Western University
Scholarship@Western

Centre for Decision Sciences and Econometrics Technical Reports

Economics Working Papers Archive

1987

Nonparametric Estimation of Econometric Functionals

Aman Ullah

Follow this and additional works at: https://ir.lib.uwo.ca/economicscdse_tr

Part of the Economics Commons

Citation of this paper:

Ullah, Aman. "Nonparametric Estimation of Econometric Functionals." Centre for Decision Sciences and Econometrics Technical Reports, 17. London, ON: Department of Economics, University of Western Ontario (1987).

ISSN: 0831-4454
ISBN: 0-7714-0922-2

THE CENTRE FOR DECISION SCIENCES AND ECONOMETRICS

Nonparametric Estimation of Econometric Functionals

Aman Ullah

TECHNICAL REPORT NO. 17
JULY 1987

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

Department of Economics Library
NOV 1 8 1987
University of Western Ontario

11731

NONPARAMETRIC ESTIMATION OF ECONOMETRIC FUNCTIONALS

by

Aman Ullah*

University of Western Ontario

June, 1987

Department of Economics Library
NOV 1 8 1987
University of Western Ontario

NONPARAMETRIC ESTIMATION OF ECONOMETRIC FUNCTIONALS

ABSTRACT

In this paper we have reviewed and explored the nonparametric density estimation approach for analyzing various econometric functionals. The applications of density estimation have been emphasized in the specification, estimation, and testing problems arising in econometrics. Some limitations of the nonparametric approach are also examined, and potential future areas of applied and theoretical research have been indicated.

1. INTRODUCTION

Since the introduction of econometrics as a subject, almost four decades ago, the estimation and testing of econometric models has been carried on under many strong parametric assumptions. For example, consider an economic model

$$y = M(x) + u$$

where y is a dependent variable, x is a vector of regressors, u is the disturbance and $M(x) = E(y|x)$ is an unspecified regression function; i.e. the conditional mean of y given x. Then the first assumption of parametric econometrics is that the form of M(x) (usually linear) is known. Secondly, forms of other econometric functions, such as the conditional variance of y given x and the autocorrelation of u, are also specified. Third, the parametric joint density (data generating process) of y and x is assumed to be, usually, normal. In addition, the x's in many econometric studies are simply considered to be nonstochastic when, in fact, they are stochastic. A serious disadvantage of parametric econometrics based on these assumptions is that it may not be robust to any slight inconsistency between the data and the particular parametric specification. Indeed, it is well known that the true functional form of the model M(x) and of the other econometric functions are, rarely, if ever, known and any misspecification in this regard may lead to erroneous conclusions. Under this scenario one might consider an alternative procedure for inference (estimation and testing). Such an alternative is data-based nonparametric modelling and inference, which is the theme of this paper.[1]

The need for a nonparametric approach to the economics of production and consumption is implicit in the important work of Afriat (1967) and Hanoch and Rothschild (1972). The initiation of this approach, however, goes back to Samuelson's (1938) introduction of revealed preference analysis. The idea has been further developed and forcefully advocated in the recent important papers by Diewert and Parkan (1978) and Varian (1984). These papers deal with nonparametric tests of, among other things, the consistency of the data with the maximization principle, homothecity, and various forms of seperability, without any functional form specification for M(x). However, a basic problem with this work is that it neither provides any econometric example nor provides an econometric framework for their proposed tests. In addition, the stochastic nature of y and x's is ignored, although the recent work of Varian (1985) and Epstein and Yatchew (1985) attempts to remove some of these difficulties. In these papers tests are given for the null hypothesis that the regression function M(x) lies in a particular compact family of functions which need not necessarily be a parametric family. It has been indicated that the common hypotheses of economic theory can be formulated in this general testing framework. However, these tests are based on the assumption that the disturbance variance is known, so it is not clear how their procedure could be implemented in practice.

A related, but different, approach is the approximation of unknown functions M(x) by flexible functional forms. The first important papers in this respect were by Diewert (1971) and Christensen et al (1973). A drawback of this literature is that it provides local Taylor series approximations which may be poor approximations on non-local sets. Recently, Gallant (1981, 1982), Gallant and Golub (1984), and Barnett and Lee (1985) among others, have suggested various useful non-local approximations of M(x). For example, while Barnett and Lee (1985) have suggested that the miniflex Laurent series expansions are a useful alternative to explore, Gallant's (1981) pioneering work introduced a general Fourier series expansions. Building on this work, Elbadawi et al (1983) demonstrate how one can estimate elasticities without knowing the functional form, and Gallant (1982) shows how the assumptions such as homotheticity and constant returns to scale can be tested by imposing restrictions on Fourier flexible approximations. This is in contrast to the work of Varian (1985) and Epstein and Yatchew (1985) where a revealed preference approach is followed. However, Gallant's test procedure is also based on the assumption of known variance of the disturbance u. Further while fourier approximations are superior to Taylor appoximations in many respects, they are heavily parametrized.

Another recent development is the semi-parametric estimation of economic models. In general, in this literature one basically estimates parametric models with less restrictive distributional assumptions. This is particularly popular in the context of estimating models with limited dependent variables. Important developments are: the maximum score estimator (Manski (1975)), the distribution-free maximum likelihood estimator (Cosslett (1983)), smoothing splines estimation (Engle et al (1986)); and the recent work by Gallant and Tauchen (1987), Horowitz (1987), Newey (1987) and Powell (1986). However, a problem with the semi-parametric approach is that it again requires limited parametric specification of certain functions.

On the other hand the data based, purely nonparametric, technique of specification and testing to be considered here is based on the specification of the unknown M(x) through its nonparametric estimation. This nonparametric specification of M(x) is truly flexible in Diewert's (1971) sense; that is, it attains arbitrary levels of first and second order derivatives at a predetermined point. The essential ingredient for the nonparmetric estimation of M(x) is the direct estimation of the unknown joint density of x and y. The first published paper in this area is due to Rosenblatt (1956). Since then a vast amount of literature has appeared on this subject in various statistics journals, including the recent work by Prakasa Rao (1983), Devroye and Györfi (1985) and Silverman (1986). However, despite the flow of articles by statisticians on density estimation over the past thirty years, particularly in the past decade, very little has been done to apply density estimation to econometrics; although see recent attempts by McFadden (1985), Bierens (1985), Ullah (1985), Ullah and Singh (1985) and Robinson (1986b). There are, perhaps two reasons for this gap. First, an important reason is that the statistical literature is highly technical, and has thus created a wide impression that density estimators are of only theoretical significance. This becomes clear when one notices that there are hardly any significant applied papers published which employ density estimates. Another important reason for the lack of application of density estimation in applied statistics, and in particular among econometricians, is the fact that the statistics literature has not dealt with nonparametric estimation and testing of functionals which are of primary interest to these practitioners., e.g. response functions, elasticities, second order partial derivatives and average partial derivatives of the regression function.

The aim of this paper is to review and develop a nonparametric approach to estimation and testing without any functional form assumption about M(x). With this in view, section 2 presents the functionals of interest in econometrics, and motivates the need for density estimation. Section 3 then briefly reviews the statistical literature on density estimation, and provides nonparametric estimators of various functionals presented in section 2. Section 4 provides the consistency, asymptotic normality and rate of convergence of various nonparametric estimators. Some practical limitations of the nonparametric approach are also described. Finally, in section 5 we present illustrative examples. It is hoped that the simplicity of the approach, and the availability of good data and computing facilities will soon make nonparametric econometrics a useful alternative to the usual parametric econometrics.

2. ECONOMETRIC FUNCTIONALS

Suppose we have n independent and identically distributed observations $(y_i, x_{i1}, \ldots, x_{ip})$, i=1,...,n from an absolutely continuous p+1 variate distribution with density $f(y, x_1, \ldots, x_p) = f(y,x)$. Here we consider y to be a dependent variable and x to be the vector of p regressors. If $E|y| < \infty$ then the conditional mean of y given x exists and it takes the form

(2.1) $E(y|x) = M(x)$

where M(x) is a real valued function of x. M(x) is called the regression function and it provides a formulation for the regression model as

(2.2) $y = M(x) + u$

where, by construction, the disturbance term u is such that $E(u|x) = 0$. Now our aim is to estimate the unknown regression function $M(x)$, and other unknown functionals which are encountered in the econometric analysis of the model (2.2) without making explicit assumptions about their functional forms. Below, we first present various functionals and then provide their estimates in Section 3.

<u>The Regression Function</u> (Conditional Mean). For the joint density $f(y,x)$ the regression function in (2.1) can be written as

$$(2.3)\quad E(y|x) = M(x) = \int y \frac{f(y,x)}{f_1(x)} dy$$

where $f_1(x)$ is the density of x marginal to $f(y,x)$. The true form of the regression function $M(x)$ can be determined if the true joint density is known. For example, if $f(y,x)$ for $p=1$ is a bivariate normal density then it is well known that $M(x)$ is linear, i.e., $M(x) = \alpha+\beta x$ where $\alpha = Ey - \beta Ex$ and $\beta = cov(x,y)/v(x)$. However since the joint density is rarely, if ever, known the true form of $M(x)$ is generally unknown. Under this scenario, in parametric econometrics, one often assumes various forms for $M(x)$. Also, ad hoc specifications of the conditional expectations of the type (2.3) are used in the parametric rational expectations models. But it is now well known that any misspecification in $M(x)$ has serious consequences for econometric inference and policy evaluation. For example, as a consequence of misspecification the estimators of the regression parameters can be seriously biased.[2] Also in a testing situation, the true rejection probability of a test can exceed its nominal rejection probability. Thus the knowledge about the true form of $M(x)$ or its consistent approximation is of utmost importance in econometrics.

<u>The Response Function</u> (Regression Coefficients). The estimation of regression parameters is one of the main objectives of econometrics analysis. The response or regression coefficient of y with respect to changes in a regressor, say $x_j (j=1,\dots,p)$, is defined as the partial derivative of $M(x)$ with respect to x_j. It is denoted by $\beta_j(x) = \beta(x)$ where

$$(2.4) \quad \beta(x) = \frac{\partial M(x)}{\partial x_j} = \lim_{h\to 0} \frac{M(x+h/2) - M(x-h/2)}{h}$$

where $M(x+h/2) = M(x_1,\dots,x_j + h/2,\dots,x_p)$.

Note that (2.4) is a varying response coefficient since it is a function of x. The fixed response coefficient can be defined as $\beta(\bar{x})$, i.e., $\beta(x)$ evaluated at $x = \bar{x} = (\bar{x}_1,\dots,\bar{x}_p)$. Alternatively, one can define $\bar{\beta} = E\beta(x)$, the average derivative, as the fixed response coefficient.

The question here is, how do we determine $\beta(x)$ without specifying the form of $M(x)$?

<u>Curvature</u> (Higher Order Derivatives of $M(x)$). Economic theory often imposes theoretical curvature conditions (concavity or convexity) on $M(x)$. For example, the expenditure function is considered to be concave, and the profit function convex in price. If the true form of $M(x)$ is known these assumptions can be verified by considering

$$(2.5) \quad C(x) = \frac{\partial}{\partial x_j}\beta(x) = \frac{\partial^2}{\partial x_j^2}M(x) = \lim_{h\to 0} \frac{\beta(x+h/2) - \beta(x-h/2)}{h}$$

and seeing whether $C(x)$ is positive or negative.

The question is whether or not the unknown M(x) satisfies the curvature conditions implied by economic theory.

