,j}}.$$

Finally, the straight line version of the histogram is defined by

$$\hat{f}_Z^{MH}(z) = b_{n,j} + c_{n,j}z, \text{ if } z \in I_j.$$

This modification of the histogram was motivated just to estimate some term in the formula for the interval width that minimizes the asymptotic mean squared error. However, it should be clear that the resulting expression is interesting by itself as an estimator of the underlying density. Another point is that there is no reason to restrict ourselves to constant or linear approximations of the density within every interval. A polynomial of an arbitrary degree or even more complex functions, like exponentials of polynomials, could be used to modify the histogram. These ideas of local parametric density estimation, in the context of the kernel method, have been studied by Loader (1996) and Hjort and Jones (1996). A relevant case is the approximation by exponential of polynomials of degree two. This is equivalent to local fit to a class that includes the normal family.

2.4. Nearest neighbours

One of the features that presents the naive density estimator concerns the fact that the smoothing parameter, h, is constant for all the points where the estimation is performed. One can try to adapt the 'amount of smoothing' to the point where we are estimating the density. In this spirit, recalling the naive estimator and defining the bandwidth h as the distance to the k-th nearest neighbour of the point where the density has to be estimated, we end up with some new estimator of $f_Z(z)$:

$$\hat{f}_Z^N(z) = \frac{1}{nh_k(z)} \sum_{i=1}^{n} 1\left(|Z_i - z| \leq h_k(z)\right) = \frac{k}{nh_k(z)},$$

where $h_k(z)$ is the distance from z to its k-th nearest neighbour. This k nearest neighbour estimator has the disadvantage of not being a density but it has the advantage with respect to the usual kernel that the bandwidth adapts for the data around the point to be estimated. When the data are disperse around z the bandwidth is large and when the data are concentrated around z the bandwidth is small.

This estimator can be extended to general kernel functions:

$$\hat{f}_Z^N(z) = \frac{1}{nh_k(z)} \sum_{i=1}^{n} K\left(\frac{Z_i - z}{h_k(z)}\right).$$

Also $r_Z(z)$ is estimated by

$$\hat{r}_Z^N(z) = \frac{1}{nh_k(z)} \sum_{i=1}^{n} Y_i K\left(\frac{Z_i - z}{h_k(z)}\right).$$

The consistency requirements are $k \to \infty$ and $k/n \to 0$ as $n \to \infty$, despite the conditions of the kernel function. Note that if Z_i is uniformly distributed $k \simeq 1/h_k(z)$. So, the regression function $m(z)$ is estimated by

$$\begin{aligned} \hat{m}^N(z) &= \frac{\sum_{i=1}^n Y_i K\left(\frac{Z_i - z}{h_k(z)}\right)}{\sum_{i=1}^n K\left(\frac{Z_i - z}{h_k(z)}\right)} \\ &= \sum_{i=1}^{n} Y_i W_i(z) \text{ with } W_i(z) \\ &= \frac{K\left(\frac{Z_i - z}{h_k(z)}\right)}{\sum_{i=1}^n K\left(\frac{Z_i - z}{h_k(z)}\right)}. \end{aligned} \qquad [10]$$

With an uniform kernel

$$W_i(z) = \frac{1}{k} 1\left(|Z_i - z| \le h_k(z)\right), \qquad [11]$$

and the nearest neighbours estimate is an average of the k $Y_i' s$ whose corresponding $Z_i' s$ are the k nearest neighbours. of z. That is

$$\hat{m}^N(z) = \frac{1}{k} \sum_{i=1}^{k} Y_{(i,z)}, \qquad [12]$$

where the $Y_{(i,z)}$, $i \ge 1$ are $Y_i' s$ sorted in the way that $\left\{\left(Z_{(i,z)}, Y_{(i,z)}\right), i = 1, ..., n\right\}$ are such that $\left|Z_{(1,z)} - z\right| \le \left|Z_{(2,z)} - z\right| \le ... \le \left|Z_{(n,z)} - z\right|$. The estimator in [12] can be extended in order to provide more weight to those Y_i which corresponding Z_i are closer to z, i.e.

$$\hat{m}^N(z) = \sum_{i=1}^{k} C_i Y_{(i)},$$

with $C_1 \geq C_2 \geq ... \geq C_k > 0$ and $\sum_{i=1}^{k} C_i = 1$. This estimation method is called k-nn. Special k-nn functions are

$$\text{Uniform}: \ C_i = \frac{1}{k}, i = 1, ..., k$$

$$\text{Triangular}: \ C_i = \frac{k-i+1}{k(k+1)/2}$$

$$\text{Quadratic}: \ C_i = \frac{k^2-(i-1)^2}{k(k+1)(4k-1)/6}.$$

The k-nn estimators are well motivated in situations where we have discrete as well as continuous regressors. If we have only discrete regressors we may have ties for the k-th nearest neighbour (i.e. we have several $Z_{(i,z)}$ identical to $Z_{(k,z)}$). In this case one applies a tie breaking rule taking observations randomly among the ties or giving the same weight to all the ties. These weights have been studied by Stone (1977).

2.5. Splines

Not far from the ideas behind the frequency polygon estimators are the splines estimators. This method can be thought as estimating some approximation of the underlying curve (the density or the regression function, for instance) which consists in some 'spline smoothing' of it. To be more precise, let us consider the density estimation problem and fix some increasing 'knots' a_j, $j = 0, 1, \ldots, k+1$, as done for the histogram estimator. One of the most popular method of splines is the degree two histospline, which consists of considering the space $C^{(2)}[a_1, a_k]$, of twice differentiable functions with continuous second derivative in the interval $[a_1, a_k]$ (where the sample falls) and to define f_Z^S as the element of this space that minimizes the functional:

$$\int_{a_1}^{a_k} g''(z)^2 \, dz, \quad \text{for } g \in C^{(2)}[a_1, a_k],$$

subject to the constrains:

$$\int_{I_j} g(z)\ dz = \int_{I_j} f_Z(z)\ dz, \quad \text{for } j = 1, 2, \ldots, k.$$

This is the 'smoothest' function in $C^{(2)}[a_1, a_k]$, of which probability masses to the intervals I_j coincide with the probability masses that the density f_Z gives to them. This restriction is equivalent to the fact that the distribution functions of f_Z and f_Z^S coincide at the knots. The solution of this problem is a so-called degree two histospline. This is a function which is piecewise polynomial of degree two, such that it is also twice differentiable, with continuous second derivative, at the knots a_j (See Schoenberg (1946) for details).

The histospline of degree two estimator, $\hat{f}_Z^S$, is obtained by replacing the underlying distribution by the empirical distribution. In other terms, this estimator is the minimizer of

$$\int_{a_1}^{a_k} g''(z)^2\ dz, \quad \text{for } g \in C^{(2)}[a_1, a_k], \qquad [13]$$

subject to:

$$\int_{I_j} g(z)\ dz = F_n(a_{j+1}) - F_n(a_j), \quad \text{for } j = 1, 2, \ldots, k.$$

This estimator is often called the cubic spline estimator, since it is the derivative of a degree 3 piecewise polynomial, $S(z)$, that interpolates F_n at the knots.

