1987

Flexible Production Function Estimation by Nonparametric Kernel Estimators

H. D. Vinod

Aman Ullah

ISSN: 0831-4454
ISBN: 0-7714-0923-0

THE CENTRE FOR DECISION SCIENCES AND ECONOMETRICS

Flexible Production Function Estimation by Nonparametric Kernel Estimators

H.D. Vinod
and
A. Ullah

TECHNICAL REPORT NO. 18
JULY 1987

Centre For Decision Science And Econometrics
Social Science Centre
The University of Western Ontario
London, Ontario N6A 5C2

FLEXIBLE PRODUCTION FUNCTION ESTIMATION

BY NONPARAMETRIC KERNEL ESTIMATORS

by

H.D. VINOD*

FORDHAM UNIV., BRONX, NEW YOURK, U.S.A.

and

A. ULLAH

UNIVERSITY OF WESTERN ONTARIO

JUNE 1987

*Some work on this paper was done when H.D. Vinod visited University of Western Ontario, whose generous facilities are greatly appreciated. Research support to A. Ullah from the NSERC is also gratefully acknowledged. Earlier versions of this paper were presented at the Rutgers University Productivity Symposium at Newark, NJ in December 85 and the Canadian Economic Study group meeting at Montreal, in September 1986. The authors thank the referees and P. Rilstone for helpful comments.

FLEXIBLE PRODUCTION FUNCTION ESTIMATION BY NONPARAMETRIC KERNEL ESTIMATORS

H. D. Vinod and A. Ullah

Fordham University and University of Western Ontario

ABSTRACT

Productivity studies can benefit from reliable estimates of production/cost functions. The assumption of homogeneity was discarded by using translog type forms in the 1970's by Christensen, Jorgenson, Lau, Vinod, Sudit and others. Barnett (1983) has suggested using flexible Laurent series to improve the second order approximations. A serious drawback of these flexible specifications has been their sensitivity to multicollinearity, and their need for several parameter estimates. This paper considers a nonparametric nonlinear amorphous functional specification which remains parsimonious in the number of parameters used. Rosenblatt suggested kernel methods which were extended by Watson and Nadaraya to conditional densities and expectations. We propose kernel estimates of analytical partial derivatives of production and cost functions. Asymptotic properties of the proposed estimator are investigated. Two illustrative examples concern the production function for the Bell System, and railroad cost function. A simulation study is also included.

1. INTRODUCTION AND THE MODEL

The implications of economic theory are usually amorphous, in the sense that they are not tied to specific functional forms or statistical assumptions about error structures. Varian (1984) and the references therein illustrate the desire of the theorists to work with nonparametric models. This paper is concerned with nonparametric estimation of various elasticities and/or partial derivatives in the context of production/cost functions, and productivity studies. The current econometric methodology for estimation depends heavily on the parametric specification of the functional forms, and for some students it is almost impossible to imagine that direct estimation is possible.

We will now review some theory to indicate the potential role of nonparametric estimation. A joint production function may be stated in an implicit form by:

$$f(Y_1, Y_2, \ldots, Y_q, X_1, X_2, \ldots, X_\ell, t) = 0 \tag{1.1}$$

where the outputs are Y_1 to Y_q and the inputs are X_1 to X_ℓ, and the time variable is denoted by t which is sometimes used as a proxy for technological change. In the single output case (q=1) there are numerous studies in the literature. In the early 1970's there was a relaxation of the homogeneity assumption associated with the Cobb Douglas and CES production functions by Christensen, Jorgenson, and Lau (1973), Diewert (1971), Vinod (1972), Sudit (1973), and others. The trans-log (TL), and generalized Leontief (GL) are examples of flexible functional forms developed during this period. More recently Gallant (1981) and Barnett (1983) have proposed a Fourier and minflex Laurent functional form respectively, with additional parameters and constraints. Although an empirical estimation of a joint production was first suggested by Vinod (1968), researchers prefer to estimate joint cost rather than joint production functions. The cost function is given by:

$$C = g(Y_1, Y_2, \ldots, Y_q, P_1, P_2, \ldots, P_\ell, t) \tag{1.2}$$

where the input prices are denoted by P_i, and total cost by C, which is also written by:

$$C = \sum_i^\ell P_i X_i \tag{1.3}$$

Note that all summations in this paper start at 1 (i=1 to i=ℓ here). Total derivative of the natural log of C from (1.3) with respect to time is given using Shepherd's lemma and familiar manipulations by:

$$(d\ \ln C/\ dt) = \sum_i^\ell (S_i)[(d/dt)\ln P_i + (d/dt)\ln X_i] \tag{1.4}$$

where $S_i = P_i X_i / C$ is the share of i-th input in total cost. Total derivative (d/dt)ln g from (1.2) is an index of productivity. Assuming that the inputs are paid according to their marginal product, and using (1.4), direct manipulation yields:

$$-(d/dt)\ \ln g = \sum_j^q (Eg{:}EY_j)(d/dt)\ln Y_j - \sum_i^\ell S_i (d/dt)\ \ln X_i \tag{1.5}$$

where $Eg{:}EY_j$ denotes the elasticity of total cost with respect to j-th output. Using first differences in logs to approximate time derivatives, Caves, et.al. (1980) rewrite the productivity index of (1.5) as:

$$-(\ln g_t - \ln g_{t-1}) = \sum_j^q [0.5(Eg{:}EY_j)_t + 0.5(Eg{:}EY_j)_{t-1}].$$

$$\cdot[\ln Y_{jt} - \ln Y_{jt-1}] - \sum_i^\ell [0.5S_{it} + 0.5S_{it-1}][\ln X_{it} - \ln X_{it-1}] \tag{1.6}$$

which can be estimated if we can estimate the elasticity of cost with respect to i-th output.

