Nonparametric Integration of Regression Functions

Christophe Bontemps[*] Arnaud Reynaud[†]

January 2000 [‡]

Abstract

For many problems in economics, econometricians only observe realizations of the derivative of the function they are interested in. This situation occurs each time one studies the objective function of an agent while only a first order condition resulting from an optimization problem is available. We propose a nonparametric procedure to estimate the integral of unknown, unspecified economic functions. A very simple integral estimator, built on Mack and Müller's kernel estimator (89), is suggested and analyzed. We propose *plug-in* and *cross-validation* criteria adapted to our bandwidth choice problem. A *factor method* developped for derivative estimators bandwidth selection is also examined ; we show how this method could be used to solve our bandwidth choice problem. An application to pollution abatement cost functions shows the potential of this procedure for many other types of economic problems.

[*]LEERNA and INRA, Toulouse. LEERNA is a joint research lab in the field of Environment and NAtural Resource Economics. The first author belongs to the INRA component (National Agronomical Research Institute), while the second is in the academic part (University of Social Sciences, Toulouse)

[†]LEERNA and GREMAQ (University of Social Sciences, Toulouse).

[‡]We would like to thank Jean-Pierre Amigues and Alban Thomas for helpful comments on this preliminary version.
Corresponding adress : LEERNA, Université des Sciences Sociales de Toulouse, 21 allée de Brienne, 31000 Toulouse, France. Email : bontemps@toulouse.inra.fr and areynaud@toulouse.inra.fr.

1 Introduction

In many economic problems, it is not possible to observe directly the function the economist is interested in. Information available is often reduced to some realizations of the derivative of this unknown function. This type of situation occurs each time econometricians [economists] are interested in the objective function of an agent whereas they only observe a first order condition resulting from an optimization program. Let us take two simple examples: the utility maximization problem of a consumer and the profit maximization program of a producer. By maximizing its utility under the budget constraint, the consumer equates the marginal utility of consumption to price. By maximizing his profits under the technological constraint, the producer equalizes the marginal product of each input to its price in term of output. In both cases, the resulting condition of this optimization behavior is a first order condition whereas the function of interest is its integral.

In the example of the consumer problem, the utility function measures direct welfare variations or unknown parameters such as risk-aversion and is of great interest. But usually, econometricians [economists] postulate a parametric form to represent the marginal condition and then use statistical methods to estimate the unknown parameters from the observed data. Since this condition results from an optimization program under constraints, the particular form has to satisfy a certain number of conditions. For example, the assumption of a quadratic utility function restricts the demand equation to be linear. But, because there is *a priori* no data nor information on the agent objective function, there are no reasons to specify a parametric form for the first order condition to be estimated. Clearly such a type of procedure suffers from the defect that the maintained hypothesis of parametric form can rarely be directly tested.

These examples clearly show that it may be inappropriate, in some cases, to use a particular form to estimate a function using data on its derivative. A natural question comes out from this discussion: *Is it possible to estimate a function without specifying any specific form, when one only observes its derivative in some points?* This is the question we are trying to answer in this paper by using a nonparametric method.

The aim of nonparametric methods is precisely to estimate functions without specifying functional forms. Historically, an area of application of these methods has been the analysis of observed choices made by economic agents in the field of consumption and production. See for example the seminal paper of Samuelson (1938) on the revealed preference theory or Diewert and Parkan (1978) and Varian (1984) for more recent nonparametric approaches to the economics of consumption and production.

More recently, nonparametric procedures have been used in regression analysis either for regression estimation where no assumption on the functional form nor on the distribution in the regression equation are specified (see the books of Härdle (1992) or Pagan and Ullah (1999)), or for regressors selection (Lavergne and Vuong (1996)), as well as for prediction in time series (see Bosq (1998)) or for testing parametric and nonparametric specifications (Härdle and Mammen (1989)) but also to compare nonparametric regression curves (Hall and Hart (1990)) and in many other fields of econometrics.

In these works, the kernel method is certainly the most popular method, mainly because its properties are best understood since its definition by Nadaraya (1964) and Watson (1964). However, the great flexibility of this method has a drawback which has been studied as often as the application of nonparametric methods themselves: "*the choice of the smoothing parameter, or bandwidth*". This quite technical issue has led to some important results concerning the optimal bandwiths related to some theoretical and applied criteria see for example the work of Rice (1984), Härdle and Marron (1989), Marron (1988), Härdle (1992), or Vieu (1991) for local bandwidths definition.

Another important issue, related to our integration problem, is the estimation of the derivative of a function as well as the function itself. The first derivative is often used, in regression analysis for example, to measure the response coefficient of the dependent variable with respect to the regressors. One may for example think of elasticities of a demand function depending on the derivatives of this unknown demand function. The second derivatives indicates the curvature of the function and are useful for testing some particular shape hypothesis. In parametric econometrics the estimation of these derivatives and testing of associated hypotheses are carried out by assuming some specific form of the relationship (either linear or non-linear). Any mispecification in this formulation may induce inconsistent estimates of the true derivatives and may affect the size and power of any test performed on them. As a response, nonparametric econometrics have focussed in the mid-eighties on methods to estimate these derivatives, without any *a priori* hypotheses on the shape of the function, see for example Gasser and Müller (1984), Härdle and Gasser (1985) and Ullah and Vinod (1993) for a review. Mack and Müller (1989) have also proposed a simplified kernel estimator allowing an easy derivation. This estimator will be used in this paper. Of course the problem of the bandwidth choice has also received a great attention in this framework (Rice (1986) and Müller, Stadtmüller, and Schmitt (1987)).