<u>Heteroskedasticity Function</u> (Conditional Variance). The conditional variance of y given x can be written as

$$(2.6)\quad V(y|x) = V(x) = \int y^2 \frac{f(y,x)}{f_1(x)} dy - (E(y|x))^2$$

where $f_1(x)$ and E(y|x) are as in (2.3). The quantity V(x) by itself is quite often of interest in economics; for example, the question of variability in inflation given the past information set. Also in many economic models, especially finance models, the conditional variance of the type (2.6) appears as an unobservable variable; e.g. risk premiums. In many other econometric problems one is interested in analysing the conditional variance V(u|x), where u = y-E(y|x) is the disturbance term as in (2.2). This is known as the heteroskedasticity problem.

Again, the question is, what is the form of V(x)?

<u>Covariance Function</u> (Autocorrelation). The covariance of y and w given x can be written as

$$(2.7)\quad \mathrm{cov}(y,w|x) = \gamma(x) = \int yw \frac{f(y,w,x)}{f_2(x)} dydw - E(y|x)\ E(w|x)$$

where $f_2(x)$ is the density of x marginal to f(y,w,x). In many macroeconomic models the quantity γ(x) may appear as one of the regressors, see, e.g. Mascaro and Meltzer (1983). In the context of the econometric model (2.2), however, one might be interested in $\mathrm{cov}(u_i,u_{i-1}|x)$, i=2,...,n, autocovariance function of the disturbance u. This is essentially (2.7) with $y = u_i$ and $w = u_{i-1}$. The question is what is the form of γ(x).

Disturbance Density Function. Consider $u = y - M(x)$ as in (2.2), and its density function as $f(u)$. Then information about the form of $f(u)$ itself is of considerable interest in understanding the behavior of the random part of the model (2.2).

The above are some of the commonly encountered functions in econometrics. There are many other such functions, for example: hazard functions, entropy functions, score functions, discriminant functions. The list is large and the discussion of each one of them requires a separate study and thus is beyond the scope of this paper. One common element in these and in the functions above, however, is that they all depend on the unknown joint density. This suggests that the questions about their forms can be answered by estimating the joint density f and its marginal and conditional densities.

We also note here that not only do the above mentioned functions but even the hypotheses generated by economic theory depend on the unknown joint density. For example if $M(x)$ in (2.2), for $p=2$, is the production function, then the hypothesis of constant returns to scale is $\partial M(x)/\partial x_1 + \partial M(x)/\partial x_2 = 1$ or $\beta(x_1) + \beta(x_2) = 1$ which depends on the unknown joint density since $M(x)$ does so. In parametric econometrics one often uses the Cobb-Douglas or translog approximations for $M(x)$, which may have nothing to do with the unknown joint density, and then tests the hypothesis $\beta(x_1) + \beta(x_2) = 1$. In the nonparametric case one can test this hypothesis by directly estimating the joint density and hence $\beta(x) = \partial M(x)/\partial x$. Thus density estimation is basically a more direct way of dealing with econometric estimation and testing problems.

Before proceeding to the estimation of the density in Section 3, we note that the functions discussed above are not confined to the regression model (2.2). In fact, although not considered here explicitly, the extension to dynamic models with stationary variables, system models, and simultaneous equations models, are straightforward. For example, for the first order autoregressive model the regression function in (2.1) becomes $E(y|y_{-1})$ where y_{-1} is the lagged value of the stationary variable y. Similarly for the system model y will be a vector of say q endogenous variables and $E(y|x)$ will be a vector regression function. We restrict ourselves to the model (2.2) for the sake of simplicity in exposition.

3. KERNEL ESTIMATION OF ECONOMETRIC FUNCTIONALS

In statistical analysis the idea of density estimation was first explored in the unpublished work of Fix and Hodges (1951). They introduced the following <u>naive estimator</u> of the density. Consider the density function of a random variable X at a point x as

$$(3.1) \quad f(x) = \lim_{h\to o} \frac{1}{h} P(x-h/2 < X < x+h/2).$$

Let $x_1,\dots,x_n$ be the sample observations. Then we can estimate $P(x-h/2 < X < x + h/2)$ by the proportion of the sample falling in the interval $(x-h/2,\ x+h/2)$. Thus an obvious estimator of the density f(x) is given by

$$(3.2) \quad f_n(x) = \frac{1}{nh} [\text{number of } x_1,\dots,x_n \text{ in } (x-h/2,\ x+h/2)],$$

where h is chosen to be a small number. Alternatively, we can write (3.2) as

$$(3.3) \quad f_n(x) = (nh)^{-1} \sum_{i=1}^{n} W\left(\frac{x_i - x}{h}\right),$$

where W is the weight function such that

$$(3.4)\quad W(z) = \begin{cases} 1 & \text{if } -\tfrac{1}{2} < z < \tfrac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\int W(z)dz = 1$.

The first published paper to deal explicitly with the density estimator was Rosenblatt (1956). He generalized (3.3) by replacing W(x) with a real positive kernel function K satisfying $\int K(x)dx = 1$. His general <u>"kernel" estimator</u> is

$$(3.5)\quad f_n = f_n(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$$

where h, the window-width (also called the smoothing parameter or band-width), is a positive function of the sample size n which goes to zero as $n \to \infty$. Usually, but not always, K will be a symmetric density function, e.g., the normal density. The choices of h and K, and the asymptotic properties of the kernal estimator are discussed in Section 4. Whittle (1958), independently of Rosenblatt, formulated the general weight function (Bayesian) class of estimators, and Parzen (1962) extended Rosenblatt's estimator to cases where the weight function need not be non-negative.

A few remarks about kernel estimations are in order. Observe from (3.4) that while the naive estimator can be considered as a sum of 'rectangles' of width h and height $(nh)^{-1}$ centered at the data points, the kernel estimator is a sum of curves placed at the data points. The kernel K determines the shape of the curves and the window width h determines their width. Another point to be noted is that so long as K is everywhere non-negative and satisfies $\int K(x)dx = 1$, f_n will be a probability density. Furthermore, f_n will possess all the continuity and differentiability properties of the kernel K. Note, however, that the naive estimator (3.3) is not continuous, but has jumps at the points $x_i \pm h/2$ and has zero derivative everywhere else. The generalization to the kernel estimator f_n in (3.5) overcomes this difficulty.

Since the published work of Rosenblatt (1956) several other density estimation techniques have appeared in the literature. Some of these are: the orthogonal series (Cencov (1962)), the nearest neighbourhood (Loftsgaarden and Quesenberry (1965)), the maximum likelihood (Wegman (1970)), the maximum penalized likelihood (Good and Gaskins (1971)), the histogram (Van Ryzin (1973)); see Prakasa Rao (1983) for details. Apart from the histogram, the kernel estimator is probably the most widely used estimator and is perhaps the most thoroughly studied in the statistics literature. This is the estimator that will be considered throughout this paper. We note, however, that although the histogram is a useful method of density estimation, it has the drawbacks of being discontinuous, sensitive to the origin and to the width of the class interval, and extremely complicated for two and more variables. The kernel estimator too suffers from a drawback. Since the window width is fixed across the data points, the estimates in the tails may show slight spurious bumps. Attempts to further smooth the tails (by changing h) may distort the middle part of the distribution. A better alternative is to consider the kernel estimator f_n with the window width varying across the data points; i.e.

$$(3.6)\quad f_n^*(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_{ni}}K\left(\frac{x_i-x}{h_{ni}}\right),$$

where h_{ni} is the distance between x_i and its k_n-th (a positive integer) nearest point in the remaining n-1 data points. The choice of a positive integer k_n determines how responsive the window width is to very local detail. A rule of thumb choice is $k_n \approx n^{1/2}$. Varying h_{ni} in this way ensures that data points in the regions where there are fewer observations will have flatter kernels. The estimator (3.6) is known as the variable kernel estimator (VKE), see e.g. Breiman, et al. (1977).

The estimator (3.6) is related to the nearest-neighbour estimator (NNE) which can be thought of as (3.6) with h_{ni} replaced by $h_n(x)$, the distance of x from its k_n-th nearest neighbour among $x_1, \ldots x_n$. See, for example, Mack and Rosenblatt (1979). Note that in the NNE the window widths depend on the point x at which the density is estimated; in the VKE the window widths are independent of the point x. Furthermore, while the VKE will itself be a probability density function this is not true for the NNE.

A useful alternative to the VKE and the NNE is the adaptive two-stage estimator (A2SE) considered by Breiman et al. (1977) and Abramson (1982). The A2SE is essentially (3.6) with $h_{ni} = h\delta_{ni}$, $\delta_{ni} = [(f_n(x_i)/G]^{-\lambda}$ where G is the geometric mean of $f_n(x_i)$ over all x_i, $0 < \lambda \leq 1$ is a sensitivity parameter and $f_n(x_i)$ is any convenient initial kernel or NNE. Brieman et al. suggested $\lambda = d^{-1}$, where d is the dimensionality of the space in which the density is being estimated. However, the numerical results of Abramson (1982) and others suggest that $\lambda = ½$ gives good results, and that for this value of λ the approximate bias is smaller than that of the kernel estimator f_n in (3.5).

When h_{ni} in (3.6) is replaced by positive smoothing factors h_i, then the estimator (3.6) can be thought of as the recursive estimator (RE). It was introduced by Wolverton and Wagner (1969) and independently by Yamato (1971). The $f_n^*(x)$, for $h_{ni} = h_i$, is recursive in the sense that when an additional data point becomes available it can be updated according to

$$(3.7) \quad f_{n+1}^*(x) = (n+1)^{-1}[nf_n^*(x) + h_{n+1}^{-1} K(\frac{x_{n+1} - x}{h_{n+1}})].$$

This estimator is especially useful in the context of time series analysis. An extensive study comparing the performances of RE, f_n, VKE, NNE and A2SE would be a useful subject of future research, although see Wegman (1972) and Kumar and Markmann (1975) for Monte Carlo Studies regarding performances of the histogram, kernel and orthogonal series estimators.

The estimator f_n in (3.5) was first generalized to the estimation of multivariate density functions by Cacoullos (1966). As in Section 2, let $z = (y,x)$ be the vector of $p+1=q$ random variables. Then the kernel estimator of the density $f(z)$ is the following straightforward generalization of (3.5)

$$(3.8)\quad f_n(y,x) = f_n(z) = \frac{1}{nh^q}\sum_{i=1}^{n} K\left(\frac{z_i - z}{h}\right).$$

Furthermore, the kernel estimator of the marginal density $f_1(x)$ is

$$(3.9)\quad f_{1n}(x) = \int f_n(z)dy = \frac{1}{nh^p}\sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h}\right)$$

where $K_1(x) = \int K(z)dy$ such that $\int K_1(x)dx = 1$ and x is a vector of p random variables. The estimator of the conditional density is then

$$(3.10)\quad f_n(y|x) = \frac{f_n(z)}{f_{1n}(x)}.$$

Similarly, the extensions of the recursive estimator, VKE, NNE and A2SE to the multivariate case are straightforward.

The use of a single h for all the q variables in z may not be appropriate if the variation in one of z's is much greater than in the others. In these situations, it may be more appropriate to use a vector or matrix of window width parameters, see e.g. Singh et al. (1987) and Singh and Ullah (1986). An attractive practical approach is to linearly transform the data to have unit covariance matrix; use (3.8) for the transformed data and finally transform back to the original metric.