In much of this literature there is an interest in estimating various elasticities of outputs, inputs, time, prices, costs, etc. with respect to other variables, all of which involve partial derivative estimation. For example there is an interest in estimating economies of scale as the scale elasticity. The current econometric methods for estimating these quantities relies heavily on parametric estimation based on specific functional forms. Once a parametric form is specified, one may implicitly rule out certain numerical estimates of elasticities, which may be economically meaningless. A truly flexible form should permit all reasonable values of estimated elasticities. According to Diewert (1971) a flexible functional form should be able to attain arbitrary levels of first- and second-order deterivatives at a predetermined point. This property is satisfied by the nonparametric specification in the sense that derivatives can take any real values. We suggest that nonparametric regression estimates discussed in the following section may provide truly flexible forms in this context.

2. NON-PARAMETRIC REGRESSION MODEL

Although our discussion can be carried out in terms of the simultaneous system of equations, we will discuss here the single equation regression model for easier exposition. Consider the usual regression model:

$$y = X\beta + \epsilon \qquad (2.1)$$

where y is an n X 1 vector of observations on the dependent variable, X is an n X p matrix of the data on p regressors $x_1, x_2, \ldots, x_p$, β is a p X 1 vector of parameters, and ϵ is an n X 1 vector of errors. If specification (2.1) is true, it is often appropriate to interpret the j-th element of β as partial derivatives of y with respect to x_j. With this interpretation in mind, econometricians transform their variables with logs, Box-Cox or Fourier transformations, or include higher powers, etc. to compute elasticity type functions of the elements of β. Recent research has focussed on achieving greater economic realism by using flexible functional forms such as translog, minflex Laurent, etc. as mentioned above.

From a purely scientific viewpoint of Sir R.A. Fisher, the partial derivative interpretation is appropriate only when the X matrix is truly a "design" matrix based on a carefully chosen experimental structure. When a social scientist uses (2.1) to infer about the partial derivatives, there is a sense in which he is "abusing" regression coefficients, Box (1966). If one wants to measure what happens to y when x_j alone is changed, one should change x_j (e.g. fertilizer) alone and measure the effect on y (e.g. yield). Unfortunately this is often impossible in econometrics.

We consider an amorphous specification:

$$y = R(x_1, x_2, \ldots, x_p) + \varepsilon \tag{2.2}$$

where the regression function R is arbitrary, nonparametric, nonlinear and unspecified. The density of y conditional on x_1 to x_p is denoted by $f(y|x_1,\ldots,x_p)$, its expectation is denoted by $E(y|x_1,\ldots,x_p)$ or $R(x_1,\ldots,x_p)$. Functions like elasticities derived from the regression function R can be estimated directly by nonparametric methods described below. McFadden (1985) notes that there are three types of methods for estimating (2.2). First, there is the cell partition method based on contingency tables familiar in mathematical statistics. Secondly, there is the quasi parametric nearest neighbor method based on the work of Stone (1977, 1984) explained by McFadden (1985). Thirdly one can use Rosenblatt-Watson-Nadaraya Kernel estimation.

McFadden notes that the three approaches are closely related and describes the second approach in detail.

This paper is concerned with direct estimation of the partial derivaties of $E(y:x_1,\ldots,x_p)$ with respect to x_j for $j=1,\ldots,p$. The conditional expectation will be estimated by the Kernel method, and the estimation of the analytic partial derivatives appears to be new. The brief description of the Kernel method is given in the following subsection.

2.1 Brief Description of the Kernel Estimator:

Consider the univariate case with the density f(x) of a random variable X as the derivative of its cumulative distribution function F(x):

$$f = f(x) = (d/dx)F(x) = \lim h^{-1}\,[F(x+h/2) - F(x-h/2)]$$
$$= \lim h^{-1}\, E[I(-\tfrac{1}{2} < (X-x)h^{-1} < \tfrac{1}{2}] \qquad (2.3)$$

where lim denotes the limit as h tends to zero and $I(a < X < b)$ denotes the indicator function which is 1 if X is in the interval (a,b) and zero otherwise. Rosenblatt (1956) considered the estimation of the density f from the observed data and noted that the indicator function can be replaced by a non-negative Kernel function K(w), where $w = (X-x)/h$. He shows that if K is such that the integral $\int K(w)dw = 1$, one can obtain a consistent estimator of the density f. Further developments were made by Parzen (1962) and Bartlett (1963). Cacoullos (1966) generalized it to the multivariate case as follows:

Let script $\mathcal{K}$ be a class of all Borel measurable real valued bounded functions K on the m dimensional Euclidian space R^m such that:

$$\int K(w)\, dw = 1, \quad \int |K(w)|dw < \infty$$
$$\|w\|^m\, K(w) \to 0 \text{ as } \|w\| \to \infty \qquad (2.4)$$

where $\|w\|$ is the usual Euclidian norm of w in R^m. Now Cacoullos estimated the joint density of m variates and it is given by

$$f_n(x_1,\ldots,x_m) = (n \prod_j^m h_j)^{-1} \sum_t^n K(w_{1t},\ldots,w_{mt}) \tag{2.5}$$

where $w_{jt} = (x_{jt}-x_j)/h_j$; x_{jt} for t=1,2,...,n are observations on the j-th (regressor) random variable. All products in this paper similar to summations start at 1.