Surprisingly, integrating a function has not received the same focus. Yet, integrating a kernel estimator uses the same methodology and ideas than those used for derivating, even if it is often said that "*it is more difficult to integrate than to derivate*". We will see that it is not the case when using a nonparametric estimation framework. In this paper, we propose a nonparametric kernel method to estimate the integral of unknown, unspecified regression functions, built on Mack and Müller's kernel estimator. This estimator is very simple and easy to compute. As usual the bandwidth choice for this estimator is crucial, and has to be studied in details. The question of the error criterion used for the selection is clearly a main issue and may lead to different bandwidth selection procedures (either *plug-in* or *cross-validation*). We will discuss several questions of interest on that problem and link it to its related problem of derivative estimators bandwidth choice. The work already done in that field may be quite useful and will be examined: if the optimal bandwidths used for a derivative estimator are related in some way to the one used for the estimator itself as it is said in Müller, Stadtmüller, and Schmitt (1987), one could take advantage of this result.

The paper is organised as follows. In the next section we present the nonparametric model and define the integral estimator we propose. In section 3, we discuss the bandwidth choice problem and propose *plug-in* and *cross-validation* criterions adapted to our estimator. We also analyse various results on bandwidth choice for derivative estimators and propose a *factor method* adaptated to our estimator. Finally, in the last section, we conclude by an application in the field of production analysis. We estimate the pollution abatement cost function on a sample of French firms. This empirical example shows, in practice, how to get the estimation of a cost function from data on marginal abatement costs.

2 Nonparametric model and integral estimator

Let us consider $q = p + 1$ economic variables (X, Y) where Y is the dependent variable and X is a $p \times 1$ vector of regressors. If $E(\mid Y \mid)$ is finite, then the nonparametric regression model can be expressed as :

$$Y = f(x) + U \tag{1}$$

where $f(x) = E[Y \mid X = x]$ and U is the error term. If we denote by $\varphi(y, x) = \varphi(y, x_1, ..., x_p)$ the unknown joint density of (X, Y) at point (x, y) and similarly $\varphi(x)$ the marginal density of X at x, we can write $f(x)$ as :

$$f(x) = E[Y \mid X = x] = \frac{\int y \cdot \varphi(y, x) dy}{\varphi(x)} = \frac{g(x)}{\varphi(x)} \tag{2}$$

for $\varphi(x) \neq 0$.

The unknown function f can be highly nonlinear and is not assumed to be expressible in some parametric form. Usually the function f is of special interest by itself and its estimation has been studied in detail in numerous papers (see Härdle (1992) or Pagan and Ullah (1999) for a recent review).

However, recent works have focussed on the partial derivative of it. If $f(x)$ is linear that is $f(x) = x_1 \cdot \beta_1 + \cdots + x_p \cdot \beta_p$ as it is specified in *parametric* econometrics, then the p first-order partial derivatives of f are simply the regression parameters $(\beta_1, \cdots, \beta_p)$. These terms are constant for x varying, while a nonparametric specification of f allows these derivatives to vary with x. The estimation of a function's integral, let say F of f, follows the same procedure. If we impose linearity for f, we will get F quadratic, the coefficients being determined by the linear estimation of f. It is one thing to impose f linear and another to hold its derivative constant or its integral quadratic. This is where nonparametric estimation helps in avoiding to specify or constrain a function, specially when it is not the function of interest.

Now, suppose that we have data $(X_i, Y_i)_{i=1,..,n}$; n i.i.d. observations upon the $q \times 1$ vector (X, Y). Then from (1), we have:

$$Y_i = f(X_i) + U_i \tag{3}$$

where the error term U_i is such that, by construction, $E[U_i \mid X_i] = 0$, and $E[U_i^2 \mid X_i] = \sigma^2(X_i) < \infty$.

The first aim of nonparametric estimation is to approximate $f(x)$ arbitrarely closely on the sample, provided that n is large enough. A large class of nonparametric estimators have been defined as a weighted sum of the dependent variable (Ullah and Vinod (1993)) like the well known Nadaraya-Watson kernel estimator:

$$\begin{aligned} \widehat{f(x)} &= \sum_{i=1}^{n} Y_i \cdot W_{NW}(x, X_i, h_n) \\ &= \frac{1}{n \cdot h_n^p} \cdot \sum_{i=1}^{n} Y_i \cdot \frac{K(\frac{X_i - x}{h_n})}{\frac{1}{n \cdot h_n^p} \cdot \sum_{i=1}^{n} K(\frac{X_i - x}{h_n})} = \frac{\widehat{g}(x)}{\widehat{\varphi}(x)} \end{aligned} \tag{4}$$

where $K(\cdot)$ is the kernel function with the scale factor or window width h_n; such that $h_n \to 0$ and $n \cdot h_n \to \infty$ as $n \to \infty$.

Provided that the weight function $W_{NW}(\cdot)$ is derivable, this estimator can be derived and one may easily obtain an estimator of the partial derivative $\frac{\partial f}{\partial x_j}$. The partial derivative's estimator is simply the estimator's partial derivative $\frac{\partial \widehat{f(x)}}{\partial x_j}$.

