1985

Nonparametric Inference in Econometrics

Radhey S. Singh

Aman Ullah

R.A.L. Carter

ISSN: 0831-4454
ISBN: 0-7714-0703-3

THE CENTRE FOR DECISION SCIENCES AND ECONOMETRICS

Nonparametric Inference in Econometrics

Radhey S. Singh
A. Ullah and R.A.L. Carter

TECHNICAL REPORT NO. 7
NOVEMBER 1985

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

NONPARAMETRIC INFERENCE IN ECONOMETRICS*

by

Radhey S. Singh

Department of Mathematics and Statistics

University of Guelph

Guelph, Ontario

and

A. Ullah and R. A. L. Carter

Department of Economics

University of Western Ontario

London, Ontario

September, 1985

*Invited paper presented at the Symposia on Statistics and a Festschrift in honor of Professor V. M. Joshi's 70th birthday held at the University of Western Ontario, London, May 27-31, 1985. Research supports to R. S. Singh and A. Ullah from NSERC are gratefully acknowledged. The authors are also thankful to V. Ho for research assistance.

ABSTRACT

In this paper, nonparametric estimators of a multivariate density, of conditional mean (regression functions) and of conditional variances (heteroskedasticity) are presented. Among other results, we establish central limit theorems for the estimators and build up confidence intervals based on these estimators. Further, some applications of these estimators are explored in econometrics.

1. Introduction

In econometrics and many other scientific disciplines (such as medical sciences, sociology and psychology) one often has to deal with several variables simultaneously, each dependent on others in some sense. A common inference problem in such sciences, especially in econometrics, is to see how a particular variable on the average is dependent on others, so that prediction (estimation) of the value (or average values) of the variable in question can be made at any specified values of the other variables. A second common inference problem in such sciences, though somewhat related to the first one, is to see how the chosen variable varies (over various spots, items, individuals as the case may be) when other variables are held fixed at certain specified values of interest. The first problem, known as the regression problem and the second problem, known as the heteroskedasticity problem in regression, are invariably handled in various sciences by postulating certain fixed model (functional form) on the regression and by assuming fixed conditional variance (homoskedasticity) of the variable in question. However, it is now well known that the set of all suitable functional forms of the regression (or of the distributions of disturbances) is quite often large, and any postulations regarding the form of the regression and the value of the conditional variance (the variance of the disturbances in the regression) are questionable, and their violations have varying effects on the econometric inferences and policy implications.

The only way of avoiding the misspecification of the functional form of the regression model or of the conditional variance is, in fact, to assume no specific parametric functional form of the regression or of the conditional variance; and to estimate the conditional mean and the conditional variance completely nonparametrically. This in turn can be achieved by estimating nonparametrically the joint probability density function (p.d.f.) of all the variables involved. For example, we can estimate the conditional mean and variance of a variable x_1 given p-1 other related variables $(x_2,\ldots,x_p)$, which in turn can obviously be achieved if we can estimate the joint p.d.f. of $(x_1,\ldots,x_p)$.

Nadaraya (1964), Watson (1964), Rosenblatt (1969), Noda (1976) and Collomb (1979, 1981) are among the first to consider estimation of a regression function nonparametrically using Rosenblatt (1956), Parzen (1962) and Cacoullos (1966)-type kernel estimates of a density function.

In this paper we present nonparametric estimates of a multivariate density, and of the conditional mean and the conditional variance of a variable given the others. The integration of the density estimation with the Monte Carlo technique of doing finite sample econometrics is explored. Also, the nonparametric estimate of the variance is used to analyze the problem of heteroskedasticity in econometrics.

The plan of the paper is as follows. In Section 2 we present the method of constructing nonparametric estimates of a multivariate p.d.f. and the conditional mean and the conditional variance of a variable given the others. In Section 3 we state various results with regard to consistencies, variances and the distributions of these estimators. The confidence intervals for the joint density, the conditional mean and the conditional variance are also presented. In Section 4 we give proofs of the main results stated in Section 3. Finally, in Section 5 we illustrate the performance of our estimators through applications to certain econometric problems.

2. Estimators of the joint p.d.f., the conditional mean and the conditional variance

Suppose we have n independent observations $w_t = (x_{t1},\dots,x_{tp})$, $t=1,\dots,n$ on p random variables $x_1,\dots,x_p$ of our interest. We are to to draw inference about the conditional mean of a variable, say x_1, given the rest $x_2,\dots,x_p$. As mentioned in Section 1, we achieve this by first considering the nonparametric estimation of the joint p.d.f., say f, of $\underset{\sim}{x} = (x_1,\dots,x_p)$, and then considering the estimation of the conditional density, say g, of x_1 given $\underset{\sim}{x}' = (x_2,\dots,x_p)$.

Throughout the remainder of this paper, we denote $\int w^j f(w_1,\underset{\sim}{w}')dw_1$ by $\ell_j(\underset{\sim}{w}')$ where $\underset{\sim}{w}' = (w_2,\dots,w_p)$ is a point in the p-1 dimensional Euclidean space R^{p-1}. We shall, however, use only ℓ_o, ℓ_1, ℓ_2 and ℓ_4. Notice that ℓ_o is the marginal p.d.f. of $\underset{\sim}{x}' = (x_2,\dots,x_p)$, $g(w_1|\underset{\sim}{w}') = f(\underset{\sim}{w})/\ell_o(\underset{\sim}{w}')$ is the conditional p.d.f. of x_1 at w_1 given $\underset{\sim}{x}' = \underset{\sim}{w}'$,

$$M(\underset{\sim}{w}') = \ell_1(\underset{\sim}{w}')/\ell_o(\underset{\sim}{w}')$$

is the conditional mean of x_1 given $\underset{\sim}{x}' = \underset{\sim}{w}'$, and

$$V(\underset{\sim}{w}') = (\ell_2(\underset{\sim}{w}')/\ell_o(\underset{\sim}{w}')) - M^2(\underset{\sim}{w}')$$

is the conditional variance of x given $\underset{\sim}{x}' = \underset{\sim}{w}'$.