3.1 Estimation of Econometric Functionals

In Section 2 we discussed how various econometric functionals depend on the unknown joint density function, and how in parametric econometrics the analysis is carried out by assuming various forms for these functionals or by implicitly assuming the form of the joint density. Here we present the nonparametric estimators of the functionals (2.3) to (2.7) by directly substituting into them the joint density estimator $f_n(y,x)$ from (3.8) and the marginal density estimator $f_{1n}(x)$ from (3.9). These nonparametric estimators, after some simplifications, are given below and their asymptotic properties are presented in the following section.

First of all consider the estimation of the regression function in (2.3). Substituting (3.8) and (3.9) into (2.3) it can easily be verified that

$$(3.11)\quad M_n = M_n(x) = \int y \frac{f_n(y,x)}{f_{1n}(x)} dy = \sum_{i=1}^{n} y_i r_i(x)$$

where $r_i(x) = K_1\left(\frac{x_i - x}{h}\right) / \sum_1^n K_1\left(\frac{x_i - x}{h}\right)$. M_n is known as the Nadaraya (1964) and Watson (1964) type regression function estimator.

Substitution of M_n into (2.2) gives the nonparametric model

$$(3.12)\quad y = M_n(x) + e.$$

where e is the nonparametric residual. We observe that this nonparametric model (3.12) has been obtained without making assumptions about the functional form of M_n, the joint density of y and the x's, or the non-stochastic behavior of the x's. Furthermore, since (3.12) has been obtained by estimating the joint density it does not require the assumption of weak exogenity of Engle et al (1983). Thus the nonparametric model is free from some of the basic assumptions required in parametric econometrics. Also, it avoids the specifications of any local or non-local flexible forms in production or consumer economics.

The model in (3.12) is useful for nonparametric forecasting. Also, since $M_n(x)$ is the conditional expectation, it provides a way out of specifying expectation variables appearing in various econometric models.

For the estimation of the <u>response function</u> $\beta(x)$ in (2.4), we take a straightforward partial derivative of $M_n(x)$ in (3.11) with respect to x_j and write

$$(3.13)\quad \beta_n^*(x) = \frac{\partial M_n(x)}{\partial x_j} = \sum_{i=1}^{n} y_i(r_{1i} - r_{2i})$$

where, for $w_i = (x_i - x)/h$, $r_{1i} = K'(w_i)/\sum_1^n K(w_i)$, $r_{2i} = K(w_i)\sum_1^n K'(w_i)(\sum_1^n K(w_i))^{-2}$ and $K'(w_i) = \partial K(w_i)/\partial x_j$.

The estimator $\beta_n(x)$ is the <u>varying response</u> or <u>varying regression coefficient estimator</u> of $\beta(x)$. Note that this nonparametric estimator has been obtained without specifying any ad hoc functional form for $M(x)$. An alternative form of $\beta_n(x)$ can be written, from (2.4), as

$$(3.14)\quad \beta_n(x) = \frac{M_n(x+h/2) - M_n(x-h/2)}{h}$$

where $M_n(x+h/2) = M_n(x_1,\ldots,x_j+h/2,\ldots,x_n)$. The estimator $\beta_n(x)$ has been studied in Rilstone (1985) and Rilstone and Ullah (1986), while the estimator in (3.13) has been used in Vinod and Ullah (1985).[3] Typically $\beta^*_n(x) \simeq \beta_n(x)$.

The <u>fixed response</u> or <u>fixed regression coefficient estimator</u> can be obtained by calculating $\beta_n(\bar{x})$, where $\bar{x}$ is the average of the sample observations. Alternatively, one can calculate either the average response coefficient $\bar{\beta}_n(x) = n^{-1}\sum_1^n \beta_n(x)$ or a weighted average response coefficient. In an important work Powell et al (1986) show that the weighted average response coefficient estimators have $n^{1/2}$ speed of consistency and normality compared to the slower speed of $(nh^{p+2})^{1/2}$ (because $h \to 0$ as $n \to \infty$) for $\bar{\beta}_n(x)$. However, in many practical applications we have found that $\beta_n(\bar{x})$ provide much more robust estimates compared to the average estimators. This is perhaps because, as mentioned earlier in Section 3, the kernel estimators do not perform well in the tails. Thus a better estimator in such situations would be to consider the median of $\beta(x)$ or any other Huber-type robust estimator. An alternative would be to calculate the average estimators by using robust regression function (M(x)) estimators in Hardle (1984).

Testing any economic restrictions can be easily carried out by using the estimator $\beta_n(\bar{x})$, and its standard error given in Section 4. This includes testing for separability and homotheticity in production and consumer economics. In addition, the estimator $\beta_n(x)$ provides a natural way to analyse and test the stability of time series models.

The estimator of <u>curvature</u> C(x) in (2.5) is also straightforward; i.e.

$$(3.15)\quad C_n(x) = \frac{\partial \beta_n(x)}{\partial x} = \frac{\beta_n(x+h/2) - \beta_n(x-h/2)}{h}$$

where $\beta_n(x)$ is as given in (3.14). Alternatively, an explicit expression of $C_n(x)$ can be written by using (3.13). For a point estimate one can calculate $C(\bar{x})$ or the average $n^{-1}\sum_1^n C(x_i)$. Using (3.15), we can empirically analyse the curvature conditions implied by economic theory without any a priori specification of the form of $M(x)$. This is especially useful in the literature on production and consumption (see section 1) where the commonly used flexible functional forms usually fail to satisfy the appropriate theoretical curvature conditions. For applications see McMillan et al (1986) and Rilstone (1985).

We now consider the estimation of the heteroskedastic conditional variance in (2.6). Using (3.8) and (3.9) the estimator of the conditional variance of y given x can be obtained as

$$(3.16)\quad V_n(x) = \sum_{i=1}^{n} y_i^2 r_i(x) - M_n^2(x)$$

where $M_n(x)$ and $r_i(x)$ are as given in (3.11). This estimator is itself of interest, e.g. in analysing the variability in an economic variable. The application of V_n has appeared in Pagan and Ullah (1985) for the risk-premium models in macro economics.

The estimator (3.16) also provides a way to analyse heteroskedasticity in econometric models without any assumption about their functional forms.[4] This estimator follows from the results in Singh and Tracy (1977) and is also given in Rose (1978). Carroll (1982) considered (3.16) for the generalised least squares (GLS) estimation of a parametric linear regression model with the unknown form of the heteroskedasticity. Some improvements over Carroll's (1982) results for consistency and rate of convergence have appeared in Carroll et al (1986), and Robinson (1986a) where the nearest neighbourhood estimator has been used to develop V_n. Singh et al (1987) provide an application of the estimator V_n for the heteroskedasticity problem.

Notice that the Carroll (1982) and Robinson (1986a) approaches are semiparametric in the sense that the parametric form of the regression function M(x) is first specified and then $V_n^{-1/2}$ is used to transform y and M(x) to obtain the GLS estimators of the regression parameters. A purely nonparametric approach would be to take M_n in (3.11), transform y and M_n by the $V_n^{-1/2}$ and then use β_n in (3.13) for this transformed y and M_n. The β_n so obtained is a new nonparametric GLS estimator under the unknown form of the heteroskedasticity. The asymptotic properties of this estimator require further work.

Now we consider the nonparametric estimator of the covariance function. This can be obtained by substituting expressions like (3.8) and (3.9) into (2.7);

$$(3.17)\quad \gamma_n(x) = \sum_{i=1}^{n} y_i w_i r_i^*(x) - M_n(x)M_n^*(x)$$

where $r_i^*(x) = K(\frac{z_i - z}{h}) / \sum_{i=1}^{n} k_2(\frac{x_i - x}{h})$ and now z = (w, y, x). The sample correlation function is then $\rho_n(x) = \gamma_n(x)/[V_n(x)V_n^*x)]^{1/2}$, where $V_n^*(x)$ is $V_n(x)$ with $y_i = w_i$. If w_i is the lagged value of y_i then $\rho_n(x)$ is the nonparametric estimator of the autocorrelation function.

Finally, for the estimation of the density of the disturbance u, we note that the nonparametric estimator of u in (2.2) is $e = y - M_n(x)$ as given in (3.12). Thus, using (3.5), the estimator of the density of e is

$$(3.18)\quad f(e) = (nh)^{-1} \sum_{i=1}^{n} K(\frac{e_i - e}{h}).$$

Note that $e = u + M(x) - M_n(x) \approx u$ since $M_n(x)$ tends to $M(x)$ for large n (see section 4.3). Thus the graph of (3.18) can provide useful information about the shape of the distribution of u; in particular, it can indicate departures from normality. If one has the least squares residuals based on a parametric model, they can be used for e in (3.18). However, the nonparametric residual e is robust against misspecified functional forms and thus, in large samples, its use in place of the usual least squares residual should generally give better results. Further, for large samples, nonparametric residuals can be used for the testing problems in Yatchew (1987), and in the various econometric diagnostic tests discussed in Pagan (1983). For small samples, little is known about the behavior of nonparametric residuals and the tests based on them.

We observe that the sum of the nonparametric residuals e is not zero. If, however, the model has an intercept, α, it can be estimated by the least squares as $\bar{y} - \bar{M}_n(x)$. The nonparametric residual e from this model will then sum to zero. Similar adjustments can be made when the model has dummy or trend variables. For example, if the model has one dummy variable its coefficient can be estimated from the least squares regression of $y-M_n(x)$ on the dummy variable. The properties of kernel estimators with discrete and continuous variables are given in Bierens (1985). However, not much is known about the properties of the residual variance based on the nonparametric residuals e, especially in the time series case.

Various other functions, mentioned in Section 2, which depend on the unknown density function can also be estimated by the nonparametric method. Some of these are: quantile functions (Parzen (1979)), hazard functions (Watson and Leadbetter (1964), Tanner and Wong (1984)), entropy and information matrix (Singh (1977)). The details of these and some other statistical applications can be found in Prakasa Rao (1983). At present, the applications to econometrics are, however, rare; although see Bierens (1985), Robinson (1986b) and Ullah (1985) for specification issues, Singh et al (1987) and Power and Ullah (1986) for Monte Carlo based finite sample econometrics, Robinson and Ullah (1987) for the simultaneous equations model, Vinod and Ullah (1985) for production economics, McMillan et al (1986) and Hildenbrand and Hildenbrand (1985) for consumer economics, and Stock (1985) for econometric policy evaluation. The areas for future applications are numerous: perhaps one can analyse any econometric issue by nonparametric methods. Some potential areas are: estimation of rational expectations and nonlinear simultaneous equations models, causality analysis, non-nested hypothesis testing, forecasting and estimation of response surfaces in Monte Carlo studies; stochastic dominance analysis, income distribution and expected utility theory; estimation of Euler equation.

4. SOME FINITE SAMPLE AND ASYMPTOTIC PROPERTIES

4.1 Results for Kernel Density Estimator[5]

There is an enormous literature on the proofs of the asymptotic properties of density estimators, the details of which can be found in the work of Prakasa Rao (1983) and Devroye and Györfi (1985). This statistical literature is highly technical and so we do not discuss it here. Instead our aim is to describe the basic assumptions and asymptotic results which may provide some understanding of the large sample behaviour of the kernel estimators and of the limitations of nonparametric approach in applied work. In addition, we present the asymptotic standard errors and normality results for various functionals considered in Section 3. These will be useful for applied work, and they are used in Section 5.

A point to be noted here is that most of the work on density estimation is for the i.i.d. case; although some work has appeared in the time dependent but identically distributed case. We first consider the i.i.d. case and state various assumptions.