In 1989, Mack and Müller proposed a modified version of this estimator where the denominator of $\widehat{f(x)}$ in (4), an estimate of the marginal density $\varphi(x)$ at the fixed point x, is replaced by an estimate of $\varphi(X_i)$ evaluated at the sample value X_i, for $i = 1, \cdots, n$.

$$\widehat{f}_{MM}(x) = \frac{1}{n \cdot h_n^p} \cdot \sum_{i=1}^{n} Y_i \cdot \frac{K(\frac{X_i - x}{h_n})}{\frac{1}{n \cdot h_n^p} \cdot \sum_{j=1}^{n} K(\frac{X_j - X_i}{h_n})} \tag{5}$$

$$= \frac{1}{n \cdot h_n^p} \cdot \sum_{i=1}^{n} \frac{Y_i}{\widehat{\varphi}(X_i)} \cdot K(\frac{X_i - x}{h_n}) \tag{6}$$

The main advantage of this estimator lies in the absence of the variable x in the denominator which allows an easy derivation. In fact, the estimation of $\varphi(X_i)$ could be done in a complete different manner, as mentioned by Mack and Müller themselves, any other consistent estimate of $\varphi(\cdot)$ will work. In particular, one could use a different kernel function and a different bandwidth for $\widehat{\varphi}(X_i)$ than the one used for the numerator of equation (5), that is for $\widehat{g}(x)$.

Derivation of Mack and Müller's estimator is straightforward, as is its integration.

Proposition 1 *:*

A functional estimator $\widehat{F}$ for the integral F of the unknown function f, up to a constant C, is the integral of the estimate $\widehat{f}$ defined by[1] *:*

$$\widehat{F(u)} = \int_{-\infty}^{u} \widehat{f(x)}\, dx + C$$

using the Mack and Müller estimator $\widehat{f_{MM}}$, leads to :

$$\begin{aligned} \widehat{F(u)} &= \frac{1}{n \cdot h_n^p} \cdot \int_{-\infty}^{u} \sum_{i=1}^{n} \frac{Y_i}{\widehat{\varphi}(X_i)} \cdot K(\frac{X_i - x}{h_n})\, dx + C \\ &= \frac{1}{n \cdot h_n^p} \sum_{i=1}^{n} \frac{Y_i}{\widehat{\varphi}(X_i)} \cdot \int_{-\infty}^{u} K(\frac{X_i - x}{h_n})\, dx + C \end{aligned} \tag{7}$$

The value of this function $\widehat{F(\cdot)}$ at one point is needed to identify the constant C, for the ease of the presentation we will omit this parameter and assume than $C = 0$ until the economic application in section 4 where we estimate it.

At first, one may think that this estimator is a Mack and Müller estimator with a new kernel $K_I(X_i, u, h_n) = \int_{-\infty}^{u} K(\frac{X_i - x}{h_n})\, dx$:

$$\widehat{F(u)} = \frac{1}{n \cdot h_n^p} \cdot \sum_{i=1}^{n} \frac{Y_i}{\widehat{\varphi}(X_i)} \cdot K_I(\frac{X_i - x}{h_n})\, dx \tag{8}$$

But it is not. The so-called kernel K_I does not satisfy the assumptions necessary for the definition of kernel estimator[2], and is not a kernel[3], therefore $\widehat{F(\cdot)}$ is not really a kernel estimator, even if this estimator is still a weighted sum of the dependent variable :

$$\widehat{F(u)} = \sum_{i=1}^{n} Y_i \cdot W_I(u, X_i, h_n) \tag{9}$$

---

[1] *For simplicity in the equation we use the above simplified notation ; of course this integral is on all components and should be written :*

$$\widehat{F}(u_1, \cdots, u_p) = \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_p} \widehat{f}(x_1, , \cdots, x_p) dx_1, \cdots, dx_p$$

This notation will be used without specific mention in the rest of paper, all the integrals refering to x and dx should be understood as $(x_1, , \cdots, x_p)$ and $dx_1, \cdots, dx_p$.

[2] In the sence of Parzen-Rosenblatt see Bosq and Lecoutre (1987).

[3] Otherwise $\widehat{F(x)}$ would converge to $\cdots f(x)$ instead of its integral.

the weight W_I being :

$$W_I(u, X_i, h_n) = \frac{\frac{1}{n \cdot h_n^p} \cdot \int\limits_{-\infty}^{u} K(\frac{X_i - x}{h_n})\ dx}{\widehat{\varphi}(X_i)} \tag{10}$$

The properties of this estimator are closely related to those of Mack and Müller estimator, and will be discused in the next section.

As for derivation, integration on one component, x_s of $x = (x_1, \cdots, x_p)$, may be of interest as we will see in our application in section 4. The estimator is straightforward, the multivariate kernel $K(\frac{X_i - x}{h_n})$ being simply integrated on that component.

Proposition 2 *:*

A functional estimator $\widehat{F_{part}}$ of f's integral on one component, say x_s for $s = 1, \cdots, p$, up to a constant C_s, is the integral of its estimate defined by :

$$\widehat{F_{part}}(x_1, \cdots, u_s, \cdots, x_p) = \int_{-\infty}^{u_s} \widehat{f(x)}\ dx_s + C_s$$

or

$$\begin{aligned} \widehat{F_{part}}(x_1, \cdots, u_s, \cdots, x_p) &= \frac{1}{n \cdot h_n^p} \cdot \int\limits_{-\infty}^{u_s} \sum_{i=1}^{n} \frac{Y_i}{\widehat{\varphi}(X_i)} \cdot K(\frac{X_i - x}{h_n})\ dx_s + C_s \\ &= \frac{1}{n \cdot h_n^p} \sum_{i=1}^{n} \frac{Y_i}{\widehat{\varphi}(X_i)} \cdot \int\limits_{-\infty}^{u_s} K(\frac{X_i - x}{h_n})\ dx_s + C_s \end{aligned} \tag{11}$$

3 Bandwidth choice

The question of the validity of any nonparametric estimator is often considered in terms of the bandwidth and the kernel choice. The two most popular choices for kernels, either the gaussian density or the Epanechnikov kernel, are both easily integrable (at least numerically). We will not deal with that question here, guessing that the answer may not differ from the literature on kernels, see Härdle (1992) or Ullah and Vinod (1993) where it is often said that this choice is much less important than the bandwidth choice. We will see, however, that the choice of the kernel for derivation may be somehow linked with the choice of the bandwidth in the factor method developped by Müller, Stadtmüller, and Schmitt (1987).