Cubic splines of equispaced knots are probably the most popular choice for splines along the literature, however, it is clear from [13] that the estimator could be extended to any class of splines of arbitrary degree q, as well to cases where the knots are random (chosen according to the data values).

The bias and variance of the cubic spline estimator with fixed knots can be found in Lii and Rosenblatt (1975) and Rosenblatt (1975b). The first reference gives also the asymptotic distribution of the spline estimator. Wahba (1976) deals with the mean squared error of the estimator when the knots are specified by order statistics determined by the sample. As with previous methods, a smoothing parameter comes up here: it is the distance between consecutive knots (when they are fixed and equispaced) or the multiplicative factor, k_n, that defines the indices $k_n, 2k_n, 3k_n, \ldots$, at which the order statistics will be used.

The method of splines has been even more popular in the regression setup. Think of the regression model $Y_i = m(Z_i) + \varepsilon_i$, $i = 1, 2, \ldots n$, where the errors ε_i are $N(0, \sigma^2)$. Then, the log-likelihood function is given by

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - m(Z_i))^2,$$

which is maximized for every function m satisfying $m(Z_i) = Y_i$. A standard tool to get rid off this problem is to add to this function a penalty functional that will account for smoothing properties of the maximizer: the curvature of m. Thus, the penalized log-likelihood function is:

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - m(Z_i))^2 - \alpha \int m''(z)^2 dz, \qquad [14]$$

This still makes sense in the more general case where no parametric structure is assumed for the regression errors.

As in the density case, the function m that maximizes the previous functional, is a spline, more precisely, it is a cubic spline. The smoothing parameter in this estimator is α. This quantity controls the amount of smoothing of the minimizer of [14]. The larger α is, the smoother the estimator is. For instance, when α is extremely large, the penalty term dominates everything in equation [14] and, hence, the minimizer of it will be a function very close to a straight line that fits the data as well as possible. On the other hand, if α is very small, the spline estimate will be very close to any function that interpolates the data. Splines of arbitrary order, q, appear when replacing the curvature in [14] by $\int m^{(q-1)2}$. These splines are just piecewise polynomials of degree $q-1$ with $q-2$ continuous derivatives.

Quite a lot of literature in nonparametric curve estimation has been devoted to the study of the spline estimators. Consistency properties were developed and asymptotic optimality conditions for some criteria to select both the smoothing parameter (α) and the degree of the spline (q), like the generalized cross validation, were proved. See Craven and Wahba (1979) for results of this type or Chapter 5 in the monograph by Eubank (1988), for a detailed study of this estimator. A recent monograph on spline estimators can be found in Wahba (1990).

2.6. Series estimates

The method of orthogonal series can be viewed as estimating some approximation of the true density function, namely, a finite linear combination of terms of a basis of the functional space were the density is assumed to belong. A natural choice for this reference space could be the space of integrable real functions $L_1(\mathbb{R})$, however, typically it will be very useful to make use of the structure of a Hilbert space. This is the reason why the space $L_2(\mathbb{R})$, of squared integrable functions, is frequently chosen.

Let us assume that we are given a complete orthonormal basis in $L_2(\mathbb{R})$: $\{e_j\}_{j=0,1,2,\ldots}$. This means that every function in this space can be represented as a (possibly infinite) linear combination of this basis. In particular, for the density f_Z, it holds:

$$f_Z(z) = \sum_{j=1}^{\infty} c_j e_j(z), \; z \in \mathbb{R}. \qquad [15]$$

Any finite linear combination of the form,

$$f_{Z,M}(z) = \sum_{j=1}^{M} c_j e_j(z) \qquad [16]$$

can be thought as some approximation to the true density function. Except for very special kind of basis, it is not true that $f_{Z,M}$ is a density function.

The coefficients c_j in expression [15] can be computed as

$$c_j = \int f_Z(z) e_j(z) dz = E(e_j(Z)), \; j = 0, 1, 2, \ldots, \qquad [17]$$

where the previous expectation is taken in the random variable Z, assumed to have density function f_Z. These expectations can be estimated by the corresponding sample means,

$$\hat{c}_{n,j} = \frac{1}{n}\sum_{i=1}^{n} e_j(Z_i), j = 0, 1, 2, \ldots \qquad [18]$$

Here, we observe that trying to use expression [15] and estimate the whole representation of f_Z, as an infinite sum, is not a well defined problem. Doing this would end in estimating an infinite number of constants with just a finite number of data values. For this reason, one should restrict to estimating the "smoothed version" of f_Z given in [16].

The orthogonal series estimator is then defined as

$$\hat{f}_{Z,M}(z) = \sum_{j=1}^{M} \hat{c}_{n,j} e_j(z). \qquad [19]$$

Here, the constant M plays the role of the smoothing parameter. In fact, as M tends to infinity, the smoothed version $f_{Z,M}$ tends to f_Z. This suggests the fact that M should depend on n in such a way that $\lim_{n\to\infty} M(n) = \infty$. However, the sequence $M(n)$, should not converge too fast to ∞, otherwise the variance of the estimator would increase dramatically (see Prakasa Rao, 1983). The technical condition that has to be imposed for the variance term to vanish asymptotically is

$$\lim_{n\to\infty} \frac{M(n)}{n} = 0.$$

An important special case is the estimator based on a Fourier expansion of a density with support on a compact interval. Assume, without loss of generality, that the support of f_Z is the unit interval $[0, 1]$. Then, the trigonometric functions,

$$\begin{aligned} e_0(z) &= 1, \\ e_{2r-1}(z) &= \sqrt{2}\cos 2\pi r z, \\ e_{2r}(z) &= \sqrt{2}\sin 2\pi r z, \\ r &= 1, 2, \ldots, \end{aligned} \qquad [20]$$

form an orthonormal basis of $L_2[0, 1]$. A detailed description of this Fourier series estimator may be found in Tarter and Lock (1993).

Some related estimators are those based on the Gram-Charlier series. In this case the standard normal density, $\phi(x)$, is introduced as a weight function in the integral that defines the inner product in L_2. Now, a typical basis of this space, $L_{2,\phi}(\mathbb{R})$, can be constructed using the Hermite polynomials (the j-th polynomial is the factor coming up when differentiating the standard normal density j times). As a consequence, the true density f_Z can be represented as the product of the standard normal density times an infinite sum of terms which are the Hermite polynomials multiplied by some constant depending on the moments of f_Z. These populational moments are estimated by the sample moments and the whole series is reduced to a finite sum, giving the final estimator. These estimates have been applied by Gallant and Nychka (1987) to the maximum likelihood estimation of certain econometric models. They called this method "semi-nonparametric".