The asymptotic properties of the density estimators can be established under some regularity assumptions about the kernel K and the density f. We also require that the window width $h=h_n$ depend on the sample size n in some way. The assumptions we make are:

A.1 Let $\mathcal{K}$ be the class of all Borel-measurable bounded real valued functions $K(z)$, $z=(z_1,\ldots,z_q)'$ such that

(i) $\int K(z)dz = 1$ (ii) $\int |K(z)|dz < \infty$ (iii) $\|z\|^q\, |K(z)| \to 0$ as $\|z\| \to \infty$ (iv) $\mathrm{Sup}|K(z)| < \infty$, where $\|z\|$ is the usual Euclidean norm of z.

A.2 $h_n \to 0$ as $n \to \infty$

A.3 $nh_n^q \to \infty$ as $n \to \infty$ (or $(nh_n^q)^{-1} = o(1)$)

A.4 $f(z)$ is continuous at any point z_o

We observe that assumption A.1 is satisfied by a large class of functions; for example, the q variate standard normal density, and the function

$$K(z_1,\ldots,z_q) = 2^{-q} \prod_1^q I(z_j),\text{ where } I(z_j) = 1 \text{ if } -1 < z_j < 1 \text{ and } 0$$

otherwise. Furthermore, the assumptions A.2 and A.3 imply that as n increases h should decrease but in a way such that nh^q is still large; see the denominitor of the last term in (3.8).

While the pointwise asymptotic unbiasedness of f_n ($\lim Ef_n \to f$ as $n \to \infty$) follows only under the assumptions A.1, A.2 and A.4, the pointwise weak consistency of f_n ($f_n \to f$ in probability, at any point, as $n \to \infty$) follows under the assumptions A.1 to A.4. The pointwise strong consistency, i.e., f_n tends to f almost surely, follows under A.1 to A.4 and

A.5 $\sum_{n=1}^{\infty} \exp(-\alpha n h_n^q) < \infty$ <u>for</u> <u>all</u> $\alpha > 0$.

Note that A.5 holds if

A.6 $n h_n^q (\log n)^{-1} \to \infty$ <u>as</u> $n \to \infty$.

The above results on the pointwise asymptotic unbiasedness and weak consistency are from Cacoullos (1966), and the strong consistency result is due to Devroye and Wagner (1976); see bibliographical notes in Prakasa Rao (1983, Ch. 3). It is important to note, however, that the assumptions A.1 and A.4 can be relaxed considerably. For example instead of A.1 if we assume the kernel K to be a bounded density with compact support and instead of A.4 we assume that every point is a Lebesgue point for f, then under A.2 and A.3 the weak consistency of f_n follows, see Deheuvels (1974) and Devroye and Györfi (1985, Ch. 6). Devroye and Györfi (1985) also discuss the pointwise consistency (weak and strong) of f_n when the window width h depends on n as well as the data on z.

Uniform weak (strong) consistency, i.e., $\sup_z |f_n(z)-f(z)|$ tends to zero as $n \to \infty$ in probability (almost surely) describes the behaviour of f_n for the entire z space, rather than just at a point z_o in the space. Cacoullos (1966) has shown that the uniform weak consistency of f_n follows under the assumptions A.1, A.2, the characteristic function of K is absolutely integrable, and

A.7 f <u>is uniformly continuous in</u> R^q

A.8 $nh_n^{2q} \to \infty$ <u>as</u> $n \to \infty$.

The weakest possible conditions for uniform strong consistency available to date are by Bertrand-Retali (1978). These are A.2, A.6, A.7 and

A.9 K <u>is almost everywhere continuous bounded kernel with compact support</u>

The other very similar condition was independently obtained by Devroye and Wagner (1980). The results of Van Ryzin (1969) and Silverman (1978) are under restrictive assumptions on the kernel K, see Devroye and Wagner (1980, p. 61). Devroye and Wagner (1980) and Deheuvels and Hominal (1980) have also obtained uniform consistency results for many practical situations where h is a function of n as well as the data on z. For further details, see Devroye and Györfi (1985, Ch. 6).

The conditions for the various consistencies described above are merely the sufficient conditions. The conditions for consistency under the global L_1 criterion, that is conditions under which $\int |f_n - f| dz \to 0$ in probability or almost surely as $n \to \infty$, are discussed in the work of Devroye and Wagner (1979) and Devroye (1983) among others. For an excellent treatment of the L_1 criterion, see Devroye and Györfi (1985). The main results of Devroye (1983) are that, under the L_1 criterion, various types of consistency are equivalent and that A.2 and A.3 are the necessary and sufficient conditions for the consistency. In general, the L_1 criterion is well defined and provides somewhat weaker conditions of consistency. Recently Bai and Chen (1987) have given the necessary and sufficient conditions for consistency under the L_p

criterion, that is conditions under which $[\int |f_n - f|^p dz]^{1/p} \to 0$ as $n \to \infty$. Note that, except for p=1, the L_p criterion is not scale invariant. The L_2 criterion is easier to work with and it has been considered extensively in the literature. We have discussed above the main results under this criterion.

The asymptotic normality of f_n has been proved by Parzen (1962), for q=1, and Cacoullos (1966) more generally. This follows from their results that

$$(4.1) \qquad \frac{f_n - Ef_n}{\sqrt{V(f_n)}} \sim N(0,1) \text{ or } (nh^q)^{1/2}(f_n - Ef_n) \sim N(0, f\int K^2)$$

as $n \to \infty$, where $V(f_n)$ is the asymptotic variance of f_n,

$$(4.2) \qquad V(f_n) \simeq (nh^q)^{-1} f\int K^2, \qquad \int K^2 = \int K^2(z)dz.$$

However, this result is not useful for constructing confidence intervals for the unknown f. For this we must replace Ef_n by f in such a way that

$$(4.3) \qquad (nh^q)^{1/2}(f_n - f) = (nh^q)^{1/2}(f_n - Ef_n) + (nh^q)^{1/2}\text{Bias}(f_n) \sim N(0, f\int k^2) \text{ as } n \to \infty.$$

This can be achieved by choosing h so that the term $(nh^q)^{1/2}$ Bias (f_n) tends to zero as $n \to \infty$to zero as $n \to \infty$. For example, we show below that for symmetric kernels $\text{Bias}(f_n(z)) \propto h^2$. Thus if we choose h such that $n\, h^{4+q}$ tends to zero as $n \to \infty$, the result in (4.3) would hold. For q=1 see e.g. Revesz (1984, p. 542). An alternative is to consider an almost-unbiased-jackknife estimator in place of f_n in (4.3). The idea is to eliminate bias by taking a linear combination of two estimators with different window widths, see e.g., Schucany and Sommers (1977).

4.2 Finite Sample Properties, Determination of h and K, and Speed of Convergence

The selection of h and K is the first step in the implimentation of the results in Section 3 to applied work. These may be determined by considering approximations to the bias, the mean sqaured error (MSE), and the integrated MSE (IMSE) of the f_n, where

(4.4) $\text{IMSE} = \int \text{MSE} = \int E(f_n - f)^2 dz = E\int (f_n - f)^2 dz$

$= I(\text{Bias})^2 + I(\text{Variance}).$

Under the i.i.d. assumption the exact mean of f_n in (3.8) is

$$Ef_n(z) = (nh^q)^{-1} E\sum_{i=1}^{n} K\left(\frac{z_i - z}{h}\right) = (h^q)^{-1} EK\left(\frac{z_1 - z}{h}\right)$$

$$= (h^q)^{-1} \int K\left(\frac{z_1 - z}{h}\right) f(z_1) dz_1 .$$

Similarly, the variance of f_n is

$$V(f_n(z)) = (nh^{2q})^{-1} V\left(K\left(\frac{z_1 - z}{h}\right)\right).$$

The expressions are, however, of no use and instead one uses approximations to the bias and the MSE which require the following assumptions:

A.10 <u>The second order derivatives of</u> f <u>are continuous and bounded in some neighbourhood of</u> z.

A.11 <u>Let $\mathcal{K}_2$ be the class of all non-negative real valued Borel-measurable bounded functions,</u> K, <u>symmetric about the origin such that</u>

$\int z_j^2 K(z) dz = \mu_2$

<u>along with</u> A.1(i) <u>to</u> A.1(iii).

Under assumptions A.10 and A.11, the approximate bias, up to the order of magnitude h^2, is

(4.5) $\text{Bias}(f_n(z)) = Ef_n - f = h^2 \lambda_1(z); \qquad \lambda_1(z) = 2^{-1}\mu_2 Df(z)$

where $Df(z) = \partial^2 f(z)/\partial z \partial z'$. This follows by first writing from above

$$Ef_n(z) = (h^q)^{-1} \int K\left(\frac{z_1 - z}{h}\right) f(z_1) dz_1 = \int K(w_1) f(hw_1 + z) dw_1 ,$$

$w_1 = (z_1 - z)/h$, and then using Taylor's series expansion of $f(hw_1 + z)$ around z. Similarly the approximate variance is

$$V(f_n(z)) = (nh^q)^{-1}\lambda_2(z),\ \lambda_2(z) = f(z)\int K^2.$$

Combining these the approximate MSE is[6]

(4.6) $MSE(f_n(z)) = h^4\lambda_1^2(z) + (nh^q)^{-1}\lambda_2(z)$

and the IMSE is

(4.7) $IMSE\ f_n(z) = h^4 \int\lambda_1^2(z)dz + (nh^q)^{-1} \int K^2.$

The idea behind the small h expansions of bias and MSE is similar to that of Kadane's (1971) small disturbance (or small σ) expansions of estimators in parametric econometrics. In fact, as indicated below, one of the ways to get small h is to have small standard deviation of z.

An intuitive choice of h can now be developed. Observe that while variance depends on h, n and K the bias depends on K, and n only through h. Another point to observe is that if, in order to eliminate the bias, a very small h is used then the variance (also Ivariance) will become large. On the other hand choosing a large h will reduce the variance at the expense of more bias. A way out of this problem is to relate h to n in a way that controls both bias and variance simultaneously. This choice is

(4.8) $h = cn^{-1/(4+q)} \propto n^{-1/(4+q)}$

which makes the order of magnitude of the MSE $n^{-4/(4+q)}$; c is a constant to be determined later. We will see later that this h minimizes MSE: i.e. any other choice of h will lead to trade off between the bias and variance which leads to higher MSE.

Regarding the choice of K satisfying (A.11), we note that any symmetric probability density can be chosen. Usually K will be a symmetric unimodel probability density, for example, the multivariate normal

(4.9) $K(z) = (2\pi)^{-q/2} \exp(-\frac{1}{2} z'z).$

Another possible kernel is the multivariate Epanechnikov (1969) kernel

(4.10) $K(z) = 2^{-1} c_q^{-1} (q+2)(1-z'z)$ if $z'z < 1$

$= 0$ otherwise

where c_q is the volume of the unit q-dimensional sphere. Some other kernels, especially useful for quick calculations, are for i = 2 and 3

(4.11) $K_i(z) = (i+1)\pi^{-1}(1 - z'z)^i$ if $z'z < 1$

$= 0$ otherwise

Notice that these kernels have higher order differentiability compared to one in (4.10).

We note now that the choices of h and K discussed above are similar to the optimal h and K, discussed below, which minimise the approximate IMSE. A simple use of calculus shows that the optimal h which minimizes the MSE given above is

(4.12) $h^* = c\, n^{-1/(q+4)}$, $c = [q\, f(z)(\mu_2 Df(z))^{-2} \int K^2]^{1/(q+4)}$

which is as given in (4.8). The h^* which minimizes IMSE is also the same as (4.12) with f(z) replaced by $\int f(z)dz = 1$ and $(Df(z))^{-2}$ by $(\int (Df(z))^2)^{-1}$. Note that h^* converges to zero as $n \to \infty$, but at a very slow rate of $n^{-1/(q+4)}$ and it is inversely related to the dimension of z.