The difficulty of the bandwidth selection lies here in the fact that we are interested in the integral of a function. Hence, any criterion for selecting the bandwidth should incorporate a measure of distance between F and its estimate $\widehat{F}$, while the data available are dealing with f and $\widehat{f}$. We may suggest several options at this stage :

- A *measure-based* solution : construct a measure of $\widehat{F(\cdot)}$'s accuracy, then find a decomposition of this measure and derive the optimal rate of convergence for the bandwidth. This is the traditional way of deriving optimal bandwidth. Theoretical asymptotical calculus may lead either to a *"plug-in* or a *"cross-validation"* method. These two options will be discussed in detail.

- A *factor method* solution : choose a bandwidth for $\widehat{f(\cdot)}$ (either by *plug-in*, *cross validation*, or any other method) that efficiently estimates f, and use it, not directly but as a basis for the bandwidth in $\widehat{F(\cdot)}$. That method is directly inspired by the *factor method* presented in Müller, Stadtmüller, and Schmitt (1987). It uses the property that there is a connection between asymptotically optimal bandwidth for the estimator $\widehat{f}$ and the one used for its ν-th derivative $\widehat{f}^{\nu}$. Estimating the terms linking the bandwidths provides a powerful tool that may be adapted to our case.

- A *data-generating* solution : estimate f with a "correct" bandwidth, and generate a pseudo-sample (potentially large), with $\widehat{f(\cdot)}$. Then use these pseudo-points $(X_c, Y_c)_{c=1,\cdots,N_c}$ for estimating F. Because of the size of the pseudo-data sample, which can be as great as desired, (and also because one may compute uniformly distributed simulated points X_c) $\widehat{F(\cdot)}$ becomes less sensitive to the bandwidth choice, that choice can be done with a classical *rule of thumb*. This solution is a kind of pragmatic way of avoiding problems by using well known tools adapted to $\widehat{f}$ even if $\widehat{F}$ is under study.

The traditional method of assessing the performance of estimators, proposed as a first option, is to consider some sort of error criterion, either pointwise or global. As most applications of function estimation call for the entire domain, instead of its value at one particular point, we may focus on global measures only.

The most popular criteria for measuring the accuracy of nonparametric estimator are the Integrated Squared Error defined as $ISE(h) = \int(f(x) - \widehat{f(x)})^2\, dx$ or its expected value, the Mean Integrated Squared Error $MISE(h) = E[\int(f(x) - \widehat{f(x)})^2\, dx]$. From these criteria[4], one may find some $Variance(h) + Bias(h)$ expression depending on the bandwidth and on some unknown functional parameters, and therefore derive the optimal rate of convergence for the bandwidth. These criteria can be applied with $(F(x) - \widehat{F(x)})^2$, and a *"plug-in"* method could be implemented. In our case, the calculus remains undone yet (details may appear in Bontemps (2000)), mainly because of the technical aspects of these calculus, but also because of the theoretical nature of the results.

The *Cross-validation*, which is also a popular tool for bandwidth selection, does not apply directly to our estimator, because the comparison between $\widehat{F(X_i)}$ and the data has no sense. However,

[4]There exist many others like the empirical version of $ISE(h)$, the Average Squared Error, defined as $ASE(h) = \sum_{i=1}^{n}\left(f(X_i) - \widehat{f(X_i)}\right)^2$ see Marron (1988) or Härdle (1992).

one may propose a method and a new cross-validation criteria, similar to the one used for estimating f itself. Following the methodology used in the setting of derivative estimation (see Härdle (1992), p.160), one may construct a classical leave-one-out estimator $\widehat{F^{(-i)}}(X_i)$:

$$\widehat{F^{(-i)}}(X_i) = \frac{1}{n \cdot h_n^p} \cdot \sum_{j=1; j \neq i}^{n} Y_j \cdot W_I(X_i, X_j, h_n) \tag{12}$$

Suppose, for the moment, that we are in the setting of ordered, fixed, equidistant predictor variable $\left(X_{(i)}, Y_{(i)}\right)_{i=1,\cdots,n}$ with $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Instead of comparing this leave-one-out estimator with the original data, one can construct another error estimator $CV^I(h)$, comparing $\widehat{F^{(-i)}}(X_i)$ with $Y^*_{(i)} = (Y_{(i)} + Y_{(i-1)})\frac{\left(X_{(i)} - X_{(i-1)}\right)}{2}$, a naïve estimate[5] of the integral at point $(X_{(i)}, Y_{(i)})$. This would lead to :

$$CV^I(h) = \sum_{i=1}^{n} \left(Y^*_{(i)} - \widehat{F^{(-i)}}(X_{(i)})\right)^2 \cdot \omega(X_{(i)}) \tag{13}$$

where $\omega(X_{(i)})$ is a weight function. This criterion has to be examined precisely, this work being beyond the scope of this paper (details will appear in Bontemps (2000)).

These two methods may be quite useful, but we would like to spend some time on the promising *factor method* developed on optimal kernels (see Müller (1984)) and used for derivative estimator bandwidth choice by Müller, Stadtmüller, and Schmitt (1987) in the case of equally spaced X_i's. Let $f^{(\nu)}$ be the ν-derivative of the regression function f. Then compute an appropriate bandwidth for $\nu = 0$, say h_0, with an appropriate kernel K_0 of order k for the estimator $\widehat{f} = \widehat{f}^{(0)}$. This bandwidth can be used with great accuracy for $\widehat{f}^{(\nu)}$ using an appropriate kernel K_ν of order k. Müller et ali showed that there is a strong connection between these bandwidths, defined as:

$$h_\nu = d_{\nu,k} \cdot h_0 \tag{14}$$

In this expression, the constant $d_{\nu,k}$ is known for various kernel and derivative orders, for example, for a kernel of order $k = 3$, and a first order derivation ($\nu = 1$), we have $d_{1,3} = 0.7083$.
As a result, provided that one uses "adapted" kernels, the bandwidth selection for derivative estimators is reduced to a "classic" problem of bandwidth choice for $\widehat{f}$.