As in Singh (1981) and Singh, Ullah and Carter (1985) we consider K in (2.5) belonging to $\mathcal{K}_r$, which is a class of all real valued Borel-measurable bounded functions K on R^m such that

$$\int (u_1)^{i_1},\ldots,(u_m)^{i_m} K(u_1,\ldots,u_m) = \begin{cases} 1 \text{ if } i_1=i_2=\ldots=i_m=0 \\ 0 \text{ if } 0 < (i_1+i_2\ldots+i_m) < r \end{cases} \tag{2.6}$$

where $\int |(u_1)^{i_1}\ldots(u_m)^{i_m} K(u_1,\ldots,u_m)| < \infty$ if $i_1+i_2+\ldots+i_m=r$

and $\|u\|^m |K(u)| \to 0$ as $\|u\| \to \infty$.

We note that for r=0,1, and 2, the m-variate standard normal density is a member of $\mathcal{K}_r$, and so is the function $K(u_1,\ldots,u_m)=2^{-m}\prod_j^m I(-1 < u_j < 1)$. In fact, functions K of the type $K(u_1,\ldots,u_m)=\prod_j^m K_j(u_j)$ where K_j are bounded symmetric (about zero) functions on R^1 with $\int K_j(u_j)du_j = 1$, $|K_j| < \infty$ and $|u_jK_j(u_j)|\to 0$ as $|u_j|\to\infty$ are members of $\mathcal{K}_r$ for r=0,1, and 2. An example of this is the m-fold product of the normal Kernels given by

$$K(u_1,\ldots,u_m) = \prod_j^m K(u_j);\; K(u_j) = (2\pi)^{-1/2}\exp[(-1/2)u_j^2]$$

which will be used in our numerical examples in section 4. The window width is chosen such that both the bias and the variance remain under control. Following Singh, Ullah and Carter (1985) and Ullah and Singh (1985) we choose:

$$h_j = \sigma_j n^L, \quad L=-1/(2r+m-1) \tag{2.7}$$

where σ_j is the standard deviation of x_j. For our estimation in Section 4 we replace σ_j^2 by its consistent estimator $s_j^2 = \sum_t^n (x_{jt} - \bar{x}_j)^2/n$.

Singh (1981) has shown that the choice of h_j in (2.7) is designed to have important desirable properties including: asymptotic normality, convergence, good speed of convergence, and mean-squared error reduction, provided r is appropriately large as explained in Remark 1 in Section 3.

Let y be denoted by x_m with m=p+1. The joint marginal density of $x_1,\ldots,x_p$ is estimated by

$$f_n(x) = f_n(x_1,\ldots,x_p) = a_n^{-1} \sum_t^n K(w_t) \quad (2.8)$$

where, from now on we denote (for brevity) $x=(x_1,\ldots,x_p)$ without the last term x_m=y, and

$$a_n = n \prod_j^p h_j, \quad K(w_t) = K(w_{1t},\ldots,w_{pt}) \quad (2.9)$$

This marginal density (2.8) is obtained by integrating out the $y=x_m$ variable. The conditional mean E(y|x) is the regression function R in (2.2)

$$R(x) = \int x_m f(x_1,\ldots,x_m)dx_m \;/f(x_1,\ldots,x_p) \quad (2.10)$$

where the numerator is the integral of y times the joint density, and denominator is the integral of just the joint density; both integrations being with respect to y. The integrals may be evaluated directly by the formulas in sections 3.323 and 3.462 of Gradshteyn et. al (1965) Tables of Integrals. Thanks to certain cancellations, Hardle (1984), the estimation of R function is easier than that of the joint and the marginal densities. The R function can be estimated by:

$$R_n(x) = \sum_t^n y_t K(w_t) / \sum_t^n K(w_t) \quad (2.11)$$

which is the Nadaraya-Watson Kernel regression function with $K(w_t)$ from (2.9). This completes our discussion of the Kernel estimator of the regression function.

We turn next to the estimation of the partial derivatives of R with respect to x_j which may be done either analytically or numerically. A simplified partial derivative of R(x) in (2.10) is denoted by

$$pd(x) = \frac{\partial R(x)}{\partial x_j} \tag{2.12}$$

and its amorphous estimate by apd(x). We have

$$apd(x) = \frac{\partial R_n(x)}{\partial x_j} = \sum_t^n y_t (K_{1t} - K_{2t}) \tag{2.13}$$

where

$$\begin{aligned} K_{1t} &= K'(w_t)(\textstyle\sum_t^n K(w_t))^{-1} \\ K_{2t} &= K(w_t)(\textstyle\sum_t^n K'(w_t))(\textstyle\sum_t^n K(w_t))^{-2} \\ K'(w_t) &= \partial K(w_t)/\partial x_j . \end{aligned} \tag{2.14}$$

For the normal kernel $K'(w_t) = h_j^{-1} w_{jt} K(w_t)$. Note that apd(x) represents the j-th response coefficient of y due to a unit change in x_j.

The numerical estimation of the amorphous estimator apd(x) will be illustrated later in Section 4, after we study its statistical properties in the following section.