The orthogonal series estimator can be extended to the regression setup in two different ways. The first approach is based on the same first steps as in the density case. Assume that the regression function m belongs to $L_2(\mathbb{R})$ and consider a complete orthonormal basis in this space: $\{e_j\}_{j=0,1,2,\ldots}$, then,

$$m(z) = \sum_{j=1}^{\infty} d_j e_j(z), \;\; z \in \mathbb{R}. \qquad [21]$$

The finite linear combination

$$m_M(z) = \sum_{j=1}^{M} d_j e_j(z) \qquad [22]$$

is a kind of smooth version of the function m. The constants d_j can be estimated by least squares, as the minimizers of

$$\sum_{i=1}^{n} (Y_i - m_M(Z_i))^2 = \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{M} d_j e_j(Z_i) \right)^2 .$$

These estimators, $\hat{d}_{n,j}$, are used to derive the final series estimator:

$$\hat{m}_M(z) = \sum_{j=1}^{M} \hat{d}_{n,j} e_j(z). \qquad [23]$$

As in the density case, the trigonometric functions, given in [20], are a popular choice for the basis. In this particular case (for $M = 2k$) the estimator has the form:

$$\hat{m}_M(z) = \hat{d}_{n,0} + \sum_{j=1}^{k} \left(\hat{d}_{n,j}^{(1)} \cos 2\pi jz + \hat{d}_{n,j}^{(2)} \sin 2\pi jz \right). \qquad [24]$$

This estimator is studied in detail in the book by Eubank (1988). This author gives asymptotic expressions for the mean squared error of this estimator as well as asymptotic normality. Rates of convergence can be found in the paper by Eubank and Speckman (1991).

An alternative approach proceeds by finding an orthogonal series estimator of the function $g(z) = m(z)f_Z(z)$ and then divide this estimator by the series estimator of the density function presented above. By means of this method, the regression function can be written in the following terms,

$$m(z) = \frac{\sum_{j=1}^{\infty} d_j e_j(z)}{\sum_{j=1}^{\infty} c_j e_j(z)},$$

where the coefficients c_j, defined as in [17], can be estimated as in [18] and the values d_j are given by

$$d_j = \int g(z)e_j(z)dz = E(m(Z)e_j(Z)) = E(Ye_j(Z)), \; j = 0, 1, 2, \ldots.$$

These expectations can be estimated using empirical averages,

$$\hat{d}_{n,j} = \frac{1}{n} \sum_{i=1}^{n} Y_i e_j(Z_i), j = 0, 1, 2, \ldots$$

leading to the final expression of the estimator,

$$\hat{m}_M(z) = \frac{\sum_{j=1}^{M} d_{n,j} e_j(z)}{\sum_{j=1}^{M} c_{n,j} e_j(z)}.$$

The Fourier and Hermite choice for the orthonormal basis in this estimator were studied by Greblicki and Pawlack (1985). These authors prove consistency properties and rates of convergence for this type of series regression estimator.

Series estimators of regression curves have been also applied in different semiparametric problems, e.g. Newey (1990) and Donald and Newey (1994) and in specification testing by Hong and White (1995).

3. Estimating other relevant nonparametric functions

3.1. Multivariate densities and regression with multivariate explanatory variables

When Z is a $\mathbb{R}^p$-valued vector of random variables, the multivariate density, $f_Z(\cdot)$, is estimated using kernels by

$$\hat{f}_Z(z) = \frac{1}{nh^p} \sum_{i=1}^{n} K_p\left((z - Z_i)/h\right), \qquad [25]$$

where $K_p(u) = \prod_{k=1}^{p} K(u_p)$, $u = (u_1, u_2, ..., u_p)'$ and $K(\cdot)$ is an univariate kernel. One could use other non multiplicative kernels (e.g. any multivariate distribution), and instead of using the same bandwidth h for all the components of Z, one could use a matrix in order to take into account the possible correlation structure between the components of Z (see Robinson 1983). That is, alternatively we could use

$$\hat{f}_Z(z) = \frac{1}{n\left\|H\right\|^{1/2}} \sum_{i=1}^{n} K_p\left(H^{-1/2}(z - Z_i)\right),$$

where H is a positive definite matrix and $\|H\|$ denotes the determinant of H. For notational convenience, we only consider [25]. Also $r_Z(z) = \int_{\mathbb{R}^p} y f_{YZ}(y, z)\, dz$ is estimated by

$$\hat{P}_Z(z) = \frac{1}{nh^p} \sum_{i=1}^{n} Y_i K_p\left((z - Z_i)/h\right). \qquad [26]$$

The resulting estimator of $m(z) = r_Z(z)/f_Z(z)$ is

$$\hat{m}(z) = \frac{\sum_{i=1}^{n} Y_i K_p\left((z - Z_i)/h\right)}{\sum_{i=1}^{n} K_p\left((z - Z_i)/h\right)}.$$

3.2. Conditional distributions

The conditional density of Y given Z, $f_{Y|Z}(y \mid z) = f_{Y,Z}(y,z)/f_Z(z)$, is estimated by

$$\hat{f}_{Y|Z}(y \mid z) = \hat{f}_{Y,Z}(y,z)/\hat{f}_Z(z). \qquad [27]$$

Then, there are two ways of estimating the conditional distribution of Y given Z, $F_{Y|Z}(y \mid z) = \int_{-\infty}^{y} f_{Y|Z}(u \mid z)\,du$. First, we can use [27], the resulting estimator is

$$\hat{F}_{Y|Z}(y \mid z) = \int_{-\infty}^{y} \hat{f}_{Y|Z}(u \mid z)\,du.$$

Using kernels,

$$\hat{f}_{Y|Z}(u \mid z) = \frac{\sum_{i=1}^{n} K((u-Y_i)/h)\,K_p((z-Z_i)/h)}{h\sum_{i=1}^{n} K_p((z-Z_i)/h)}.$$

Hence,

$$\hat{F}_{Y|Z}(y \mid z) = \frac{\sum_{i=1}^{n} \mathbb{K}((y-Y_i)/h)\,K_p((z-Z_i)/h)}{\sum_{i=1}^{n} K_p((z-Z_i)/h)}, \qquad [28]$$

where $\mathbb{K}(u) = \int_{-\infty}^{u} K(u)\,du$. The fact that $F_{Y|Z}(y \mid z) = E[1(Y \leq y) \mid Z = z]$, suggests to estimate the conditional distribution as a regression function,

$$\tilde{F}_{Y|Z}(y \mid z) = \frac{\sum_{i=1}^{n} 1(Y_i \leq y)\,K_p((z-Z_i)/h)}{\sum_{i=1}^{n} K_p((z-Z_i)/h)}. \qquad [29]$$

Both, [28] and [29] yield the same regression estimate, since

$$\begin{aligned} \hat{m}(z) &= \int_{\mathbb{R}} y\hat{F}_{Y|Z}(dy \mid z) = \\ &= \frac{\sum_{i=1}^{n} K_p((z-Z_i)/h)\int_{\mathbb{R}} yK((y-Y_i)/h)\,dy}{\sum_{i=1}^{n} K_p((z-Z_i)/h)} \\ &\simeq \frac{\sum_{i=1}^{n} Y_i K_p((z-Z_i)/h)}{\sum_{i=1}^{n} K_p((z-Z_i)/h)} = \int_{\mathbb{R}} y\tilde{F}_{Y|Z}(dy \mid z). \end{aligned}$$

The distribution function estimate is specially useful for studying the specification of discrete choice models like probit or logit.