A straitforward application of this result may be to use that method for integration : Using the familly of "adapted" kernels defined by Müller (1984) and a "reference" bandwidth h_1 for the first-order regression, $\widehat{f}^1$, (the latter opeation may be done in our case with classic *cross-validation* for example), and using the above relation (14), in reverse, gives a good estimation of the optimal bandwidth, h_0, for $\widehat{F} = \widehat{f}^{(0)}$.

[5] This estimation is done using simple *trapezoid* or *midpoint* rule for example, see Judd (1998) for details. For derivative, Härdle used the first difference of the $Y_{(i)}$'s : $\frac{Y_{(i)} - Y_{(i-1)}}{X_{(i)} - X_{(i-1)}}$ in its criterion.

4 An application to industrial cost function of water pollution abatement

In this section, we use the methodology presented in section(2) to estimate water pollution abatement cost functions. We use a large database on French industry for the period 1994-96.

4.1 Background

Industrial water pollution in France has been a major concern for the past two decades. Some recent reports of French Water Agencies have identified industrial pollution as the source of approximatively 50% of organic pollution. For toxic pollution, including heavy metals and some persistent organic pollution, industry is at the origin of the most part of effluent emissions. French current regulatory system is built on a combination of two instruments: emission charges and individual contracts between polluters and the regulatory agency. Emission taxes are based on effluent pollution emissions which correspond to the level of pollution generated by the production process minus the abated pollution. Table 4.1 presents the evolution for some industrial effluents during the two last decades in France.

Table 1: Evolution of industrial pollution 1974-1996

	Gross pollution		Net pollution		Abatement	
	COD (t/day)	Tox. (K.equ/day)	COD (t/day)	Tox. (K.equ/day)	COD (t/day)	Tox. (K.equ/day)
1974	4670	110000	3630	76500	1040	33500
1996	6039	119294	1604	17705	4435	101589
Evolution	+29%	+8%	–56%	–77%	+326%	+203%

COD: Chemical Oxygen Demand.

Tox.: biological measure of effluents toxicity.

Industrial firms can abate water pollution by scaling back polluting activities or by diverting resources to cleanup. In both cases, pollution reduction entails costs. Traditionally, abatement cost estimates have been based on plant's reported direct costs of setting up and operating pollution control equipment. Coupled with information about the benefits of reducing pollution, such cost estimates can provide a basis for setting efficient regulatory schemes, as showed by Thomas (1995) and McConnell and Schwarz (1992). But the scarcity of appropriate plant-level data often prevents detailed empirical studies of average and marginal abatements costs of pollution. Moreover, firms can adjust to the threat of higher pollution-related costs along many dimensions including new process technology, pollution control equipment, improved efficiency. It follows that policy analyses have frequently developed abatement estimates from engineering models. However, failure to rely on behavioral data may led to considerable errors.

As mentioned above, estimates of water pollution cost abatement function are very costly in terms of data information (input, prices, quantities, technologies). In general, the only thing observed by public authorities is the result of the industrial minimization of costs. The question we are going to answer can be stated as follows: *is it possible to infer from this condition the characteristics of industrial cost abatement function?*

4.2 A simple model of abatement choice by industrials

In order to answer this question, we need to describe abatement choices of a firm. We consider a representative firm with a production process generating a level of *gross pollution* P_G. The representative firm is facing a unit emission tax τ. It chooses a level of pollution *abatement* $A \in [0, P_G]$. Pollution abatement may be achieved by investing in and by operating an external abatement plant. It follows that the *net pollution* taxed by the Water Agency is equal to: $P_N = P_G - A$. We first assume that production and abatement process are independent. It is justified by the fact that abatement costs are small in comparison with production costs. Hence it is likely that emission tax does no affect allocation of inputs in the production process. We assume, next, that the abatement cost function depends on the gross pollution and the abatement level. Let C and K denote the variable and fixed cost of abatement process. The profit of the firm in the abatement activity is:

$$\prod = \tau(P_G - A) - C(P_G, A) - K$$

The representative firm maximizes his profit with respect to A under the constraints $A \geq 0$ and $P_G - A \geq 0$. For firms with non-binding constraints, the first order condition of this maximization program is:

$$\frac{\partial C}{\partial A}(P_G, A) = \tau \tag{15}$$

We also have some information concerning the unknown function C since for any value of P_G :

$$C(P_G, 0) = 0 \tag{16}$$

Using notations of section 2, this observed marginal condition may be written as $Y = f(X)$ with $Y = \tau$, $X = (X^1, X^2) = (P_G, A)$ and the unknown function $f(.) = \frac{\partial C}{\partial A}(.)$.

4.3 Econometric issues: data and estimation

4.3.1 Data

Data used were made available from the different French Water Agencies: Adour-Garonne, Artois-Picardie, Loire-Bretagne, Rhin-Meuse, Rhône-Méditerranée-Corse and Seine-Normandie. We use a sample of 153 plant-level observations.