3. PROPERTIES OF AMORPHOUS PARTIAL DERIVATIVE ESTIMATOR

In this section we discuss the consistency, bias, mean squared error, asymptotic normality, confidence intervals and related properties of the amorphous estimator apd(x) defined in (2.13). It is assumed that the econometrician is directly interested in the first r partial derivatives of the conditional expectation of y given x_1 to x_p without explicitly specifying the form of the relationship. We use the normal big oh, small oh notation so that $z_n = O(n^{\lambda})$ if $|n^{-\lambda}z_n|$ is eventually bounded, and $z_n = o(n^{\lambda})$ if $|n^{-\lambda}z_n|\to 0$, and the subscript p on o or O denotes "in probability". For some results we need the assumption A_r defined by:

A_r: For some integer r the r-th order partial derivatives of R(x) are continuous and bounded in some neighborhood of x.

Before obtaining the results for apd(x) we summarize the known asymptotic results for the kernel regression function $R_n(x)$ in a following Theorem (all the following convergence and rate results are with respect to $n\to\infty$).

THEOREM 1. <u>At every continuity point x</u>

$$\text{Bias } R_n(x) = ER_n(x) - R(x) = o(1) \tag{3.1}$$

<u>and with</u> $a_n = n\prod_i^p h_i \to\infty$ as $n\to\infty$

$$R_n(x)-R(x) = o_p(1) \tag{3.2}$$

$$(a_n)^{1/2}(R_n(x) - ER_n(x))\to N(0, \Lambda^*(x)) \tag{3.3}$$

<u>where</u>

$$\Lambda^*(x) = \sigma^2(x)(f(x))^{-1}\int K^2(u)du,\ \sigma^2(x) = V(\varepsilon|x) = E(\varepsilon^2|x) \tag{3.4}$$

<u>Further, if for some</u> $r \geq 1$, A_r <u>holds and with this</u> r, K <u>in</u> (2.11) <u>belongs to</u> $\mathcal{K}_r$, <u>then</u>

$$\text{Bias } R_n(x) = ER_n(x) - R(x) = O(h^r) \tag{3.5}$$

$$R_n(x) - R(x) = O_p(a_n^{-1/2} + h^r) \tag{3.6}$$

<u>and if</u> h_i's <u>are chosen in such a way so that</u>

$$a_n^{1/2} h^r = o(1) \tag{3.7}$$

we have

$$a_n^{1/2}(R_n(x) - R(x)) \rightarrow N(0, \overset{*}{\Lambda}(x)); \tag{3.8}$$

$h^r = \max(h_1^r, \ldots, h_p^r)$.

PROOF: For the case when $\{y_t, x_{1t}, \ldots, x_{pt}\}$, t=1,...,n are i.i.d., the proof of the Theorem 1 is given in Singh, Ullah and Carter (1985), Theorem (3.2). Also for the special case p=1 see Schuster (1972) and Prakasa Rao (1983, p. 283). For the time series case, under the strict stationarity of the $\{y_t, x_{1t}, \ldots, x_{pt}\}$ and certain conditions on ϕ-mixing coefficient, see Bierens (1985).

We also note another useful result for $R_n(x)$ from Schuster (1972) and Prakasa Rao (1983, p. 283) for p=1, and for $p \geq 1$ Bierens (1985). If ξ and η are two distinct points of x then, as $n \rightarrow \infty$,

$$a_n \operatorname{cov}(R_n(\xi), R_n(\eta)) \rightarrow \sigma^2(x)(f(\eta))^{-1} \int k(u)k(u + \frac{\xi-\eta}{h}) = o(1).$$

From this result it follows that, for two points $\xi = x + h_j/2$ and $\eta = x - h_j/2$ which are not distinct in the limit,

$$a_n \operatorname{cov}(R_n(x+h_j/2, R_n(x-h_j/2)) \rightarrow \sigma^2(x)(f(x))^{-1} \int K(u)K(u+1)du, \tag{3.9}$$

where $x \pm h_j/2 = x_1, \ldots, x_j \pm h_j/2, \ldots, x_p$ and $K(u+1) = K(u_1, \ldots, u_j+1, \ldots, u_p)$.

Now we present the results for our proposed estimator apd(x) of pd(x).

THEOREM 2. At every continuity point x.

$$\text{Bias }(x) = E(apd(x)) - pd(x)) = o(1) \tag{3.10}$$

and with $b_n = h_j^2 a_n = h_j^2 n \prod_i^p h_i \rightarrow \infty$ as $n \rightarrow \infty$

$$apd(x) - pd(x) = o_p(1) \tag{3.11}$$

$$b_n^{1/2}[apd(x) - Eapd(x)] \rightarrow N(0, 2\Lambda(x)) \tag{3.12}$$

where $\Lambda(x) = \sigma^2(x)(f(x))^{-1}(\int K^2(u)du - \int K(u)K(u+1)du)$. Further, if for some r ≥ 1, A_r holds and with this r, K in (3.4) belongs to $\mathcal{K}_r$, then

$$\text{Bias}(x) = E(\text{apd}(x)) - \text{pd}(x) = O(h_j^{-1} h^r) \tag{3.13}$$

$$\text{apd}(x) - \text{pd}(x) = O_p(b_n^{-1/2} + h_j^{-1} h^r) \tag{3.14}$$

and if h_i's are chosen according to (3.7)

we have

$$b_n^{1/2}[\text{apd}(x) - \text{pd}(x)] \to N(0, 2\Lambda(x)); \tag{3.15}$$

$h^r = \max(h_1^r \ldots h_p^r)$.