To see h^* explicitly, consider, for example, the case where the true density f is multivariate normal, with $V(z) = \sigma^2 I$. Then for the case of the normal kernel in (4.9) we get h* which minimizes IMSE as

(4.13) $h^* = \sigma(4(q+2)^{-1})^{1/q+4}\, n^{-1/(q+4)}$.

In practice σ^2 can be replaced by $\sigma^2 = q^{-1} \sum_1^q s_{ii}$; s_{ii} is the sample variance of z_i.

To obtain the optimal kernel we can first substitute h^* in the IMSE and write it as $\{\int K^2(z)dz\}^{4/(4+q)}$. Now we minimize the IMSE, which is the same as the minimization of $\int K^2(z)dz$ subject to $\int K(z)dz = 1$ and $\int z_i^2 K(z)dz = \mu_2$. This gives the optimal kernel given in (4.10), see Deheuvels (1977). For q=1, its graph looks like a parabola, and it was first discovered by Bartlett (1963). Strictly speaking, we should therefore call it Bartlett's and not Epanechnikov's kernel as usually referred to in the literature. Epanechnikov (1969) compared the relative efficiency of various kernels with the optimal kernel and found that any reasonable kernel gives almost optimal results. However, negative-valued kernels may improve the performance of the density estimator; see remarks in 4.4. Also, see Davis (1975, 1977) for the comparison of the MSE of various density estimators.

In general, h^* in (4.12) which minimizes IMSE, depends on the unknown $Df(z)$. However, an operational h^*, say $\hat{h}$, can be constructed by using the estimator f_n for f, and $Df_n(z)$ for $Df(z)$. At the first stage, the estimator f_n can be evaluated by choosing an arbitrary initial value of h. Recently Hall and Marron (1987), and Marron (1987) have pointed out that the rate of convergence of $\hat{h}$ is slow though better than that of the well known least squares cross-validation (cv) procedure (Rudemo (1982) and Bowman (1984)). In particular, for q = 1, $(\hat{h}-h^*)/h^* \sim n^{-4/13}$ whereas $(h_{cv} - h^*)/h^* \sim n^{-1/10}$. Thus for small and moderate size samples, the cross-validation procedure may be subject to more sample noise. This is despite the comforting result that, for large samples, cross-validation is optimal in some sense (Hall (1983) and

Stone (1984)). Marron (1987) and Scott and Terrell (1987) have suggested ways to improve upon the performance of the cross-validation procedure in small samples. There are various other methods of choosing h for the kernel and other density estimators; for details see Prakasa Rao (1983), Silverman (1986), and the review articles by Titterington (1985) and Scott (1986).

4.3 Results for Other Functions

From (3.11) we observe that the regression function estimator $M_n(x)$ depends on the density estimators. Thus, in principle, using the asymptotic results of section 4.1 and with some additional assumptions, the consistency and asymptotic normality results for the $M_n(x)$ should follow. These can be found in Schuster (1972), Prakasa Rao (1983, Ch. 4), Singh et al (1987) and Bierens (1985) among others. Similarly, using the results for the regression function estimator $M_n(x)$ the results for the response or regression coefficient $\beta_n(x)$ and the curvature $C_n(x)$ can be developed, see Vinod and Ullah (1985) and Rilstone and Ullah (1986) for the response coefficient and McMillan et. al. (1986) and Rilstone (1986) for the curvature. Finally, since the residual $e = y-M_n(x) = u-(M_n(x)-M(x)) \simeq u$ for large n, the asymptotic results in 4.1 hold for the density of the residual in (3.18). We present here the results for the response coefficient estimator in (3.14) since these will be of most interest to applied researchers.

The consistency and asymptotic normality result corresponding to (4.3) is

$$(4.14) \quad \beta_n(x) - \beta(x) \sim N(0, \Lambda(x))$$

where

$$(4.15) \quad \Lambda(x) = V(\beta_n(x)) \simeq \frac{V(x)(f(x))^{-1} \int (K')^2}{nh^{p+2}};$$

$V(x) = V(y|x)$ is as given in (2.6) and $K'(w) = \partial K(w)/\partial x_j$ as in (3.13). In practice $\Lambda(x)$ can be estimated by $\Lambda_n(x)$, which is $\Lambda(x)$ with $V(x)$ and $f(x)$ replaced by $V_n(x)$ and $f_n(x)$, respectively, as given in (3.16) and (3.9). The result in (4.14) is useful for constructing confidence interval and testing restrictions on $\beta(x)$ implied by economic theory.

If we consider the x's to be non-stochastic, or the analysis conditional on the x's in Engle et. al. (1983) sense, as usually considered in parametric econometrics, then the nonparametric estimator β_n in (3.13) will be linear in y's. Further β_n (in (3.13)) will be asymptotically normal with the variance

$$(4.16)\quad V(\beta_n(x)) = V(x)\sum_{i=1}^{n}(r_{1i} - r_{2i})^2.$$

For practical implementation of (4.15) or (4.16) we again require h and K. To this end we first note that the approximate MSE of $\beta_n(x)$ (corresponding to the result in (4.6)) is

$$\text{MSE } \beta_n(x) = h^2\, \lambda_3^2(x) + \Lambda(x)$$

where $\lambda_3(x)$ is free from h, and $\Lambda(x)$ is as in (4.15). It is clear that the choice of h for which MSE $\beta_n(x)$ is minimum is

$$h^* = c_1(x) n^{-1/(p+4)} \propto n^{-1/(p+4)}$$

which is similar to h^* in (4.12). However, unlike in (4.12) we do not have an explicit expression for $c_1(x)$. Hence an operational h^* cannot be implemented here. An alternative is to determine c_1 by the cross validation procedure; that is by minimizing

$$(4.17)\quad \sum_{i=1}^{n}(\beta_n(x_i) - \beta_{-i,n}(x_i))^2$$

with respect to c_1 after replacing $h = c_1 n^{-1/(p+4)}$, $\beta_{-i,n}$ represents the response coefficient estimator based on all the observations except x_i. An initial value of h can be taken by considering $c_1 = 1$. A better alternative, in view of (4.13), is to consider $c_1 = \sigma$ such that

(4.18) $$h^* = \sigma n^{-1/(p+4)}.$$

In the situation where the p components of x have different variances a useful value of $h_i (i=1,\ldots,p)$ is

(4.19) $$h_i = \sigma_i n^{-1/(p+4)}$$

where $\sigma_i = V(x_i)$ can be replaced in practice by the sample standard deviation of x_i. An extensive study on the choices of h, for other econometric functions in section 3, would be a useful subject of future research.

4.4. Remarks on Speed of Convergence (Limitations of Nonparametric Approach)

It has been shown in 4.2 that the rate of convergence for the MSE of the density estimator cannot be better than $n^{-4/(q+4)}$, where q is the dimension of the distribution. Similarly, from 4.3., in the case of response coefficient β_n, the rate of convergence is $n^{-2/(p+4)}$ where p is the dimension of x's. This implies that in the case of single x, p=1, the rate cannot be better than 2/5. Further, the higher the dimension the slower the rate, which is the "curse of dimensionality".

The slow rate of convergence implies that the standard errors of the nonparametric estimates may turn out to be large for moderate size samples. Also the tests based on nonparametric residuals may be inefficient, at least in the sense that the rate of convergence in distribution is less than $n^{1/2}$. In this regard the idea of averaging nonparametric estimates in Powell et. al. (1986) and Stock (1985), also see 3.1, is useful. These average estimates converge in distribution at rate $n^{1/2}$.

Another alternative to improve upon the rate of convergence is to consider kernels which take negative as well as positive values and whose $(r-1)^{th}$ order moments are zero (generalization of A.10), see Parzen (1962) and Bartlett (1963). For such kernels, assuming that r^{th} order derivatives of f are continuous around z (generalization of A.11.) Johns and Van Ryzin (1972) for q=1 and later Singh (1981) have shown that the rate of convergence of the MSE can be increased to $n^{-2r/(2r+q)}$. The results of section 4.2 are for r=2. In fact by choosing r large enough one can get the n^{-1} convergence for the MSE and $n^{1/2}$ convergence in distribution. However, for large r it will be difficult to choose kernels satisfying zero moment conditions. For example, if r=11 we would require a kernel whose first ten moments are zero. For details on the choice of kernels whose (r-1) moments are zero, see Muller (1984), Singh (1981), and Ullah and Singh (1985).

4.5 Dependent Observations

Most of the asymptotic results for the i.i.d. case discussed above go through for the case of dependent observations. As expected, one can achieve these asymptotic results by making certain assumptions, along with those in i.i.d. case, on the nature of the dependence. To understand this, let us write from (3.8)

$$(4.20)\quad f_n - Ef_n = (nh^q)^{-1} \sum_{i=1}^{n} \eta_i$$

where $\eta_i = K\left(\frac{z_i - z}{h}\right) - EK\left(\frac{z_i - z}{h}\right)$. Then

$$(4.21)\quad V(f_n) = (nh^q)^{-2} \left[\sum_{i=1}^{n} V(\eta_i) + 2 \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \mathrm{cov}(\eta_j, \eta_{j+i})\right].$$

It is clear that for the i.i.d. observations $\text{cov}(\eta_j, \eta_{j+i}) = 0$ for $i \neq 0$, and the variance of f_n tends to zero as $n \to \infty$ (provided $nh^q \to \infty$). Thus MSE consistency is achieved as mentioned before. One way to achieve this MSE consistency for dependent observations is to impose conditions on the serial dependence ($\text{cov}(\eta_j, \eta_{j+i})$) such that as $n \to \infty$, the second term on the right of (4.21) tends to zero. Such conditions are discussed in the work of Robinson (1983) where it has been shown that for the asymptotic normality of f_n and M_n, respectively, we require

$$(4.22) \sum_{j=n}^{\infty} \alpha_j = o(n^{-1}) \text{ and } \sum_{j=n}^{\infty} \alpha_j^{1-2/\theta} = O(n^{-1}), \ \theta > 2$$

as $n \to \infty$; α_j is called the mixing coefficient of $\{z_t\}$, and is a measure of dependence of the processes $\{z_t\}_{-\infty}^{t}$ and $\{z_t\}_{t+j}^{\infty}$ defined as

$$\alpha_j = \sup_{A,B} |P(A \cap B) - P(A)P(B)|,$$

where $A \in \mathcal{B}_{-\infty}^{t}$ and $B \in \mathcal{B}_{t+j}^{\infty}$; $\mathcal{B}_{\ell}^{\ell'}$ is the σ-field of events generated by $z_\ell, z_{\ell+1}, \ldots, z_{\ell'}$. It is assumed that the process $\{z_t\}$ is strictly stationary, and strongly mixing in the sense that $\alpha_j \to 0$ as $j \to \infty$ (distant future is virtually independent of the past and present, and vice versa). Strongly mixing processes have also been used in density estimation by Ahmad (1979), in regression estimation by Ahmad and Barry (1987), and in other contexts by Ibragimov (1970) and Pham and Tran (1980) among others. Bierens (1983) has considered the estimators f_n and M_n for a more restricted stationary process, namely the ϕ-mixing process. The process $\{z_t\}$ is said to be ϕ-mixing if the mixing coefficient $\phi_j \to o$ as $j \to \infty$, where

$$\phi_j = \sup_{A,B} |P(B|A) - P(B)|$$

and A and B are defined above. Notice that every ϕ-mixing process is a strong-mixing. For further details on the properties of nonparametric estimators under ϕ-mixing, see Bierens (1985) and Abdulal (1984).