Table 2: Descriptive statistics

Variable	Mean	Std Dev.	Variance	Min.	Max.
P_G	67706	56337	3173.10^6	2266	196389
A	45171	48620	2363.10^6	292	179504
τ	0.9678	0.0448	0.0020	0.8800	1.0400

The file includes yearly information at plant-level on various effluent emissions, emission charges and abatements for the 1994-96 period. The Water Agency data set contains levels of effluent emissions, both before and after treatment, for five categories of pollutants: nitrogen, suspended solids, chemical oxygen demand, phosphorous and absorbable organic halogens. A price index is computed as a weighted sum of average emissions charges in order to construct a sixth as an agregate of the others. The level of pollution abatement is then computed for each firm as the difference between gross and net effluent emissions.

4.3.2 Nonparametric estimator

As defined in section (2), the estimator for F will be based on a weighted sum of the data (X_i, Y_i) previously introduced. Since $p = 2$ in this application, we will use a bivariate kernel $K(\frac{X_i^1 - x^1}{h_n^1}, \frac{X_i^2 - x^2}{h_n^2})$ defined as the product of two univariate kernels with, *a priori* two different bandwidths, h_1 and h_2 since the marginal densities of x^1 and x^2 may be different.

$$K(\frac{X_i^1 - x^1}{h_n^1}, \frac{X_i^2 - x^2}{h_n^2}) = K(\frac{X_i^1 - x^1}{h_n^1}) \cdot K(\frac{X_i^2 - x^2}{h_n^2})$$

The bivariate version of Mack and Müller estimator is :

$$\begin{aligned}
\widehat{f}_{MM}(x^1, x^2) &= \frac{1}{n \cdot h_1 h_2} \cdot \sum_{i=1}^{n} Y_i \cdot \frac{K\Big(\frac{X_i^1 - x^1}{h_n^1}\Big) \cdot K\Big(\frac{X_i^2 - x^2}{h_n^2}\Big)}{\frac{1}{n \cdot h_1 h_2} \cdot \sum_{j=1}^{n} K(\frac{X_j^1 - X_i^1}{h_n^1}) \cdot K(\frac{X_j^2 - X_i^2}{h_n^2})} \\
&= \frac{1}{n \cdot h_n^p} \cdot \sum_{i=1}^{n} \frac{Y_i}{\widehat{\varphi}(X_i^1)\widehat{\varphi}(X_i^2)} \cdot K\Big(\frac{X_i^1 - x^1}{h_n^1}\Big) \cdot K\Big(\frac{X_i^2 - x^2}{h_n^2}\Big)
\end{aligned}$$

For the purpose of the proposed application, we are interested in the integral relative to only one component, the abated pollution, A, denoted here by x^2. We will therefore use a partial integration relative on the variable x^2, as defined by the first-order equation (15). According to proposition (2) the integration estimator, denoted $\widehat{F}_2(\cdot)$, is:

$$\widehat{F}_2(x^1,u^2) = \frac{1}{n\cdot h_n^p}\cdot\sum_{i=1}^{n}\frac{Y_i}{\widehat{\varphi}(X_i^1)\widehat{\varphi}(X_i^2)}\cdot K\Big(\frac{X_i^1-x^1}{h_n^1}\Big)\cdot\int_{-\infty}^{u^2}K\Big(\frac{X_i^2-x^2}{h_n^2}\Big)\,dx^2 \;+C_2 \qquad (17)$$

Since we have some information on the unknown fonction F (equation 16), we can estimate the constant C_2 by adjusting the estimator $\widehat{F}_2(\cdot)$ in that point, therefore C_2 must satisfy[6] :

$$\begin{aligned}\widehat{F}_2(x^1,0) &= 0\\ &so\\ C_2 &= -\frac{1}{n\cdot h_n^p}\cdot\sum_{i=1}^{n}\frac{Y_i}{\widehat{\varphi}(X_i^1)\widehat{\varphi}(X_i^2)}\cdot K\Big(\frac{X_i^1-x^1}{h_n^1}\Big)\cdot\int_{-\infty}^{0}K\Big(\frac{X_i^2-x^2}{h_n^2}\Big)\,dx_2\end{aligned}$$

As we mentioned in section (3), we have to choose a kernel and the bandwidth for the estimation. We must say here that we have not followed the recommendations we made. Using a kernel of order (1) as defined by Müller (84) for $\widehat{f}$ and of order 0 for $\widehat{F}_2$ would have been more clever and could have reduced our bandwidth choice. However this method has not been defined nor tested for random predictor variable, and some work has to be done before using it in our framework. The *plug-in* and the *cross validation* method are also under study and are not operational yet.
So, we have used a traditional bivariate gaussian kernel while the banwidths for $\widehat{F}_2$ have been chosen in two steps ; first we have estimated the cross-validated bandwidths for $\widehat{f}$, then we have constructed a grid of simulated points $(X_c,Y_c)_{c=1,\cdots,N_c}$. Since the estimator $\widehat{F}_2$ applied to this *pseudo-sample* has revealed to be very insensitive to bandwidths, we have applied an *ad-hoc* bandwidth choosen by a classic *rule of thumb*. As we will see in the next section the results are quite promising.

4.4 Results of estimations

4.4.1 The marginal cost function

The economic theory suggests that use of marked-based instruments, such as effluent charges, allows to realize significant efficiency gains by confronting polluters to the price of polluting at the margin. But the theory says nothing about the actual magnitude of variations in marginal abatement costs. In practise, these have to be large enough to convince policy makers that the efficiency benefits of market-based instrument will outweight the cost of the transition to a new regulatory regime. Figure 1 presents an estimation of the marginal abatement cost function based on a sample of 153 industrial observed from 1994 to 1996.

[6] More precisely we should use $\widehat{C_2}$ instead of C_2 since the latter constant is estimated.