Economic cross-section data sometime present a lot of zero responses. A typical example is when estimating consumption curves of goods like alcohol, tobacco, cloth or recreation. The conditional density function, in this case, is a mixture of a continuous conditional density and a discrete conditional point of mass at zero. The situation could be represented by means of the following Tobit model, $Y = \max\{m(Z) + \varepsilon, 0\}$, where $m(\cdot)$ is an unknown function and ε is an error term. The conditional distribution of Y given Z will be estimated in this case by the mixture distribution with continuous part and discrete part given by

$$\begin{aligned}\tilde{f}_{Y|Z}(y \mid z) &= \left[1 - \tilde{F}_{Y|Z}(0 \mid z)\right]^{-1} \hat{f}_{Y|Z}(y \mid z) \text{ for } y > 0 \\ \tilde{p}_{Y|Z}(y \mid z) &= \tilde{F}_{Y|Z}(0 \mid z) \text{ for } y = 0,\end{aligned}$$

where now,

$$\hat{f}_{Y|Z}(y \mid z) = \frac{\sum_{i=1}^{n} 1(Y_i > 0) K((y - Y_i)/h) K_p((z - Z_i)/h)}{\sum_{i=1}^{n} K_p((z - Z_i)/h) 1(Y_i > 0)}.$$

3.3. Hazard rate

A relevant function in survival analysis and reliability is the hazard rate function. Given the random variable of interest, Z, with distribution function F_Z, the hazard rate at a given value z is defined by

$$r_Z(z) = \lim_{\epsilon \to 0^+} \frac{P(Z \le z + \epsilon | Z > z)}{\epsilon}. \qquad [30]$$

When the variable Z measures the life time of a component (or a living being), the hazard rate function at time z represents the instantaneous rate of the probability that this component breaks down immediately after time z. The larger the hazard rate is, at a given time z, the more probable that the failure occurs immediately after that time. If there exists a density function of Z, f_Z, straight forward calculations starting from expression [30] show that the hazard rate admits the following representation:

$$r_Z(z) = \frac{f_Z(z)}{1 - F_Z(z)}.$$

Obvious estimators of this function come up just by combining one estimator for the density (for instance the kernel estimator) and another for the survival function $1 - F_Z(z)$ (by using the empirical distribution function, for example). According to this, given a random sample $Z_1, Z_2, \ldots, Z_n$, one could estimate the hazard rate by means of

$$\hat{r}(z) = \frac{\hat{f}(z)}{1 - F_n(z)} = \frac{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{z-Z_i}{h}\right)}{\frac{1}{n}\sum_{i=1}^{n} 1(Z_i > z)}.$$

Another possibility consists in incorporating the denominator in the additive structure of the kernel density estimator, ending up with a method where the survival function is not directly estimated but introduced into the smoothing mechanism. This is the Watson and Leadbetter (1964a, b) estimator:

$$\hat{r}_{WL}(z) = \frac{1}{nh}\sum_{i=1}^{n} \frac{K\left(\frac{z-Z_i}{h}\right)}{1 - F_n(Z_i)}.$$

This estimator is not well defined at the largest data value. For this reason, it may be replaced by

$$\frac{1}{nh}\sum_{i=1}^{n} \frac{K\left(\frac{z-Z_i}{h}\right)}{1 - \frac{n}{n+1}F_n(Z_i)}.$$

3.4. Conditional quantiles and other conditional location functionals.

From the above estimator of the conditional distribution, the conditional quantile $Q_\theta(Y \mid Z = z) = \inf\left\{y : F_{Y|Z}\,(y \mid z) \geq \theta\right\}$ is estimated by

$$\hat{Q}_\theta(Y \mid Z = z) = \inf\left\{y : \hat{F}_{Y|Z}\,(y \mid z) \geq \theta\right\}.$$

This conditional quantile estimate is a reasonable alternative to the regression estimates when the dependent variable has many zeros, as it has been motivated by Powell (1984, 1986) in a parametric context. Indeed, if $Y = \max(m(Z) + \varepsilon, 0)$, where ε has conditional median zero, $m(\cdot)$ is estimated by the conditional median, and, under assumptions on ε, conditional quantiles provide parallel translations of the conditional median. Also, a number of conditional quantiles provide a parsimonious information on the conditional distribution much more informative than a mere conditional location estimate, like the conditional mean or median.

These conditional quantiles have also a M-estimation representation. In fact,

$$\begin{aligned} Q_\theta(Y \mid Z = z) &= \arg\min_q E\{[\theta \cdot 1(Y - q < 0) + \\ &\quad + (1-\theta) \cdot 1(Y - q \geq 0)] \, |Y - q| \mid Z = z\}, \end{aligned}$$

or, alternatively, as solving the equation

$$E\{1(Y - Q_\theta(Y \mid Z = z) < 0) \mid Z = z\} = \theta.$$

Then, the quantile estimators can be defined as

$$\begin{aligned} \hat{Q}_\theta(Y \mid Z = z) &= \arg\min_q \sum_{i=1}^n \{[\theta \cdot 1(Y_i - q < 0) + \\ &\quad + (1-\theta) \cdot 1(Y_i - q \geq 0)]\} \, |Y_i - q| \, W_i(z), \end{aligned}$$

or

$$\sum_{i=1}^n 1\left(Y_i - \hat{Q}_\theta(Y \mid Z = z) < 0\right) W_i(z) = \theta,$$

where $W_i(z)$ are nonparametric weights.

In general, a location functional implicitly defined as

$$E[\psi(Y - m(z)) \mid Z = z] = 0,$$

is estimated by $\hat{m}(z)$, defined as the solution to

$$\sum_{i=1}^n \psi(Y_i - \hat{m}(z)) \, W_i(z) = 0.$$

Note that when $\psi(u) = u$, $\hat{m}(z)$ is the usual kernel estimate.

3.5. Derivatives of regression curves

For simplicity, suppose that Z is scalar. In general $m^{(r)}(z) \equiv \partial^r m(z)/\partial z^r$ is estimated by $\hat{m}^{(r)}(z) = \partial^r \hat{m}(z)/\partial z^r$. Also, applying a Taylor expansion,

$$m(u) \simeq$$

$$m(z)+m^{(1)}(z)(u-z)+\frac{1}{2}m^{(2)}(z)(u-z)^2+....+\frac{1}{r!}m^{(r)}(z)(u-z)^r. \quad [31]$$

Define $\beta_0 = (\beta_{00}, \beta_{01}, ..., \beta_{0r})'$, where $\beta_{0k} = m^{(k)}(z)$, $\beta_{00} = m^{(0)}(z)$ and x' denotes the transpose a matrix x, then [31] suggests the estimator

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left\{ Y_i - \sum_{k=1}^{r} \beta_k (Z_i - z)^k \right\}^2 K\left(\frac{Z_i - z}{h}\right).$$

Hence,

$$\hat{\beta} = [\mathbb{X}'\mathbb{W}\mathbb{X}]^{-1}\mathbb{X}'\mathbb{W}\mathbb{Y},$$

where

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \ \mathbb{X} = \begin{pmatrix} 1 & (Z_1 - z) & \cdots & (Z_1 - z)^r \\ 1 & (Z_2 - z) & \cdots & (Z_2 - z)^r \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (Z_n - z) & \cdots & (Z_n - z)^r \end{pmatrix},$$

$$\mathbb{W} = \begin{pmatrix} K\left(\frac{Z_1 - z}{h}\right) & 0 & \cdots & 0 \\ 0 & K\left(\frac{Z_2 - z}{h}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K\left(\frac{Z_n - z}{h}\right) \end{pmatrix}.$$

Then, $\hat{\beta}_k$ is the estimate of $m^{(k)}(z)$. It is easy to realize that $\hat{\beta}_0$ estimates $m(z)$.