PROOF: From (2.2)

$$y_t = R(x_{1t}, \ldots, x_{pt}) + \varepsilon_t \tag{3.16}$$

Substituting this in (2.11) and writing $r(w_t)=K(w_t)/\sum_t^n K(w_t)$ we have

$$\begin{aligned} R_n(x) &= \textstyle\sum_t^n R(x_t) r(w_t) + \sum_t^n \varepsilon_t\, r(w_t), \\ &= \textstyle\sum_t^n R(w_t h + x) r(w_t) + \sum_t^n \varepsilon_t\, r(w_t) \\ &= R(x) + \textstyle\sum_t^n \varepsilon_t\, r(w_t) + o_p(1), \end{aligned} \tag{3.17}$$

where the third equality follows because $R(w_t h + x) = R(w_{1t}h_1 + x_1, \ldots, w_{pt}w_p + x_p) = R(x) + o(1)$ by using Taylor series expansion and the fact that $h \to 0$ as $n \to \infty$.

The partial derivative with respect to x_j on both sides of (3.17) is

$$\text{apd}(x) = \text{pd}(x) + \sum_t^n \varepsilon_t \frac{\partial r(w_t)}{\partial x_j} + o_p(1) \tag{3.18}$$

Thus the result in (3.10), showing asymptotic unbiasedness, follows by taking expectations on both sides of (3.18), and noting that $E(\varepsilon_t | x_t) = 0$ from (2.2).

Next consider the consistency result (3.11). This follows by noting that $E\sum_t^n \varepsilon_t r(w_t) = 0$ and $V(\sum_t^n \varepsilon_t r(w_t)) \simeq V(R_n(x)) = O(a_n^{-1})$ from (3.3). Thus by Tchebychev's inequality $\sum_t^n \varepsilon_t r(w_t) = O_p(a_n^{-1/2})$, and hence $\sum_t^n \varepsilon_t \partial r(w_t)/\partial x_j$ in

(3.18) is $-h_j^{-1} \sum_t^n \varepsilon_t \frac{\partial r(w_t)}{\partial w_{jt}} = O_p(h_j^{-1} a_n^{-1/2}) = O_p(b_n^{-1/2}) = o_p(1)$.

Now we consider the normality result in (3.12). For this, first we write from (2.13)

$$h_j(apd(x) - Eapd(x)) = [(R_n(x+h_j/2) - ER_n(x+h_j/2)]-[(R_n(x-h_j/2) - ER_n(x-h_j/2))]$$

or multiplying both sides by $a_n^{1/2}$

$$b_n^{1/2}(apd(x) - Eapd(x)) = a_n^{1/2}[(R_n(x+h_j/2) - ER_n(x+h_j/2)]-a_n^{1/2}[(R_n(x-h_j/2) - ER_n(x-h_j/2)]$$

Thus following the proof of (3.3) in Schuster (1972) and using (3.9) the normality result in (3.12) follows.

For the results (3.13), (3.14) and (3.15) we note that under the assumption A_r, and K belonging to $\mathcal{K}_r$ (3.17) becomes (also see (3.6))

$$R_n(x) = R(x) + \sum_t^n \varepsilon_t r(w_t) + O_p(h^r),$$

where, as indicated before, $\sum_t^n \varepsilon_t r(w_t) = O_p(a_n^{-1/2})$. Thus

$$apd(x) = pd(x) - h_j^{-1} \sum_t^n \varepsilon_t \partial r(w_t)/\partial w_{jt} + O_p(h_j^{-1} h^r).$$

Using this, and following the arguments used to prove (3.10) and (3.11) the results in (3.13) and (3.14) follow immediately. For the result (3.15) we note that

$$\begin{aligned} b_n^{1/2}(apd(x) - pd(x)) &= b_n^{1/2}(apd(x) - Eapd(x)) + \\ &+ b_n^{1/2} (Eapd(x) - pd(x)) \\ &= b_n^{1/2}(apd(x) - pd(x)) + O_p(a_n^{1/2} h^r) \end{aligned}$$

by using (3.13). Thus if (3.7) holds then the result (3.15) follows.

Now we include some comments regarding Theorem 2.

<u>REMARK 1</u>. The result in (3.10) shows the asymptotic unbiasedness of apd(x), whereas (3.13) gives the speed of convergence for the asymptotic unbiasedness. This speed gets faster as r gets large. However, note that the

construction of the kernel K belonging to $\mathcal{K}_r$ in (2.6) (and hence the estimator apd(x)) becomes difficult as r gets large. For example if r=10 we would need a kernel whose first nine order moments are zero.

REMARK 2. While (3.10) suggests that we can reduce the bias of the estimator apd(x) to any desired order by choosing our h_j converging to zero sufficiently fast, (3.12) suggests that we pay a price. That is, we may inflate the variance of (apd(x)) = $b_n^{-1} \Lambda(x) = O(b_n^{-1})$. However, if we control $h_j \propto n^{-1/(2r+p)}$ defined in (2.7), then both bias and variance (and hence the MSE) can be controlled simultaneously. In this case the speed of covergence of the MSE = $O(n^{-2r/(2r+p)})$, and the convergence speed in (3.14) becomes $apd(x) - pd(x) = O_p(n^{-r/(2r+p)})$. This suggests that the speed of convergence is inversely related to the number of regressors.