Strong consistency results are not yet available for stationary processes nor are results available for non-stationary processes. An extensive study on the asymptotic as well as small sample properties of various econometric functionals in Section 3, for the dependent observations, would be a useful subject of future research.

5. ILLUSTRATIVE EXAMPLES

Below we present two examples; first related to industrial organization and the second to finance. These examples illustrate the estimation and testing of economic parameters of interest in these models without specifying their functional forms.[7]

5.1 Industrial Organization Example

Classical economic theory suggests that the "profit maximization hypothesis" should dictate the compensation of business executives. In the literature on industrial organization, Baumol (1967) and others have proposed the alternative "corporate growth hypothesis" of executive compensation. Empirical evidence by Ciscel and Carroll (1980) and Guerard and Horton (1984), hereafter denoted as GH84, suggests an eclectic view whereby sales, profits and employment all have a significant impact on the executive compensation.

The compensation committee of a large chemical firm was interested in evaluating whether their executives are paid according to the norm in the industry. We use the data based on compustat tapes for 1980 for a sample of 33 peer firms, GH84. If the sales, profits and/or employment changes by one percent, the committee would like to know the appropriate adjustment in the executive compensation in thousands of dollars. The partial derivatives of the following semi-log nonparametric regression model provides the desired answer.

Let y_i be the compensation in thousands of dollars per year for the i-th firm. Similarly, x_{i1} denotes log sales, x_{i2} denotes log profits and x_{i3} denotes log of the number of employees. Now the nonparametric specification of the model is

$$(5.1)\quad y_i = M(x_{i1}, x_{i2}, x_{i3}) + u_i = E(y_i | x_{i1}, x_{i2}, x_{i3}) + u_i$$

which is a special case of (2.2) for p = 3. Note that the estimate of M and its partial derivatives with respect to x_1, x_2 and x_3 can be calculated by using (3.11) and (3.13), respectively. For the calculations the kernel used was the normal kernel given in (4.9), and following the results in section 4 h_j taken was $s_j n^{-1/7}$ where for j=1,2,3, $s_j^2 = \Sigma_1^n (x_{ji} - \bar{x}_j)^2/n$.

A parametric approach to estimating the model by GH84 is to specify:

$$(5.2)\quad E(y_i | x_{i1}, x_{i2}, x_{i3}) = b_o + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} .$$

Table 1 reports the estimates based on ordinary least squares (OLS) and ridge regression from GH84, as well as, our nonparametric estimates of the partial derivatives, evaluated at the mean values of the regressors.

Note that the compensation committee is specifically interested in the estimates of the partial derivatives, not in the regression coefficients of a linear model, per se. The linearity is an artifact of our specification. The fact that the coefficient (=-98.144) of the log Sales variable has a negative sign does not mean that typical chemical industry executives are punished by a \$98,144 reduction in their salary when the Sales of their corporations increase by one percent. The partial derivative of log Sales variable need not be negative if the underlying model is highly nonlinear. Our nonparametric method estimates the partial derivatives directly by actually keeping the variation in other regressors at zero. We find that a one percent increase in Sales, Profits and Employment respectively leads to $60,874, $63,706 and $60,131 increase in executive compensation. Also, these numbers are found to be subject to smaller standard error compared to OLS. Note, however, that the usual OLS standard errors are conditional, i.e., appropriate under the assumption of fixed regressors, whereas our nonparametric standard errors from (4.15) are unconditional. For a more direct comparison with OLS we have also computed nonparametric conditional standard errors using (4.16). These are 21.09, 21.24 and 20.48 for sales, profits and employment respectively, which are again smaller. In GH84 ridge regression is used to alleviate the multicollinearity, and has yielded all positive estimates similar to those of nonparametric model. Since the nonparametric estimation has avoided some of the difficulties associated with multicollinearity as well as linear specifications, applied econometricians may find nonparametric approach attractive in a variety of other problems.

5.2 Finance Example

Estimation of the systematic risk or the beta coefficient has attracted the attention of many researchers in finance. For a given stock, the central parametric model in most of the research has been the single index market model

$$y_i = \alpha + \beta x_i + u_i \tag{5.3}$$

where y_i is the return on security at time i, x_i is the market return at time i, u_i is the disturbance term, and β is the beta coefficient on the systematic risk of the stock. One of the several assumptions that model (5.3) is based on is that β is constant across time. However, many recent studies have examined (5.3) under various specifications of random betas and have found that they are in fact not constant across time, see e.g., Fabozzi and

Francis (1978) and Sunder (1980). Recently Fabozzi et. al. (1984), using ridge regression procedure and the data set in Fabozzi and Francis (1978), claimed that betas are in fact fixed and not random as claimed in earlier studies. All these parametric studies are based on ad hoc specifications of $E(y|x) = M(x)$, as in (5.3), and also of $\beta(x) = \partial M(x)/\partial x$. To avoid these specifications we explore below the question of randomness of beta by the nonparametric approach.

To illustrate the nonparametric method we consider 73 months of return = y (price change plus dividends) and market returns = x data from December 1965 to December 1971 on the two stocks as given in Fabozzi et. al. (1984, p 159). The market returns are based on the Standard and Poor's composite index. The nonparametric estimates of the systematic risk β and its standard error can be calculated by using (3.13) and (4.15) respectively. For the calculations the kernel used was the normal kernel in (4.9) and h taken was $sn^{-1/5}$ as described in section 4; s is the sample standard deviation of x. The estimates and their standard errors in parenthesis for two stocks are

Stock 1: $\beta_n = .5617$ (.2182) Stock 2: $\beta_n = 1.3453$ (.3145)

The OLS parametric estimates and their standard errors are Stock 1: .9445(.1574) and Stock 2: 1.1811(.1685). These parametric and nonparametric estimates indicate that the systematic risk β is significant for both stocks.

Next, for Stock 2, we also obtained nonparametric estimates of $\beta_n(x)$ across the values of x and these are plotted in Figure 1. The figure shows that the systematic risk is a nonlinear function of the market return. This result has not been explored in the parametric literature. We also observe from Figure 2 that β_n across time is random. This indicates that the parametric result of Fabozzi et. al. (1984) that β's are fixed may not be correct. Similar results were obtained for Stock 1.

Table 1: OLS, Ridge and Nonparametric Regression Results, Chemical Industry

Cross Section of 33 Large Firms for 1980

Coeff. of:	Intercept	Sales	Profits	Employment	R^2
OLS by GH84	-360.67	-98.144	109.825	99.617	.56
	(214.88)	(69.08)	(48.19)	(50.98)	
Ridge by GH84	-284.69	15.774	46.710	38.229	
		(11.83)	(14.48)	(12.89)	
Nonparametric		60.874	63.706	60.131	
		(32.76)	(32.90)	(30.69)	

Note: Residual Sum of Squares for OLS is 371432, and for the nonparametric model it is 299179.7. The ridge estimates of standard errors are only suggestive because they ignore the bias of the ridge estimator, and also they are conditional on fixed regressors.

FIGURE 1

BETA COEFFICIENTS

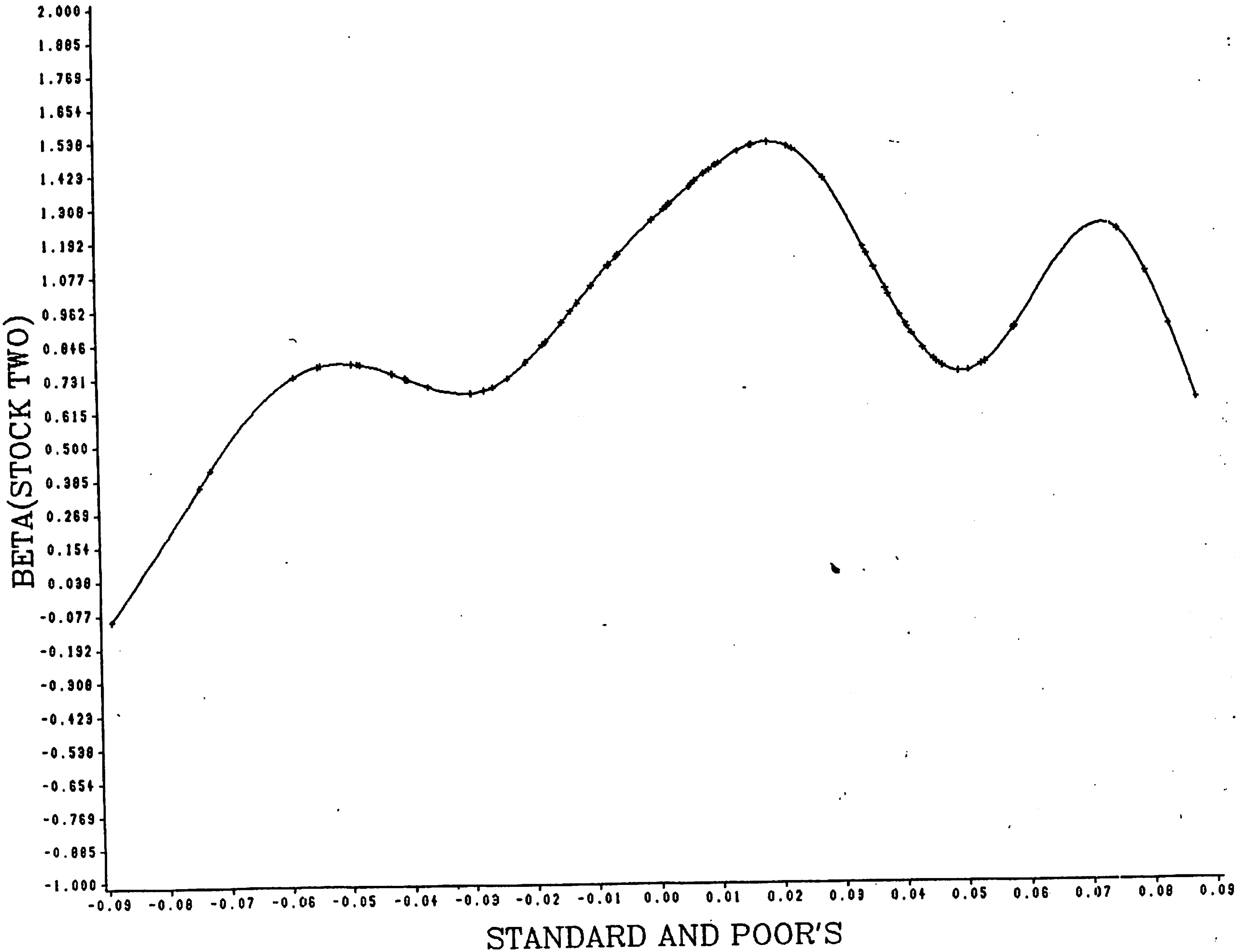
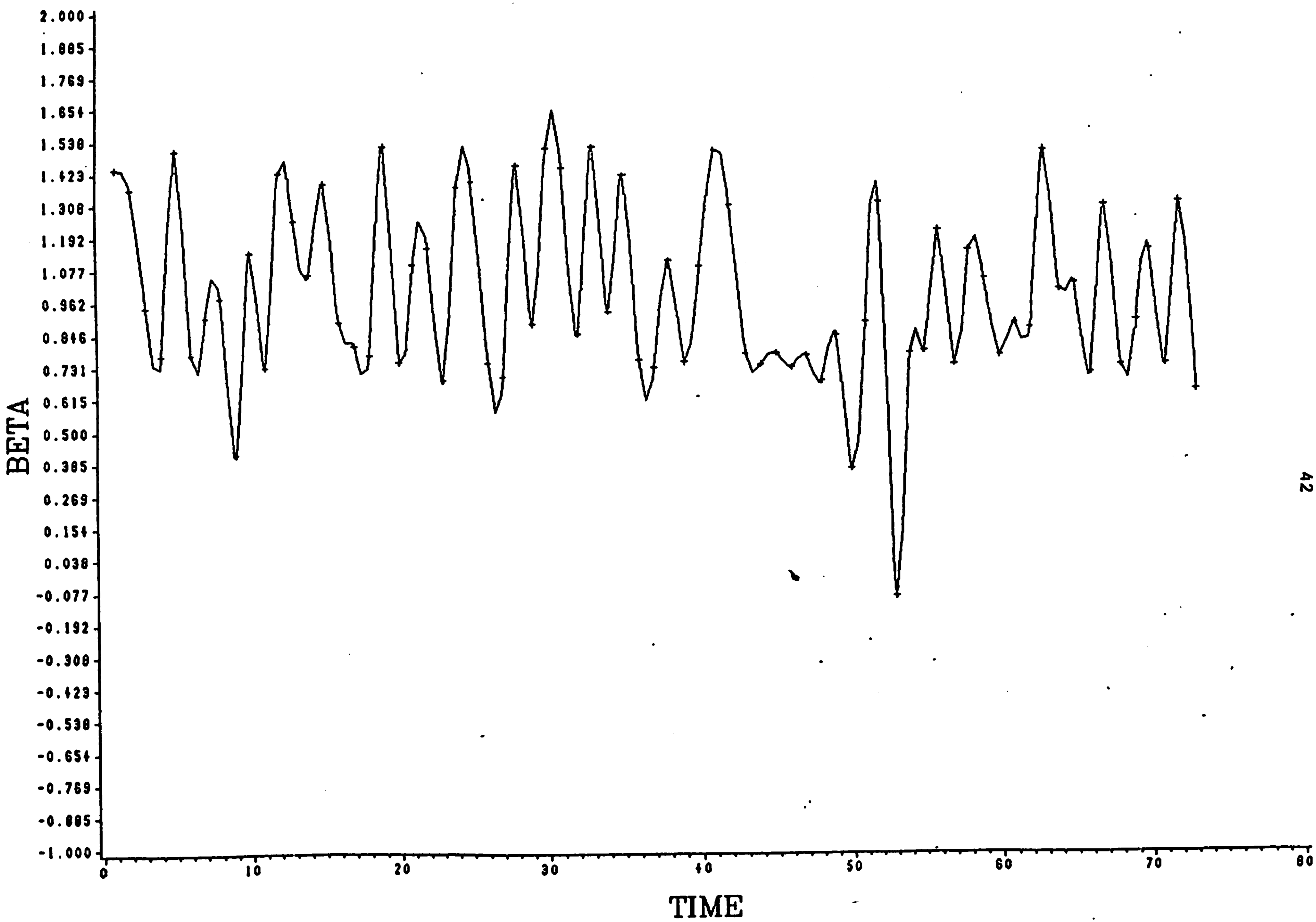