4. Choice of the smoothing parameter

An important issue in nonparametric curve estimation is the problem of selection of the smoothing parameter. All the estimators presented above depend on such a quantity. Typically, the bias of the nonparametric estimators increases with this smoothing parameter while the variance decreases as this parameter gets large. Most of the solutions to this problem along the literature consist in defining some (unobservable) measure of performance of the estimator and then try to estimate the value of the smoothing parameter which is optimal with respect to this measure. Since this value is also unobservable, statistical procedures have to be designed to estimate it.

For clarity and conciseness we will concentrate on the kernel estimator in the context of nonparametric density estimation. However, most of the ideas that will be presented in this section can be applied to the curves in the previous sections.

4.1. The ISE and MISE criteria

Let us consider a random sample $Z_1, Z_2, \ldots, Z_n$ coming from a population with cumulative distribution function F_Z (F, to avoid unnecessary subindex in this section) and density f_Z (or simply, f). A popular measure of performance of the kernel density estimator:

$$\hat{f}_h(z) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{h}\right)$$

is the integrated squared error criterion (ISE):

$$ISE = \int \left(\hat{f}_h(z) - f(z)\right)^2 \, dz, \qquad [32]$$

which is nothing else but the square of the L_2-distance between the estimator and the underlying density. One of the sharpest problems of considering this criterion as a measure of discrepancy is that it is random (since it depends on the sample). A simple (and very popular) way to get rid off this randomness consists in using the expected value of ISE, namely the mean integrated squared error (MISE), given by

$$MISE = E\left(\int \left(\hat{f}_h(z) - f(z)\right)^2 \, dz\right). \qquad [33]$$

Some properties of asymptotic equivalence between these two criteria indicate that both methods should be very close for large sample sizes.

Of course, the choice of the L_2-distance in the previous criteria is not the only possibility. The L_1 distance has also been used in the literature (see the monograph by Devroye and Györfi, 1985) with also promising results concerning the bandwidth selection problem (see Devroye, 1989).

Simple calculations can be carried out in expression [33] in order to decompose the MISE into two terms: one coming from the integrated squared bias and the other incorporating the integrated variance,

$$MISE(h) = B(h) + V(h),$$

where

$$B(h) = \int \left(E\left(\hat{f}_h(z)\right) - f(z)\right)^2 \, dz$$

and

$$V(h) = \int Var\left(\hat{f}_h(z)\right) \, dz.$$

To derive asymptotic properties in this context, one has to control the amount of smoothing in order that it vanishes asymptotically ($h \rightarrow 0$) when the sample size gets large ($n \rightarrow \infty$), otherwise we would estimate average quantities in a certain neighbourhood, instead of local quantities. Under this assumption, standard expectation calculations and Taylor expansions lead to some asymptotic representations,

$$B(h) = \frac{1}{4} R(f'') d_K h^4 + o(h^4)$$

and

$$V(h) = \frac{1}{nh} R(K) + O(\frac{1}{n}),$$

where $R(g) = \int g(z)^2 \, dz$ represents the square of the L_2-norm of a function g and $d_K = \int z^2 K(z) dz$.

It turns out that the integrated variance will tend to zero whenever $nh \rightarrow \infty$ as $n \rightarrow \infty$. These expressions result in an asymptotic representation for the MISE,

$$MISE(h) = AMISE(h) + o(h^4) + O(n^{-1}),$$

where

$$AMISE(h) = \frac{R(f'')d_K h^4}{4} + \frac{R(K)}{nh}. \qquad [34]$$

An attempt to solve the bandwidth selection problem could be to obtain the value that minimizes expression [34],

$$h_{AMISE} = \left(\frac{R(K)}{nR(f'')d_K}\right)^{1/5},$$

that decreases to zero at the rate $n^{-1/5}$. However, the value h_{AMISE} is not observable since it depends on $R(f'')$, the curvature of the underlying density.

4.2. Plug-in selectors

A possible way to proceed with h_{AMISE} is to estimate the curvature of f by using a preliminary kernel estimator of the density with some pilot bandwidth g. This means to replace the term $R(f'')$ by $R(\hat{f}''_g)$ in the formula for this asymptotically optimal bandwidth. The general expression for the plug-in selector is

$$h_{PI} = \left(\frac{R(K)}{nR(\widehat{f''})d_K}\right)^{1/5},$$

where $R(\widehat{f''})$ may be the estimator $R(\hat{f}''_g)$ or even a more sophisticated term (see Sheather and Jones, 1991). At some point, a subjective choice of a pilot bandwidth has to be done. For instance, one may find the best bandwidth, g, in the sense of the mean squared error of $R(\hat{f}''_g)$ as an estimator of $R(f'')$. Unfortunately, this bandwidth depends on the constant $R(f''')$ and we are in a never ending process. A common solution to this problem is to assume an arbitrary parametric structure (for instance, normal) at some stage of this process and compute the unknown constant, depending on f, in terms of the parameters of the family. These parameters are estimated and a final value for the pilot bandwidth is found. This may be even done at the beginning just by estimating the curvature in a parametric way.

4.3. Least squares cross-validation

This was one of the most popular methods in the early days of bandwidth selection. Nowadays it is known to exhibit a poor rate of convergence to the optimal bandwidth, h_{MISE}, that minimizes the MISE. The idea behind this method, proposed by Rudemo (1982) and Bowman (1984), is just to decompose the exact expression of ISE into three terms,

$$ISE(h) = \int \hat{f}_h(z)^2 \, dz + \int f(z)^2 \, dz - 2 \int \hat{f}_h(z) f(z) \, dz.$$

Since the term $\int f^2$ does not depend on h, it is not relevant into the minimization procedure. On the other hand, the first term is observable, while the cross product term, $\int \hat{f}_h(z) F(dz)$, may be estimated by

$$\int \hat{f}_h(z) F_n(dz) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}_h(Z_i),$$

or, even better, by a leaving-one-out unbiased estimator

$$\frac{1}{n} \sum_{i=1}^{n} \hat{f}_h^i(Z_i),$$

where

$$\hat{f}_h^i(z) = \frac{1}{(n-1)h} \sum_{j=1, j \neq i}^{n} K\left(\frac{z - Z_j}{h}\right).$$

The bandwidth, h_{CV}, that minimizes the resulting expression:

$$CV(h) = \int \hat{f}_h(z)^2 \, dz - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_h^i(Z_i)$$

is a natural ISE-oriented selector. The main problem of this selector is its large variability. This was partially solved by the biased cross-validation method, proposed by Scott and Terrell (1987). However, the slow convergence to the optimum is still present in this version.

4.4. *Smoothed cross-validation*

This method consists in finding the minimum of a new estimator, namely $SCV(h)$, of the MISE criterion. This function incorporates the first order term of the integrated variance $R(K)/nh$, that is already known, and uses a smooth estimator of the exact formula for the integrated squared bias:

$$\hat{B}(h) = \int \left(\int K(u) \hat{f}_g(z - hu) du - \hat{f}_g(z) \right)^2 dz,$$

where g is a pilot bandwidth, that plays the same role as in the plug-in selectors. The smoothed cross-validation selector, h_{SCV}, is defined as the minimizer of

$$SCV(h) = \frac{R(K)}{nh} + \hat{B}(h).$$

The method was proposed by Hall, Marron and Park (1992) and the relative rate of convergence of h_{SCV} has been studied by Jones, Marron and Park (1991). These authors prove that this rate is the best attainable: $n^{-1/2}$.