REMARK 3. From the result in (3.12), 100(1-α)% confidence interval for the true partial derivative, pd(x), is given by

$$apd(x) \pm z_{\alpha/2}(b_n)^{-½}[2\Lambda(x)]^{½}$$

where $z_{\alpha/2}$ is such that, if Z is the univariate standard normal random variable, then the probability $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$. In practice, a consistent estimator of $\Lambda(x) = \sigma^2(x)(f(x))^{-1}\int K^2(u)du$ can be used in (3.12) as $\Lambda_n(x) = \sigma_n^2(x)(f_n(x))^{-1}(\int K^2(u)du - \int K(u)K(u+1)du)$ where $f_n(x)$ is given by (2.8), $\sigma_n^2(x) = \sum_1^n \hat{\varepsilon}_t^2 r(w_t)$; $\hat{\varepsilon}_t = y_t - R_n(x_t)$, and $\int K(u)K(u+1)du = (\exp(-1/4))$ $\int K^2(u)du = (\exp(-1/4))\ (1/2\sqrt{\pi})^p$ for the normal kernel. Thus we have all the necessary properties needed for practical applications of apd(x).

4. NUMERICAL EXAMPLES

In this section we illustrate the methodology of the previous sections with examples. The first example regarding Bell System single output production function is chosen for simplicity. The second example based on multiple output railroad cost function is discussed in subsection 4.1. We are interested in the direct estimates of the regression function R of (2.10), which is the expectation of the dependent variable, conditional on the values of the regressors. From a scatter diagram of the dependent variable against its conditional expectation we note that the fit is good in the middle part of the range and somewhat distorted near the end points of the range. This is typical in most nonparametric estimation by Kernel methods as in Ullah and Singh (1985), Ullah (1985), among others.

A parametric approach to estimating the linear regression model is to specify:

$$E(y_t | x_{1t},\ldots,x_{3t}) = b_o + b_1 x_{1t} + b_2 x_{2t} + b_3 x_{3t} \tag{4.1}$$

The following is the familiar translog "flexible" production function for all Bell System telephone companies for 1947-1977, well before the recent divestiture:

$$\ln y = b_o + b_1 \ln K + b_2 \ln L + b_3 \ln K \ln L + b_4 (\ln K)^2 + b_5 (\ln L)^2 + b_6 \ln x_6 \tag{4.2}$$

where ln is natural log, y is the "valued added" output, K is the capital input based on the deflated value of the plant corrected for depreciation and age distribution, L is the labor input measured by employee hours adjusted for seniority, and x_6 is an index of technology based on Poisson weighted index of Research and Development expenditures. We have used an updated version of the data set used by Vinod [1976b, Sec. 5].

The Cobb Douglas production function is a special case of (4.2) when $b_3 = b_4 = b_5 = 0$. The scale elasticity (SCE) of a Cobb Douglas production function is given by simply $b_1 + b_2$, which measures the economies of scale, or the percentage change in output when both inputs are changed by one percent. The SCE for (4.2) is:

$$SCE = b_1 + b_2 + b_3 (\ln K + \ln L) + 2b_4 \ln K + 2b_5 \ln L \qquad (4.3)$$

If SCE is significantly larger than unity it indicates that there are scale economies, and lends support to the argument that the divestiture of the Bell System may not have been in the best interest of the telephone customer. This is because the divestiture has reduced the scale of operations and the economies of scale may have been lost. There are several papers attempting to estimate Bell System's SCE referenced in Vinod and Ullah [1981, p. 203], Docket 20003 of the Federal Communications Commission-(FCC).

Table 1 reports the estimates based on ordinary least squares (OLS) for linear regression based on a Cobb Douglas specification, as well as, our flexible amorphous estimates of the partial derivatives, apd(x), evaluated at the mean values of the regressors. The absolute values of the t ratios are also reported. In nonparametric estimation it is well known that the fit is poor near the end points of the data, because of a lack of information in their neighborhoods. Hence a fair comparison of residual sum of squares between OLS and amorphous models should exclude the end points. Table 1 reports the residual sum of squares after omitting two observations at each end, being roughly five percent trimming at each end.

Note that economic decision makers are specifically interested in the estimates of the partial derivatives, not in the regression coefficients of a linear or nonlinear model, per se. The linearity is an artifact of our specification, and the flexible forms attempt to mitigate the consequences of assuming log linearity, to achieve greater realism. Our amorphous method estimates the partial derivatives directly by actually keeping the variation

Table 1: OLS Cobb Douglas and Flexible Amorphous Regression Results, Bell System Data

Coeff. of:	Intercept	Capital	Labor	Technology	R^2
OLS	-3.303	0.52	0.70	0.49	0.997
Absolute t value	4.51	9.20	4.82	11.33	
Amorphous (r=1)		0.44	0.98	0.59	
(coeff/SE) ratio		3.16	1.90	4.56	
Amorphous (r=2)		0.43	0.96	0.54	
(coeff/SE) ratio		3.62	2.20	4.98	

Note: r is the largest order of derivative of interest used in defining the window h_j of (2.7). Trimmed Residual Sum of Squares for OLS is 0.0302, and for the amorphous model it is 0.0649 and 0.0915 for r=1 and r=2, respectively. We have excluded the first and last two observations, since it is well known that nonparametric methods do not yield good fits near the end points.

in other regressors at zero. We find that a one percent increase in inputs leads to a larger than one percent increase in the output, suggesting that SCE > 1, or economies of scale. In Vinod (1976b) ridge regression is used to alleviate the multicollinearity of the translog model, and has yielded meaningful estimates of partial elasticities. The choice r=2 is the highest permissible for the normal kernel by Remark 1 above. The choice r=1 yields a lower residual sum of squares (better fit), but lower ratio of the coefficient to its standard error. An appealing feature of our nonparametric estimates of elasticities is that they are based on a most "flexible" specification.