FIGURE 2

BETA COEFFICIENTS



STOCK TWO

FOOTNOTES

*I thank R. Carter, Y. Fan, G. MacDonald, M. Parkin, A. Pagan, P. Rilstone, C. Robinson, D. Scott, H.D. Vinod and the participants of the workshops at the Universities of New South Wales, Canterbury, Guelph, Indiana and Yale for their comments on the subject matter of this article. I also thank two referees for their useful comments. The support of the NSERC is gratefully acknowledged.

[1] The nonparametric methods considered in this paper refer to the nonparametric density estimation methods, not the nonparametric methods in the sense of distribution-free methods (method based on ranks, signs or permutations). For the latter, see for example, Puri and Sen (1985).

[2] This is a classical misspecification problem where the true $E(y|x)$ is, say, $\alpha+\beta x+\gamma x^2+\delta x^3$ but the investigator estimates instead, say, $\alpha+\beta x+\gamma x^2$. We also note here that although it may always be possible (provided $f(y, x)$ is stable across observations and second moments exist) to write $M(x)$ in (2.2) as the linear regression (best linear predictor) of y given x, it may not be the same as $E(y|x)$ which is the best predictor of y given x in the mean square error sense. See Amemiya (1985, p. 3) for further discussion on this point. In this paper we consider the direct estimation of $E(y|x)$ without specifying its form.

[3] In a related work Gasser and Muller (1984) considered the derivatives of a univariate regression where the regressor is a fixed design variable such as time or age.

[4] It should however be noted that the heteroskedasticity has been assumed to be a stable function of x.

[5] The author is grateful to a referee for helpful suggestions and bringing to his notice several important references related to the material in this section.

[6] For a good treatment of the IMSE, see Tapia and Thompson (1978) and Deheuvels (1977). For the comparisons of the rates of MSE and IMSE of various density estimates, see Davis (1975, 1977) and Watson and Leadbetter (1963).

[7] The program used for the calculations in this section is available on request from the author.

REFERENCES

Abdulal, H. I. (1984), On Density Estimation, Ph.D. Thesis, Colorado State University.

Abramson, I. S. (1982), On Bandwidth Variation in Kernel Estimates - A Square Root Law. Annals of Statistics, 10, 1217-1223.

Afriat, S. (1967), The Construction of a Utility Function from Expenditure Data. International Economic Review 8, 67-77.

Ahmad, I. A. (1979), Strong Consistency of Density Estimation by Orthogonal Series Methods for Dependent Variables With Applications. Annals of the Institute of Statistical Mathematics, 31, 279-288.

Ahmad, I. A. and A. M. Barry (1987), Nonparametric Regression Estimation By Orthogonal Series under Generalized Conditions, Research Report, University of South Florida.

Amemiya, T. (1985), Advanced Econometrics (Boston: Harvard University Press).

Bai, Z. D. and X. R. Chen (1987), Necessary and Sufficient Conditions for the Convergence of Integrated and Mean-Integrated P-th Order Error of the Kernel Density Estimates, Centre for Multivariate Analysis, Technical Report No. 87-06, University of Pittsburgh.

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. A 25, 245-254.

Baumol, W. J. (1967), Business Behaviour, Value and Growth (New York: Harcourt, Brace and World).

Bertrand-Retali, M. (1978), Convergence uniforme d'un estimateur de la densité par la méthode du Noyau, Revue Roumaine de Mathématiques Pures et Appliquées 23, 361-385.

Bierens, H.J. (1983), Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions, Journal of the American Statistical Association, 77, 699-707.

Bierens, H. J. (1985) Kernel Estimators of Regression Function, forthcoming in Advances in Econometrics, Cambridge University Press.

Bowman, A. (1984), An Alternative Method of Cross-Validation for the Smoothing of Density Estimates, Biometrika, 65, 521-528.

Breiman, L., W. Meisel and E. Purcell (1977), Variable Kernel Estimates of Multivariate Densities, Technometrics, 19, 135-144.

Cacoullos, T. (1966), Estimation of a Multivariate Density, Annals of the Institute of Mathematical Statistics, 18, 176-189.

Carroll, R. J. (1982), Adapting for Heteroscedasticity in Linear Models, Annals of Statistics 10, pp. 1224-1233.

Carroll, R. J., D. Ruppert and L. A. Stefanski, (1986), Adapting for Heteroscedasticity in Regression Models, Research Report 1702, University of North Carolina.

Cencov, N.N (1962), Evaluation of an Unknown Distribution Density From Observations, Soviet Mathematics, 3, 1559-62.

Christensen, L., D. Jorgenson, and L. Lau (1973), Transendental Logarithmic Production Frontiers Review of Economics and Statistics 55, 28-45.

Ciscel, D. H., and T. M. Carroll (1980) The Determinants of Executive Salaries: An Econometric Survey, The Review of Economics and Statistics, 60, 7-13.

Cosslett, S. (1983), Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model Econometrica 51, 765-82.

Davis, K. B. (1975), Mean Square Error Properties of Density Estimates, Annals of Statistics 3, 1025-30.

Davis, K. B. (1977), Mean Integrated Square Error Properties of Density Estimates, Annals of Statistics 5, 530-535.

Deheuvels, P. (1974), Conditions necessaires et suffisantes de convergence ponctuelle presque sure et uniforme presque sure des estimateurs de la densite, Comptes Rendus Académie des Sciences de Paris Ser A 178, 1217-1220.

Deheuvels, P. (1977), Estimation nonparametrique de la densite par histogrammes generalises, Revue de Statistique Appliquée 25, 5-42.

Deheuvels, P. and P. Hominal (1980), Estimation Automotique De La Densite, Revue de Statistique Appliquée 28, 25-55.

Devroye, L. and T. J. Wagner (1976), Nonparametric Discrimination and Density Estimation, Technical Report 183, University of Texas, Austin.

Devroye, L. and T. J. Wagner (1979), The L_1 Convergence of Kernel Density Estimates, Annals of Statistics, 7, 1136-1139.

Devroye, L. P. and T. J. Wagner (1980), The Strong Uniform Consistency of Kernel Density Estimates. In P. R. Krishnaiah (ed). Multivariate Analysis - V (New York: North Holland).

Devroye, L. (1983), The Equivalence of Weak, Strong and Complete Convergence in L_1 for Kernel Density Estimates, Annals of Statistics, 11, 896-904.

Devroye, L. and L. Györfi, (1985), Nonparametric Density Estimation (New York: John Wiley).

Diewert, W.E. (1971), An Application of the Sheperd Duality Theorem: A Generalized Leontif production function, Journal of Political Economy, 49, 481-507.

Diewert, E. W. and C. Parkan, (1978), Tests for the Consistency of Consumer Data and Nonparametric Index Numbers, Discussion paper 78-27 (University of British Columbia, Vancouver).

Elbadawi, I., A. R. Gallant and G. Souza (1983), An Elasticity Can Be Estimated Consistently Without a Priori Knowledge of Functional Form, Econometrica 51, 1731-1751.

Engle, R. F., D. F. Hendry and J. F. Richard (1983), Exogeneity, Econometrica 51, 277-304.

Engle, R. F., C.W.J. Granger, J. Rice, and A. Weiss (1986), Semiparametric Estimates of the Relation Between Weather and Electricity Sales, Journal of the American Statistical Association, 81, 310-320.

Epanechnikov, V. A. (1969), Nonparametric Estimates of a Multivariate Probability Density, Theory of Probability and Applications, 14, 153-158.

Epstein, L. G., and A. J. Yatchew (1985), Non-parametric Hypothesis Testing Procedures and Applications to Demand Analysis, Journal of Econometrics, 30 (1/2), 149-69.

Fabozzi, F. J., B. Raj and H. D. Vinod, (1984), The Stability of the Systematic Risk of Individual Stocks: Ap Application of Ridge Regression: Communications in Statistics 13(2), 151-161.

Fabozzi, F. J., and J. C. Francis (1978), Beta as a Random Coefficient. Journal of Financial and Quantative Analysis, 13, 101-116.

Fix, E. and J. L. Hodges (1951), Discriminatory Analysis, Nonparametric Estimation: Consistency Properties. Report No. 4, Project No. 21-49-004, USAF School of Aviation Medicine, Raudolph Field, Texas.