4.5. *Bootstrap bandwidth selectors*

The bootstrap method goes back to Efron (1979) and can be summarized as follows. Given a random variable $R = R(Z_1, Z_2, ..., Z_n; F)$ depending on the true distribution function, F, of a given population and on a random sample $Z_1, Z_2, ..., Z_n$ from that population, replace the distribution function F by some estimator of it, $\hat{F}$, and the random sample by a random resample $Z_1^*, Z_2^*, ..., Z_n^*$ drawn from $\hat{F}$. The resampling distribution of $R^* = R\left(Z_1^*, Z_2^*, ..., Z_n^*; \hat{F}\right)$ is used to approximate the sampling distribution of R.

In the context of bandwidth selection, the bootstrap is used to obtain different types of selectors by minimizing some kind of bootstrap estimate of the MISE. The differences among them come from the type of resampling used in the bootstrap approach: resampling from the empirical (see Hall, 1990) or smoothed resampling, either with pilot bandwidth (see Cao, 1993) or without it (see Taylor, 1989).

Among all these possibilities, the smoothed bootstrap with pilot bandwidth, proposed by Cao (1993), seems to be the best choice. The method consists in using a bootstrap mechanism to draw resamples from a kernel density estimator with a pilot bandwidth $\hat{f}_g$. Since the MISE is just an expectation, it can be estimated by means of a bootstrap expectation that, of course, will be observable. In other words, we define

$$MISE^*(h) = E^*\left(\int \left(\hat{f}_h^*(z) - \hat{f}_g(z)\right)^2 \, dz\right),$$

where $\hat{f}_h^*(z)$ is the kernel density estimator based on the bootstrap resample.

Simple calculations of the bootstrap bias and variance show that the bootstrap version of MISE can be computed directly, as a function of the sample, without the need of any Montecarlo. In fact, it may be shown that,

$$MISE^*(h) = V^*(h) + \hat{B}(h),$$

where $\hat{B}(h)$ is the same term that appeared in the smoothed cross-validation criterion and $V^*(h)$ is an estimator of the whole integrated variance:

$$V^*(h) = \frac{R(K)}{nh} - \int \left(\int K(u)\hat{f}_g(z-hu)du\right)^2 \, dz.$$

Now, the bootstrap bandwidth selector, h^*, is defined as the minimizer of $MISE^*(h)$.

The asymptotic formula for the optimal pilot bandwidth, in the sense of the mean squared error of the curvature estimator, may be found to be

$$g_0 = \left(\frac{R(K'')}{nd_K R(f''')}\right)^{1/7}.$$

Since this value depends on the underlying density, some preliminary estimation steps are needed. This will end with a parametric assumption, as commented above for the plug-in selectors.

5. An application to the Spanish expenditure survey

Nonparametric estimation techniques are available in many computers packages like GAUSS, MATLAB, AXUM or S^+. XPLORE is a computer package specialized in nonparametric estimation (see Härdle et al, 1995). Nonparametric estimation offers a powerful tool in specification, providing valuable hints on the underlying form of the function to be estimated. We provide some examples which illustrate the use of nonparametric estimation in practice.

Consider the Spanish FES for 1973, 1980 and 1990 on household consumption habits over more than 20,000 households. All the data is translated to 1990 prices. During this period different income distribution policies have been developed. The income distribution is usually parameterized by means of a lognormal. In Figure 1, we present maximum likelihood estimates based on the lognormal specification as well as kernel estimates based on different automatic bandwidth choice criteria, i.e. cross validation and a bootstrap method. As usual, income is proxied by the total expenditure rather than reported net income.

Figure 1 shows that the nonparametric estimators behave similarly with the different bandwidth selectors discussed in Section 5. It is observed that the mode is different for the lognormal specification indicating that this model is not suitable for representing the income family distribution. The behavior of the income distribution during the three examined periods is interesting. The same exercise was performed by Hildenbrand and Hildenbrand (1986) and Härdle et al (1991) (see also Härdle 1990, p. 6-9 for a discussion of these applications) using the UK FES in the period 1968 to 1983. They find a clear bimodality in the income distribution of each period as well as a clear change in shape, Härdle (1990, p 7) observes that '...more people enter the 'lower income range' and the 'middle class' peak becomes less dominant'. Our estimated income distributions are clearly unimodal and their shapes change in an opposite way than in the UK case: more people leave the 'lower income range' and the mode of the distribution moves to the right, that is, to higher income levels.

Next we have estimated an Engel curve relating mean expenditure on food and recreation with total expenditure for the three surveys. In

FIGURE 1
Density Estimates of Income Based on the Spanish Expenditure Survey

order to avoid heterogeneity, we select families with two children and the head of the household employed. The bandwidth is chosen via an iterative plug-in from pilot estimates of the densities, regression and derivatives. Figure 2 presents the estimates of Engel curves for food for the three surveys.

It is observed that expenditure on food has decreased from 1973 to 1990. This fact can be due to a decrease of food relative prices as well as changes in total income. Figures 2 and 3 suggest that the Working Leser model $E[Y \mid X = x] = \beta_0 + \beta_1 \log x$ is a reasonable approximation in practice, where Y is expenditure on a good, X is total income and β_0 and β_1 are given parameters.

Figure 3 provides kernel estimates of the conditional median of food

FIGURE 2
Nonparametric Food Engel Curves Estimates Based on the Spanish Expenditure Survey

expenditure given total expenditure. These estimates were discussed in Section 3. The conditional median, $q_{0.5}(x) = Q_{0.5}(Y \mid X = x)$, (i.e. the 0.5. conditional quantile of Y given X) is computed as,

$$\hat{q}_{0.5}(x) = \frac{\left(\hat{q}^{L}_{0.5}(x) + \hat{q}^{U}_{0.5}(x)\right)}{2},$$

where

$$\begin{aligned} \hat{q}^{L}_{0.5}(x) &= \max_{1 \le i \le n} \left\{ Y_i \mid \sum_{j=1}^{n} 1(Y_j \le Y_i) W_j(x) \le \frac{1}{2} \right\} \\ \hat{q}^{U}_{0.5}(x) &= \min_{1 \le i \le n} \left\{ Y_i \mid \sum_{j=1}^{n} 1(Y_j \le Y_i) W_j(x) > \frac{1}{2} \right\}. \end{aligned}$$

The conditional median has also decreased form 1973 to 1990. In this case we are using the same bandwidth as in Figure 2, an there is possibly some oversmoothing.

FIGURE 3

Conditional Median Estimates of Expenditure on Food Given Total Income Based on the Spanish Expenditure Survey

The next Figure represents Engel curve estimates of recreation goods. Apparently, the expenditure on recreation, given total income, is greater in 1980 than in previous years for incomes lower than 3.25 millions. However, there is not a clear relationship for higher incomes. That is, there is not a systematic pattern for the Engel curve behaviour for the different years. This fact, could be explained by the heterogeneous behaviour of individuals, depending on their income levels, with respect to consumption on recreation goods. The presence of many zeros in the dependent variable can also explain this erratic behaviour. So, it seems reasonable to look at regression quantiles as it has been done by Koenker and Bassett (1982) for studying the behaviour of Engel curves, and by Chamberlain (1994) and Buckinsky (1994) in the context of estimating wage equations.