4.1 Railroad Productivity Index Based on a Cost Function

The multiple output production function of (1.1) can be estimated by our nonparametric methods, without assuming separability of outputs and inputs, which was necessary for Vinod (1968, 1976a) and similar attempts in the literature. The cost function of (1.2) is often estimated by fitting a system of share equations. The substantive economic interest is usually in estimating output elasticities, cost elasticities, etc. which are partial derivatives, and for which our methods are applicable..

The productivity index (1.5), or its discrete version (1.6) are based on time derivatives. Its estimation involves elasticites of total cost with respect to the j-th output, ($Eg:EY_j$). In Caves et.al. (1980) these elasticities were estimated from cross sectional study for 1955, 1963 and 1974, interpolated and extrapolated for the remaining years, and combined with the time series data reported there for 1951 to 1974. Our nonparametric methods may be used to avoid having to mix cross section results with time series, by directly estimating the elasticities ($Eg:EY_j$) evaluated at the mean of the data.

There are four outputs: freight ton miles, freight length of haul, passenger (trip) miles, and the length of trip; and five inputs: Labor, Way and structures, equipment, fuel, and materials. Generalized translog multi-product cost function of outputs and input prices is used to estimate elasticities of total cost with respect to the outputs after imposing Shepherd's Lemma and related conditions. Caves et al (1980, p. 175) find that the Hessian matrix is not negative semidefinite, althought the violation is not regarded as significant.

Our results for this example reported in Table 2 are claimed to be reasonable. For the ton-miles output Caves et. al's elasticity is 0.772 compared to our estimates of 0.361 (for r=2), and 0.367 (for r=1). For the passenger miles output Caves et. al's elasticity based on a cross sectional study is 0.201, which says that the railroads having one percent larger passenger mile output tend to have 0.20 percent higher costs. From Table 4 of Caves et al freight revenue as a share of total operating revenue increased from 0.906 in 1951 to 0.967 in 1974. Since passenger business is unimportant, the magnitude of their passenger elasticity (0.201) relative to freight elasticity (0.772) seems to be large. By contrast, our amorphous estimates of the passenger elasticities based on time series data are statistically significant, and negative: -0.039 (for r=2) and -0.038 (for r=1) meaning that if the passenger miles increased by one percent over time, the total cost will decrease slightly, by less than 0.04 percent. As passengers abandoned the railroads, the quality of passenger service deteriorated during this period, encouraging further decline in use of railroads by the travelling public. It is conceivable that the negative elasticity reflects the observable fact in the time series data that the passenger miles declined sharply, by more than five percent per year between 1951 and 1974, which may have permitted the railroads to curtail the schedules, thereby slightly reducing the total cost. On the other hand, it is possible that our time series estimate is no match for a detailed cross sectional study of Caves and others. Since Caves et. al. do not report the standard errors associated with their estimates in their Table 4, or goodness of fit etc., a comparison is inconvenient. We have also estimated the rate of change in railroad productivity with $E(g:Y_j)_t$ of (1.6) above replaced by the amorphous elasticities for each observation. Our average is 0.6% compared to Caves et al's 1.5% (Table 6). Further details about numerical results may be obtained from the authors upon request. In any case, we provide a simpler alternative which may be of interest in applications where the distinction between the interpretations of cross-sectional and time-series elasticities mentioned above is important.

4.2 A Simulations of Elasticity Estimation

We generate 100 observations on quantity produced by uniform random numbers between 0 and 1 multiplied by 100. The 100 observations on the costs are obtained by the square root of outputs. The average output is 55.4 units ranged between 0.5 to 98.1 units, with a sample standard deviation of 29.3528. The conditional expectation of costs at the mean output is 7.4782. Since costs are square roots of the outputs, we know the partial derivative of costs with respect to output is the reciprocal of twice the square root of output at the mean output. The correct value of the partial derivative should be 0.0672, whereas the amorphous estimate is 0.0666. This suggests that amorphous methods can yield reasonable estimates of elasticities, without assuming any knowledge about functional forms.

The estimates may not be as good for arbitrary forms having a fixed, cost of 50 and random errors. To illustrate this we use:

$$\text{cost} = 50 + \text{output}^{1/2} + \text{error} \qquad (4.4)$$

where error represents 0.00001 times unit normal random numbers. In this example, the partial derivative of cost with respect to output should be 0.0683, and the amorphous estimate is 0.0954. To consider the small sample case we use the first 30 observations based on (4.4) and find that sample mean output is 62.3002, the amorphous estimate (r=2) of partial derivative is 0.0581, whereas the correct value is 0.0633. The results for our method applied to White's (1980) simulation of CES production function are also encouraging. From simulations and other experience we conclude that amorphous estimates of partial derivatives from typical sample sizes in econometrics are in general agreement with our asymptotic results. More research is needed to develop further experience with larger models.

5. FINAL REMARKS

The production/cost function estimation research in recent years has increased our ability to obtain realistic estimates of elasticities and partial derivatives by using "flexible" functional forms which do not impose arbitrary restrictions on their values. However the flexibility was purchased at the cost of increased number of parameters leading to multicollinearity and other statistical difficulties. Since the nonparametric estimation can avoid some of the difficulties associated with multicollinearity, as well as linear parametric specifications, applied econometricians may find amorphous specification attractive in a variety of problems. One novelty here is that we have directly and analytically estimated the partial derivatives of conditional expectations, whereas other applications of nonparametric regression merely give conditional expectations or forecasts of the dependent variables. A theorem proved in Section 3 gives the asymptotic properties of the amorphous estimator.