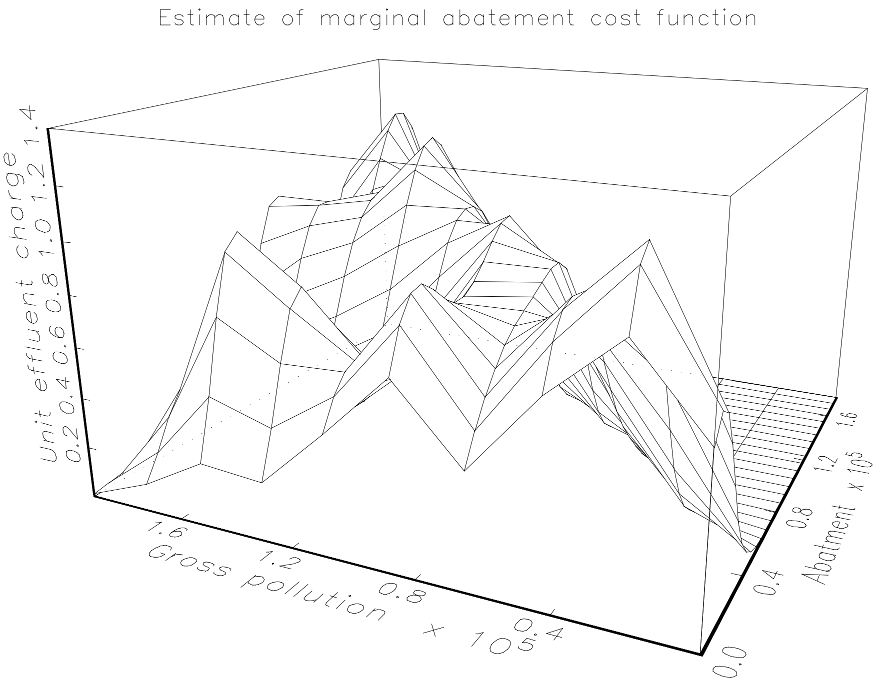

Figure 1: Estimation of the marginal abatement cost function $\widehat{c}(P_G, A) = \frac{\partial C}{\partial A}(P_G, A)$.

This graphic clearly highlights the variability in marginal abatement costs. Marginal abatement cost of pollution exhibits very large differences by scale and degree of abatement. For example, given a small level of abatement, the higher is the gross pollution, the smaller is the marginal abatement cost.

4.4.2 The cost function

Figure 2 presents an estimation of the abatement cost function based. It gives for every level of gross pollution the cost of depollution depending upon the level of abatement. This estimate allows in particular to analyse the effects of setting up effluent standards on abatement costs. The shape of this function is in accordance with the results obtained by Thomas (1995) and McConnell and Schwarz (1992) where parametric specifications of the cost function were done. A more detailed analysis of these results have to be done in the next future. Specially, we have used all the data available, while separate estimations of cost functions on subsets (depending on the level of effluent emissions, or on the industry), an therefore a comparison of these cost functions, would be very interesting.

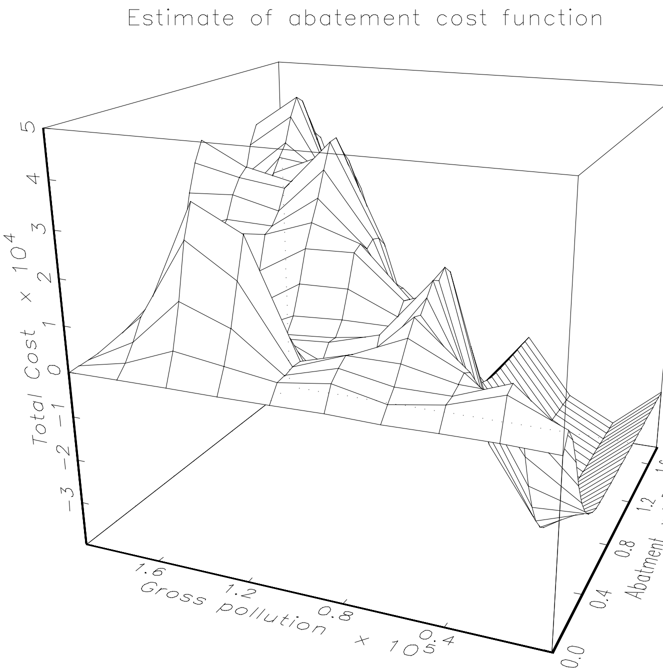

Figure 2: Estimation of the abatement cost function $\widehat{C}(P_G, A)$.

5 Conclusion

In this paper, we propose a nonparametric kernel method to estimate the integral of unknown, unspecified regression functions, built on Mack and Müller's kernel estimator (89). The latter estimator, which was designed for estimating derivatives of regression functions, is simply integrated. The estimator we propose for the integral of regression functions is simply the integral of Mack and Müller estimator. Our estimator is therefore very simple and easy to compute. The bandwidth choice for this estimator is crucial, and has been studied in details. The question of the error criterion used for the selection is still a main issue and has led to different bandwidth selection procedures (either *plug-in* or *cross-validation*). We have discussed several options on that problem and linked it to its related problem of derivative estimators bandwidth choice. The work already done in that field may be quite useful and has be examined. A *factor method* based on the work of Müller, Stadtmüller, and Schmitt (1987) has been proposed. It uses the result that there is a connection between asymptotically optimal bandwidth for a kernel estimator and the ones used for its derivatives.

As an application of the estimator proposed, we provided a numerical example in the field of production analysis. We have estimated the pollution abatement cost function on a sample of French firms. This empirical example shows, in practice, how to obtain the estimation of a cost function from data on marginal abatement costs.

We may propose several extensions to this paper either from a methodological point of view or for applied work in the field of our specific application.