Figure 5 provides estimates of conditional medians of recreation expenditure with respect to total income. For any income level, the median expenditure in recreation increases with respect to time, just the opposite behaviour than the expenditure on food.

FIGURE 4
Nonparametric Recreation Engel Curves Based on the Spanish Expenditure Survey

Finally, we have estimated the probability of non expenditure on tobacco (non smoking) given total expenditure using a probit model and the kernel method. The conditional probability $P(x) = \Pr(Y = 0 \mid X = x)$ is estimated nonparametrically by $\hat{P}(x) = \sum_{j=1}^{n} 1(Y_j = 0)\, W_j(x)$. In Figure 6 we represent the probit (discontinuous lines) and the nonparametric estimates (continuous lines). We observe that the probability of smoking has increased, at constant 1990 prices, for all income levels. As total expenditure increases, the probability of smoking decreases, except at the higher income levels. The probit estimates are, in the tails, quite different from the nonparametric ones. This fact is probably due to boundary bias and low number of observations at the tails. Once, one correct for these factors, it is expected that the parametric and nonparametric curves will not differ too much.

6. Final Comments

In this paper we have offered an overview to nonparametric esti-

Figure 5
Conditional Median Estimates of Expenditure on Recreation Given Total Income Based on the Spanish Expenditure Survey

mation, illustrating the discussion with some applications using the Spanish FES.

Nonparametric estimators are useful in model specification, but they are hard to compare with parametric estimators based on a correctly specified model. A nonparametric estimator always converges at a slower rate than an estimator based on a correctly specified model (usually $n^{-1/2}$). The rate of converge of nonparametric estimators becomes slower, at an exponential rate, as the nonparametric function depends on an increasing number of variables, problem known as 'curse of dimensionality'. When the curve to be estimated depends on more than two variables, we need enormous sample sizes in order to obtain trustable nonparametric estimates.

The nonparametric estimators have the advantage of providing an approximation to the true underlying curve, leaving the data 'speak for themselves'. These preliminary estimates can be useful in order to find a suitable parametric function. When a parametric form is imposed to the curve to be estimated, we cannot be sure on the

Figure 6
Conditional Probability Estimates of Non Tobacco Consumption given Income Based on the Spanish Expenditure Survey

meaning of the parameters and the estimated curve can be a very poor approximation to the true one.

References

Bean, S.J. and C.P. Tsokos (1980): "Developments in nonparametric density estimation", *International Statistical Reviews* 48, pp. 267-287.

Bierens, H.J. and H.A. Pott-Butter (1991): "Specification of household Engel curves by nonparametric regression", *Econometric Reviews* 9, pp. 123-184.

Bowman, A.W. (1984) : "An alternative method of cross-validation for the smoothing of density estimates", *Biometrika* 71, pp. 353-360.

Buchinsky, M. (1994): "Changes in the U.S. wage structure 1963-1967: Application of quantile regression", *Econometrica* 62, pp. 405-458.

Cao, R. (1993): "Bootstrapping the mean integrated squared error", *Journal of Multivariate Analysis* 45, pp. 137-160.

Chamberlain, G. (1994): "Quantile regression, censoring, and the structure of wages", in C.A. Sims (ed.), *Advances in Econometrics, Sixth World*

Congress vol 1. Cambridge University Press.

Craven, P. and G. Wahba (1979): “Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation”, *Numer. Math.* 31, pp. 377-403.

Collomb, G. (1981): “Estimation nonparamétrique de la régresion: review bibliographique”, *International Statistical Review* 49, pp. 75-93.

Collomb, G. (1985): “Non-parametric regression: an up-to-date bibliography”, *Statistics* 16, pp. 309-324.

Delgado, M.A. and D. Miles (1997): “Household characteristics and consumption behaviour: a nonparametric approach”, *Empirical Economics* forthcoming.

Delgado, M.A. and P.M. Robinson (1992): “Nonparametric and semiparametric methods for economic research”, *Journal of Economic Surveys* 6, pp. 201-249.

Delgado, M.A. and T. Stengos (1994): “Semiparametric testing of non-nested econometric models”, *Review of Economic Studies* 75, pp. 345-367.

Devroye, L. (1987): “A Course in Density Estimation”, Birkhauser.

Devroye, L. (1989): “The double kernel method in density estimation”, *Annals of the Institute of Henri Poincaré* 25, pp. 533-580.

Devroye, L. and . Györfi (1985): “Nonparametric Density Estimation: The L_1-view”, *Wiley* .

Diebold, F.X. and J.A. Nason (1990): “Nonparametric exchange rate prediction?”, *Journal of International Economics* 28, pp. 315-332.

Donald, S.G. and W.K. Newey (1994): “Series estimators of semilinear models”, 50, pp. 30-40.

Efron, B. (1979): “Bootstrap methods: another look at the jackknife”, *Annals of Statistics* 7, pp. 1-26.

Eubank, R.L. (1988): “Spline Smoothing and Nonparametric Regression”, *Marcel Dekker* 86, pp. 112-132.

Eubank, R.L. and P. Speckman (1991): “Convergence rates for trigonometric and polynomial-trigonometric regression”, *Statistics and Probability Letters* 11, pp. 119-124.

Eubank, R.L. and C.H. Spiegelman (1990): “Testing the goodness of fit in regression via order selection criteria”, *Journal of the American statistical Association* 85, pp. 387-392.

Fan, Y. and Q. Li (1996): "Consistent model specification tests: omitted variables and semiparametric functional forms", *Econometrica* 64, pp. 865-890.

Fryer, M.J. (1977): "A review of some nonparametric methods of density estimation", *Journal of the Institute of Mathematics and Applications* 20, pp. 335-354.

Gallant, A.R. and D.W. Nychka (1987): "Semi-nonparametric maximum likelihood estimation", *Econometrica* 55, pp. 363-390.

Greblicki, W. and Greblicki M. Pawlak (1985): "Fourier and Hermite series estimates of regression functions", *Annals of the Institute of Statistical Mathematics* 37, pp. 443-454.

Green, P.J. and B.W. Silverman (1994): "Nonparametric Regression and Generalized Linear Models", *Chapman and Hall.*

Haavelmo, T. (1944): "The probability approach in econometrics", *Econometrica* 12, Monograph.

Hall, P. (1990): "Using the bootstrap to estimate mean square error and select smoothing parameters in nonparametric problems", *Journal of Multivariate Analysis* 32, pp. 177-203.

Hall, P., J.S. Marron and B. Park (1992): "Smoothed cross-validation", *Probability Theory and Related Fields* 92, pp. 1-20.

Härdle, W. (1990): "Applied Nonparametric Regression", *Cambridge University Press.*

Härdle, W., W. Hildenbrand and M. Jerison (1991): "Empirical evidence on the law of demand", *Econometrica* 59, pp. 1525-1549.

Härdle, W., S. Klinke and B.A. Turlach (1995): "XploRe: an interactive statistical computing environment", Springer-Verlag.