The ability to estimate nonparametric partial derivatives is particularly useful in econometrics. Further study of nonparametric estimators seems to be worthwhile.

TABLE 2

Elasticities of Total Cost by Caves et al and by amorphous estimation

	Output Variable				Residual sum of squares
	Ton Miles	Average Length of Haul	Passenger Miles	Average Length of Trip	
Caves et al	0.772	-0.082	0.201	-0.031	
amorphous r=1	0.367	0.267	-0.038	-0.058	0.006772
coeff/SE ratio	6.718	3.604	2.304	2.090	
amorphous r=2	0.361	0.269	-0.039	-0.059	0.007413
coeff/SE ratio	6.974	3.828	2.500	2.242	

Notes: Caves et al's (1980) results are for 1951-74 average, from the bottom line of Table 4 in their paper. The residual sum of squares do not involve trimming because a comparable number is not reported by Caves et al.

REFERENCES

Barnett, W.A., 1983, Definitions of second order approximation and of flexible functional form, Economics Letters, 12, 31-35.

Barnett, W.A. and Y.W. Lee, 1985, The global properties of the minflex laurent, generalized Leontief, and translog flexible functional forms, Econometrica, 53, 1421-1437.

Bartlett, M.S., 1963, Statistical estimation of density function, Sankhya, Ser. A29, 373-382.

Bierens, H.J., 1985, Kernel estimators of regression functions, forthcoming in Advances in Econometrics.

Box, G.E.P., 1966, Use and abuse of regression, Technometrics, 8, 625-629.

Caves, D.W., L.R. Christensen and J.A. Swanson, 1980, Productivity in U.S. railroads, The Bell Journal of Economics, 11, 166-181.

Cacoullos, T., 1966, Estimates of a multivariate density, Annals of the Institute of Statistical Mathematics, 18, 176-189.

Christensen, L.R., D.W. Jorgenson and L.J. Lau, 1973, Transcendental logarithmic production frontier, Review of Economics and Statistics, 55, 28-45.

Diewert, W.E., 1971, An application of the Sheperd duality theorem: A generalized Leontief production function, Journal of Political Economy, 79, 461-507.

Elbadawi, I., A.R. Gallant and G. Souze, 1983, An elasticity can be estimated consistently without a priori knowledge of functional form, Econometrica, 51, 1731-1751.

Gallant, A.R., 1981, On the bias in flexible functional forms and an essentially unbiased form: the Fourier flexible form, Journal of Econometrics, 15, 211-245..

Gradshteyn, I.S. and I.M. Ryzhik, 1965, Table of integrals, series and products, New York: Academic Press.

Hardle, W., 1984, Robust regression function estimation, Journal of Multivariate Analysis, 14, 169-180.

McFadden, D., 1985, Specification of econometric models, Presidential address, econometric society, Fifth world congress, Cambridge, Mass.: Economics Dept., Mass. Institute of Technology.

Parzen, E., 1962, On the estimation of probability density and model, Annals of Mathematical Statistics, 33, 1065-1076.

Prakasa, Rao B.L.S., 1983, Nonparametric functional estimation, Orlando: Academic Press.

Rosenblatt, M., 1956, Remarks on some nonparametric estimates of density, Annals of Mathematical Statistics, 27, 832-837.

Schuster, E.F., 1972, Joint asymptotic distribution of the estimated regression function at a finite number of distinct points, Annals of Mathematical Statistics, 43, 84-88.

Singh, R.S., 1981, Speed of convergence in nonparametric estimation of a multivariate mu-density and its mixed partial derivatives," Journal of Statistical Planning and Inference, 5, 287-298.

Singh, R. S., A. Ullah and R. A. L. Carter, 1985, Nonparametric inference in econometrics: new applications in I. MacNeill and G. Umphrey (eds.) Time series and econometric model, D. Reidell.

Stone, C., 1977, Consistent nonparametric regression, Annals of Statistics, 5, 549-645.

Stone, C., 1984, An asymptotically optimal window selection rule for kernel density estimates, Annals of Statistics, 12, 1285-1297.

Sudit, E.F., 1973, Additive nonhomogeneous production functions in telecommunications, Bell Journal of Economics, 4, 499-514.

Ullah, A., 1985, Specification analysis of econometric models, Journal of Quantitative Economics, 1, 187-210.

Ullah, A., and R.S. Singh, 1985, The estimation of probability density functions and its applications in econometrics, Economics Discussion Paper, Dept. of Economics, University of Western Ontario, London, Ontario, Canada.

Varian, H., 1984, The nonparametric approach to production analysis, Econometrica, 52, 579-597.

Vinod, H. D., 1968, Econometrics of joint production, Econometrica, 36, 322-336.

Vinod, H.D., 1972, Non homogeneous production functions and applications to telecommunications, Bell Journal of Economics, 3, 531-543.

Vinod, H.D., 1976a, Canonical ridge and econometrics of joint production, Journal of Econometrics, 4, 147-166.

Vinod, H.D., 1976b, Application of new ridge regression methods to a study of Bell system scale economies, Journal of the American Statistical Association, 71, 835-841.

Vinod, H.D. and A. Ullah, 1981, Recent advances in regression methods, New York: Marcel Dekker.

White, H., 1980, Using least squares to approximate unknown regression functions, International Economic Review, 21, 149-170.