Despite its simplicity, the estimator properties have not been discussed in detail, we will derive its statistical properties soon. This work may lead to improvement, specially on the choice and relative speed of convergence for the two bandwidths involved in our estimator: one for the numerator and one for the marginal density estimator (denominator). The three methods proposed for bandwidth selection (*plug-in*, *cross-validation*, and the *factor method*) have also to be developped more precisely for our integration estimator. Their statistical properties must be properly caracterized, and simulations have to be done before being fully operational.

It would also be interesting to compare the results done here on the estimation of the abatement cost function, with those presented in Thomas (1995). Such a "*parametric versus nonparametric*" comparison of cost functions may reveal important features on the impact of parametric specifications of first order relationships. A possible extension would be to use this nonparametric framework for testing specification (or misspecification) of objective functions in economic optimization procedures.

A Appendix

An algorithm to compute $\widehat{F(x)}$.

This algorithm has a $O(n^2)$ complexity but can be ran on parallel.

Let $\Phi(X_i, u, h_n) = \int\limits_{-\infty}^{u} K(\frac{X_i - x}{h_n})\, dx$ be the kernel integral.

The algoritm can be easily implemented as follow :

- Generate the vector $(X_c)_{c=1,\cdots,N_c}$ of the N_c points where you want $F(X_c)$ to be computed
- For $i = 1, \cdots, n$, compute the density estimation at the sample points $\widehat{\varphi(X_i)}$
- Define the new vector $V_i = \frac{Y_i}{\widehat{\varphi(X_i)}}$
- Create the weight matrix $W_{i,c} = \Big(\Phi(X_i, X_c, h_n)\Big)_{i=1,\cdots,n}^{c=1,\cdots,N_c}$
- The matrix product $Y_c = W_{i,c} \cdot V_i$ gives the vector of all $F(x_c) = Y_c$.

References

BONTEMPS, C. (2000): "Bandwidth selection for nonparametric integration estimators," mimeo LEERNA-INRA Toulouse.

BOSQ, D. (1998): *Nonparametric statistics for stochastic processes.* Springer.

BOSQ, D., AND J. P. LECOUTRE (1987): *Théorie de l'estimation fonctionnelle.* Economica.

DIEWERT, W. E., AND C. PARKAN (1978): "Tests for the Consistency of Consumer Data and Nonparametric Index Numbers," Discussion Paper, 78-27, University of British Columbia.

GASSER, T., AND H. G. MÜLLER (1984): "Estimating regression functions and their derivatives by the kernel method," *Scandinavian Journal of Statistics*, 11, 171–185.

HALL, P., AND J. D. HART (1990): "Bootstrap test for differnece between means in nonparametric regression," *Journal of the American Statistical Association*, 85, 1039–1049.

HÄRDLE, W. (1992): *Applied nonparametric regression.* Cambridge University Press.

HÄRDLE, W., AND T. GASSER (1985): "On robust kernel estimation of derivatives of regression functions," *Scandinavian Journal of Statistics*, 12, 233–240.

HÄRDLE, W., AND E. MAMMEN (1989): "Comparing nonparametric versus parametric regression fits," *Annals of Statistics*, 21(4), 1926–1947.

HÄRDLE, W., AND J. S. MARRON (1989): "Optimal bandwidth selection in nonparametric procedure regression function estimation," *Annals of Statistics*, 13(4), 1465–1481.

JUDD, K. L. (1998): *Numerical methods in economics.* M. I. T. Press.

LAVERGNE, P., AND Q. H. VUONG (1996): "Nonparametric selection of regressors : The nonested case," *Econometrica*, 64, 207–219.

MACK, Y., AND H. G. MÜLLER (1989): "Derivative estimation in nonparametric regression with random predictor variable," *Sankhya*, 51, 59–72.

MARRON, J. S. (1988): "Automatic smoothing parameter selection: A survey," *Empirical Economics*, 13, 187–208.

MCCONNELL, V., AND G. SCHWARZ (1992): "The supply and demand for pollution control : evidence from wastewater treatment," *JEEM*, 23, 54–77.

MÜLLER, H. (1984): "Smooth Optimum kernel estimators of regression curves, densities, and modes," *Annals of Statistics*, 12, 766–774.

Müller, H., U. Stadtmüller, and T. Schmitt (1987): "Bandwidth choice and confidence intervals for derivatives of noisy data," *Biometrika*, 74, 743–749.

Nadaraya, E. A. (1964): "On estimating regression," *Theory of Probability and its Applications*, 9, 141–142.

Pagan, A., and A. Ullah (1999): *Nonparametric econometrics*. Cambridge University Press.

Rice, J. A. (1984): "Bandwidth choice for nonparametric regression," *Annals of Statistics*, 12-4, 1215–1230.

——— (1986): "Bandwidth choice for differentiation," *Journal of Multivariate Analyis*, 19, 251–264.

Samuelson, P. (1938): "A Note on the Pure Theory of Consumer Behaviour," *Econometrica*, 5, 61–71.

Thomas, A. (1995): "Regulating polution under assymetric information : The case of industrial wastewater treatment," *Journal of Environmental Economics and Management*, 28, 357–373.

Ullah, A., and H. D. Vinod (1993): "General nonparametric regression estimation and testing in econometrics," in *Handbook of statistics*, ed. by G. Maddala, C. Rao, and H. D. Vinod. Elsevier science publishers.

Varian, H. (1984): "The Nonparametric Approach to Production Analysis," *Econometrica*, 52, 579–98.

Vieu, P. (1991): "Nonparametric regression : Optimal local bandwidth choice," *Journal of the Royal Statistical Society*, 53, 453–464.

Watson, G. S. (1964): "Smooth regression analysis," *Sankhya*, 26, 359–372.