Härdle, W. and E. Mammen (1993): "Comparing nonparametric versus parametric regression fits", *Annals of Statistics* 21, pp. 1926-1947.

Hildenbrand, K. and W. Hildenbrand (1986): "On the mean income effect: a data analysis of the UK Family Expenditure Survey", In W. Hildenbrand and A. Mas-Colell (eds) *Contributions to Mathematical Economics* , North-Holland.

Hjort, N.L. and M.C. Jones (1996): "Locally parametric nonparametric density estimation", *Annals of Statistics* 24, pp. 1619-1647.

Hong, Y. and H. White (1995): "Consistent specification testing via nonparametric series regression", *Econometrica* 63, pp. 1133-1159.

Izenman, A.J. (1991): "Recent developments in nonparametric estimation", *Journal of the American Statistical Association* 86, pp. 205-224.

Jones, M.C., J.S. Marron and B. Park (1991): “A simple root n bandwidth selector”, *The Annals of Statistics* 19, pp. 1919-1932.

Koenker, R. and G. Bassett (1982): “Robust tests for heteroskedasticity based on regression quantiles”, *Econometrica* 50, pp. 43-61.

Kozek, A.S. (1991): “A non-parametric test of fit of a parametric model”, *Journal of Multivariate Analysis* 37, pp. 66-75.

Lavergne, P. and Q. Vuong (1996): “Nonparametric selection of regressors: the nonnested case”, *Econometrica* 64, pp. 207-219.

Leonard, T. (1978): “Density estimation, stochastic processes and prior information”, *Journal of the Royal Statistical Society, Series B* 40, pp. 113-146.

Lii, K.S. and M. Rosenblatt (1975): “Asymptotic behavior of a spline estimate of a density function”, *Comput. Math. Appl.* 1, pp. 223-235.

Loader, C.R. (1996): “Locally likelihood density estimation”, *Annals of Statistics* 24, pp. 1602-1618.

Nadaraya, E.A. (1964): “On estimating regression”, *Theory of Probability and its Applications* 9, pp. 141-142.

Newey, W.K. (1990): “Efficient instrumental variable estimation of nonlinear models”, *Econometrica* 58, pp. 809-837.

Pagan, A.R. and G.W. Schwert (1990): “Alternative models for conditional stock volatility”, *Journal of Econometrics* 45, pp. 267-290.

Parzen, E. (1962): “On estimation of a probability density function and mode”, *The Annals of Mathematical Statistics* 33, pp. 1065-1076.

Prakasa Rao, B.L.S. (1983): “Nonparametric Functional Estimation”, *Academic Press* .

Powell, J. (1984): “Least absolute deviations estimation for the censored regression model”, *Journal of Econometrics* 25, pp. 303-325.

Powell, J. (1986): “Censored regression quantiles”, *Journal of Econometrics* 32, pp. 143-155.

Robinson, P.M. (1983): “Nonparametric estimators for time series”, *Journal of Time Series Analysis* 4, pp. 185-207.

Robinson, P.M. (1991): “Consistent nonparametric entropy-based testing”, *Review of Economic Studies* 58, pp. 437-453.

Rosenblatt, M. (1956): “Remarks on some nonparametric estimators of a density function”, *The Annals of Mathematical Statistics* 27, pp. 832-837.

Rosenblatt, M. (1971): “Curve estimates”, *Annals of Mathematical Statistics* 42, pp. 1815-1842.

Rosenblatt, M. (1975a): “A quadratic measure of deviation of two-dimensional density estimates and a test of independence”, *Annals of Statistics* 1, pp. 1-15.

Rosenblatt, M. (1975b): “A note on the local behavior of the derivative of a cubic spline interpolator”, *Journal of Approximation Theory* 86, pp. 112-132.

Rosenblatt, M. (1991): “Stochastic Curve Estimation”, Institute of Mathematical Statistics, Hayward, California.

Rudemo, M. (1982): “Empirical choice of histograms and kernel density estimation”, *Scandinavian Journal of Statistics* 9, pp. 65-78.

Schoenberg, I.J. (1946): “Contributions to the problem of approximation of equidistant data by analytical function”, *Quarterly of Applied Mathematics* 4, pp. 45-99.

Scott, D.W. (1992): “Multivariate Density Estimation: Theory, Practice and Visualizati”, *Wiley* .

Scott, D.W and G.R. Terrell (1987): “Biased and unbiased cross-validation in density estimation”, *Journal of the American Statistical Association* 82, pp. 1131-1146.

Sheather, S.J. and B.C. Jones (1991): “A reliable data-based bandwidth selection method for kernel density estimation”, *Journal of the Royal Statistical Society, Ser. B* 53, pp. 683-690.

Silverman, B.W. (1986): “Density Estimation for Statistics and Data Analysis”, *Chapman and Hall.*

Simonoff, J.S. (1996): “Smoothing Methods in Statistics”, Springer-Verlag.

Stone, C.J. (1977): “Consistent nonparametric regression”, *Annals of Statistics* 5, pp. 595-620.

Tapia, R.A. and J.R. Thompson (1978): “Nonparametric Probability Density Estimation”, Johns Hopkins University Press.

Tarter, M.E. and R.A. Kronmal (1976): “An introduction to the implementation and theory of nonparametric density estimation”, *American Statistician* 30, pp. 105-112.

Tarter, M.E and M.D. Lock (1993): “Model-Free Curve Estimation”, Chapman and Hall.

Taylor, C.C. (1989): “Bootstrap choice of the smoothing parameter in kernel density estimation”, *Biometrika* 76, pp. 705-712.

Wahba, G. (1976): “Histosplines with knots which are order statistics”, *Journal of the Royal Statistical Society, Series B* 38, pp. 140-151.

Wahba, G. (1990): "Spline Models for Observational Data", SIAM, Philadelphia.

Wand, M.P. and Jones, M.C. (1995): "Kernel Smoothing", Chapman and Hall.

Watson, G.S. (1964): "Smooth regression analysis", *Sankhyā* Ser. A 26, pp. 359-372.

Watson, G.S. and M.R. Leadbetter (1964a): "Hazard analysis I", *Biometrika* 51, pp. 175-184.

Watson, G.S. and M.R. Leadbetter (1964b): "Hazard analysis II", *Sankhyā* Ser. A 26, pp. 101-116.

Wegman, E.J. (1982): "Density estimation", Kotz and Johnson (eds), *Encyclopedia of Statistical Sciences,* Vol. 2 Wiley, pp. 309-315.

Zheng, X. (1996): "A consistent test of functional form via nonparametric estimation techniques", *Journal of Econometrics* 75, pp. 263-289.

Resumen

Este artículo presenta una revisión de los métodos no paramétricos de estimación de curvas, orientada especialmente a las aplicaciones econométricas. La discusión se centra, sobre todo, en la estimación de funciones de densidad y curvas de regresión, pero se consideran también otras funciones relevantes (derivadas de curvas de regresión, funciones de riesgo, densidades y cuantiles condicionales). Discutimos brevemente las ideas en las que descansan estas técnicas y, en particular, el problema práctico de elección del parámetro de suavizado. El trabajo concluye con unos ejemplos ilustrativos basados en datos de la Encuesta de Presupuestos Familiares española.