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.

Gallant, A. R. (1982), Unbiased Determination of Production Technologies, Journal of Econometrics 20, 285-323.

Gallant, R. A. and G. H. Golub (1984), Imposing Curvature Restrictions on Flexible Functional Forms, Journal of Econometrics, 26, 295-321.

Gallant, A. R. and G. Tauchen (1987), Semi-nonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications, Research Report, North Carolina State University.

Gasser, T. and H. G. Muller (1984), Estimating Regression Functions and Their Derivatives by the Kernel Method, Scandinavian Journal of Statistics, 11, 171-185.

Good, I. J. and R. A. Gaskins (1971), Nonparametric Roughness Penalties for Probability Densities, Biometrika 58, 255-277.

Guerard, J. B. and R. L. Horton, (1984), The Management of Executive Compensation in Large, Dynamic Firms: A Ridge Regression Estimation Communications in Statitstics, Theory & Meth., 13(2), 183-190.

Hall, P. (1983), Large Sample Optimality of Least Squares Cross-Validation In Density Estimation, Annals of Statistics, 11, 1156-1174.

Hall, P. and J. S. Marron (1987), The Amount of Noise Inherent in Bandwidth Selection for a Kernel Density Estimator, Annals of Statistics, 15, 163-181.

Hanoch, G. and M. Rothschild, (1972), Testing the Assumptions of Production Theory: A Nonparametric Approach, Journal of Political Economy 80, 256-275.

Hardle, W. (1984). Robust Regression Function Estimation, Journal of Multivariate Analysis, 14, 169-180.

Hildenbrand, K. and W. Hildenbrand, (1985), On the Mean Income Effect: A Data Analysis of the U.K. Family Expenditure Survey, Research Report, University of Bonn.

Horowitz, J. (1987), Semiparametric M Estimation of Censored Linear Regression Models, Research Report, University of Iowa.

Ibragimov, I. A. (1970), On the Spectrum of Stationary Gaussian Processes Satisfying the Strong Mixing Condition II. Sufficient Conditions. Mixing Rate. Theory of Probability and Applications, 15, 23-36.

Johns, M. F. and J. Van Ryzin (1972), Convergence Rates for Empirical Bayes Two-Action Problems II. Continuous Case, Annals of Mathematical Statistics 43, 934-947.

Kadane, J. B. (1971), Comparison of K-Class Estimators When the Disturbances are Small, Econometrica 39, 723-37.

Kumar, T. K. and J. M. Markmann (1975), Estimation of Probability Density Function: A Monte Carlo Comparison of Parametric and Nonparametric Methods (Preprint).

Loftsgaarden, D. O. and C. P. Quesenberry (1965), A Nonparametric Estimate of a Multivariate Density Function, Annals of Mathematical Statistics 36, 1049-1051.

Mack, Y. P. and M. Rosenblatt (1979), Multivariate K-Nearest Neighbor Density Estimates, Journal of Multivariate Analysis, 9, 1-15.

Manski, C. (1975) Maximum Score Estimtion of the Stochastic Utility Model of Choice Journal of Econometrics 3, 205-228.

Marron, J. S. (1987), Partitioned Cross-Validation, Mimeo, University of North Carolina, Chapel Hill.

Mascaro, A. and A. H. Meltzer (1983), Long-and Short-Term Interest Rates in a Risky World, Journal of Monetary Economics, 25, 221-247.

Marron, J. S. (1987), Partitioned Cross-Validation, Research Report University of North Carolina.

McFadden, D. (1985), Specification of Econometric Models, Presidential Address, Econometric Society, Fifth World Congress, Cambridge, Mass.: Economics Dept., Mass. Institute of Technology.

McMillan, J., A. Ullah, H. D. and Vinod (1986), The Shape of the Demand Curve, University of Western Ontario, manuscript.

Muller, H. G. (1984), Smooth Optimum Kernel Estimators of Densities, Regression Curves and Modes, Annals of Statistics 3, 1329-1348.

Nadaraya, E. (1964), On Regression Estimators, Theory of Probability and Applications 9, 157-159.

Newey, W. K. (1987), Adaptive Estimation of Regression Models Via Moment Restrictions, Econometric Research Program Research Memorandum No. 330, Princeton University.

Pagan, A. (1983) Model Evaluation by Variable Addition. In D. F. Hendry and K. F. Wallis (eds.) Econometrics and Quantitiative Economics, (Oxford: Basil Balckwell).

Pagan, A. and A. Ullah (1985), The Econometric Analysis of Models with Risk Terms, Centre for Economic Policy Research Discussion Paper (London).

Parzen, E. (1962), On the Estimation of Probability Density and Model, Annals of Mathematical Statistics, 33, 1065-1076.

Parzen, E. (1979), "Nonparametric Statistical Data Modelling, Journal of the American Statistical Association, 74, 105-131.

Pham, T. D., and L. T. Tran (1980), The Strong Mixing Property of the Autoregressive Moving Average Time Series Model. Seminaire de Statistique, Grenoble, 59-76.

Powell, J. L. (1986), Symmetrically Trimmed Least Squares Estimation of Tobit Models, Econometrica, 54, 1435-1460.

Powell, J., J. M. Stock and T. M. Stoker, (1986), Semiparametric Estimation of Weighted Average Derivatives, Working paper, 1793-86, M.I.T.

Power, S. and A. Ullah (1986), Nonparametric Monte Carlo Density Estimation of Rational Expectations Estimators and Their t-Ratios, Research Report 13, University of Regina; forthcoming in Advances in Econometrics.

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

Puri, M.L. and P.K. Sen (1985), Nonparametric Methods in General Linear Models, (New York: John Wiley).

Revesz, P. (1984), Density Estimation, in P. R. Krishnaiah and P. K. Sen (eds) Handbook of Statistics, Elsevier Science Publishers (North-Holland).

Rilstone, P. (1985), Nonparametic Partial Derivative Estimation, University of Western Ontario, Manuscript.

Rilstone, P. and A. Ullah (1986), Nonparametic Estimation of Response Coefficients, University of Western Ontario Manuscript.

Robinson, C. and A. Ullah (1987), Nonparametric Estimation of the Simultaneous Equations Model, Mimeo, University of Western Ontario.

Robinson, P. M. (1983), Nonparametric Estimators for Time Series, Journal of Time Series Analysis, 4, 85-208.

Robinson, P. M., (1986a), Asymptotically Efficient Estimtion in the Presence of Heteroskedasticity of Unknown Form. Forthcoming Econometrica.

Robinson, P.M. (1986b), Nonparametic Methods In Specification, The Economic Journal--Conference Papers, 96, 134-141.

Rose, R. L. (1978), Nonparametric Estimation of Weights in Least-Squares Regression Analysis, Doctoral Dissertation, University of California at Davis.

Rosenblatt, M. (1956), Remarks on Some Nonparametric Estimates of Density Function, Annals of Mathematical Statistics. 27, 832-837.

Rudemo, M. (1982), Empirical Choice of Histograms and Kernel Density Estimators, Scandinavian Journal of Statistics, 9, 65-78.

Samuelson, P. A. (1938), A Note on the Pure Theory of Consumer Behaviour, Economica, 5, 61-71.

Schucany, W. R. and J. P. Sommers (1977), Improvement of Kernel Type Density Estimators, Journal of the American Statistical Association, 71, 420-433.

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.

Scott, D. W. (1986), Choosing Smoothing Parameters for Density Estimators. Computer Science and Statistics: in D. M. Allen (ed.) The Interface, Elsevier Science Publishers (North-Holland).

Scott, D. W. and G. R. Terrell (1987), Biased and Unbiased Cross-Validation in Density Estimation, Technical Report 87-02, Rice University.

Silverman, B. W. (1978), Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and Its Derivatives, Annals of Statistics 6, 177-184.

Silverman, B. W. (1986), Density Estimation for Statistics and Data Analysis, (New York: Chapman and Hall).

Singh, R. S. (1977), Applications of Estimators of a Density and its Derivatives to Certain Statistical Problems, Journal of Royal Statistical Society Ser. B 39, 357-363.

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. and D. S. Tracy (1977), Strongly Consistent Estimators of kth Order Regression Curves and Rates of Convergence. Z. Wahrsch Verw. Gebiete. 40, 339-348.

Singh, R. S. and A. Ullah (1986), "Nonparametric Recursive Estimation of a Multivariate, Marginal and Conditional DGP with an Application to Specification of Econometric Models, Communications in Statistics, Theory and Methods, 15, 3489-3513.

Singh, R. S., A. Ullah and R. A. L. Carter (1987), Nonparametric Inference in Econometrics: new applications, in I. MacNeill and G. Umphrey (eds) Time Series and Econometric Modelling (Holland: D. Reidell).

Stock, J. H. (1985), Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Clean Up Benefits, Kennedy School of Government, Harvard University, manuscript.

Stone, C. J. (1984), An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates, Annals of Statistics 12, 1285-1297.

Sunder, S. (1980), Stationarity of Market Risk: Random Coefficient Tests for Individual Common Stocks. Journal of Finance, 35, 883-896.

Tanner, M. A. and W. H. Wong (1984), Data-Based Nonparametric Estimation of the Hazard Function with Applications to Model Diagnostic and Exploratory Analysis, Journal of the American Statistical Association 79, 174-182.

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

Titterington, D. M. (1985), Common Structure of Smoothing Techniques in Statistics, International Statistical Review, 53, 141-170.

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, Technical Report 6, University of Western Ontario.

Van Ryzin, J. (1969), On Strong Consistency of Density Estimates, Annals of Mathematical Statistics, 40, 1765-1772.

Van Ryzin, J (1973), On Histogram Method of Density Estimation, Communications in Statistics, 12, 493-506.

Varian, H. (1984), The Nonparametric Approach to Production Analysis, Econometrica, 52, 579-598.

Varian, H. (1985), Non-Parametric Analysis of Optimizing Behavior With Measurement Error. Journal of Econometrics, 30 (1/2), 445-458.

Vinod, H. D. and A. Ullah (1985), Flexible Production Function Estimation By Nonparametric Kernel Estimators, Research Report, University of Western Ontario.

Watson, G. S. (1964), Smooth Regression Analysis, Sankhya, Series A 26, 51, 175-184.

Watson, G. S. and M. R. Leadbetter (1963), On the Estimation of Probability Density I, Annals of Mathematical Statistics, 34 480-491.

Watson, G. S. and M. R. Leadbetter (1964), Hazard Analysis I. Biometrika,

Wegman, E. J. (1970), Maximum Likelihood Estimation of Unimodal Density Function, Annals of Mathematical Statistics, 41, 457-571.

Wegman, E. J. (1972), Nonparametric Probability Density Estimation II: A Comparison of Density Estimation Methods. Journal of Statistical Computation and Simulation 1, 225-245.

Whittle, P. (1958), On Smoothing of Probability Density Functions, Journal of Royal Statistical Society Ser. B, 20, 334-343.

Wolverton, C. T., and T.J. Wagner (1969), Asymptotically Optimal Discriminant Functions for Pattern Classification. IEEE Transaction of Information Theory, IT-15, 258-265.

Yamato, H. (1971), Sequential Estimation of a Continuous Probability Function and Mode, Bulletin Mathematical Statistics, 14, 1-12.

Yatchew, A. (1987), Some Tests of Nonparametric Regression Models, forthcoming, in W. Barnett, E. Berndt, H. White, (eds), Symposia in Econometrics, Cambrige University Press.