As in Singh (1981), for an integer s > 1 and for i=1,...,p, let $\mathcal{K}_i^s$ be the class of all Borel-measurable real valued bounded function, on the real line, symmetric about zero, such that for a $K_i \in \mathcal{K}_i^s$,

$$(2.1)\qquad \int y^j K_i(y)dy = \begin{cases} 1 \text{ if } j=0 \\ 0 \text{ if } j=1,\ldots,s-1, \end{cases}$$

$\int |y^s K_i(y)| dy < \infty$ and $|yK_i(y)| \to 0$ as $|y| \to \infty$. For example, for s=2, take $K_i(y) = 1/2I(-1 < y < 1)$ or $(2\Pi)^{-1/2}\exp(-y^2/2)$ for all i; and for s = 3 or 4 take $K_i(y) = (2\Pi)^{-1/2}[2\exp(-y^2/2) - 2^{-1/2}\exp(-y^2/4)]I(-\infty < y < \infty)$ or $(2\Pi)^{-1/2}(1/2)(3-y^2)\exp(-y^2/2)I(-\infty < y < \infty)$ for all i. Other examples of function K_i may be found in Singh (1979 and 1981). Define K on R^p by

$$K(y_1,\ldots,y_p) = K_1(y_1) \cdot K_2(y_2) \ldots K_p(y_p)$$
$$= \prod_{i=1}^{p} K_i(y_i)$$

<u>Remark 2.1</u> We have chosen the above kernel only for the sake of simplicity. All the results of this paper remain valid if K in various results is replaced by a more general K, namely a Borel-measurable real valu bounded function on R^p symmetric about the origin such that

$\int y_1^{j_1},\ldots,y_p^{j_p} K(y_1,\ldots,y_p) = 1$ or 0 according to $j_1 = \ldots = j_p = 0$ or $0 < j_1 + \ldots + j_p < s-1$;

$\int |y_1^{j_1} \ldots y_p^{j_p} K(y_1,\ldots,y_p)| < \infty$ if $j_1 + \ldots + j_p = s$ and $\|\underset{\sim}{y}\| |K(\underset{\sim}{y})| \to 0$ as $\|\underset{\sim}{y}\| \to \infty$ where $\|\underset{\sim}{y}\|$ is the usual Euclidean norm.

For i-1,...,p, let $0 < h_i = h_i(n)$ be functions of the sample size n such that $h_i \to \infty$ as $n \to \infty$. (A suitable choice of h_i's will be given later.)

We estimate the joint p.d.f. f of $\underset{\sim}{x} = (x_1,\ldots,x_p)$ at $\underset{\sim}{w} = (w_1,\ldots,w_p)$ by

$$(2.2)\qquad \hat{f}(\underset{\sim}{w}) = n^{-1}\sum_{t=1}^{n}\left(\prod_{i=1}^{p} h_i^{-1}\right)K\left(\frac{x_{t1}-w_1}{h_1},\ldots,\frac{x_{tp}-w_p}{h_p}\right)$$

$$= n^{-1}\sum_{t=1}^{n}\left\{\prod_{i=1}^{p}\left(h_i^{-1}K_i\left(\frac{x_{ti}-w_i}{h_i}\right)\right)\right\}$$

Singh (1981) and Singh and Ullah (1984) have considered estimator (2.2) with $h_1 = \ldots = h_p = h$. For other estimators, we define, for j=0,1,

$$\hat{\ell}_j(\underset{\sim}{w}') = |n^{-1}| \sum_{t=1}^{n}(x_{t1}^{j})\left\{\prod_{i=2}^{p}\left(h_i^{-1}K_i\left(\frac{x_{ti}-w_i}{h_i}\right)\right)\right\}$$

$$= n^{-1}\sum_{t=1}^{n} x_{t1}^{j}\left(\prod_{i=2}^{p} h_i^{-1}\right)K'\left(\frac{\underset{\sim}{x}'_t-\underset{\sim}{w}'}{\underset{\sim}{h}'}\right)$$

where $\underset{\sim}{x}'_t = (x_{t2},\ldots,x_{tp})$, $\underset{\sim}{w}' = (\underset{\sim}{w}_2,\ldots,w_p)$, and

$$K'\left(\frac{\underset{\sim}{x}'_t-\underset{\sim}{w}'}{\underset{\sim}{h}'}\right) = \prod_{i=2}^{p}\left(K_i\left(\frac{\underset{\sim}{x}_{ti}-\underset{\sim}{w}_i}{h_i}\right)\right)$$

A nonparametric estimate of the marginal p.d.f. of $\underset{\sim}{x}' = (x_2,\ldots,x_p)$ evaluated at $\underset{\sim}{w}'$ is therefore

$$\int \hat{f}(w_1,\ldots,w_p)dw_1 = \hat{\ell}_o(\underset{\sim}{w}').$$

The estimate of the conditional density of x_1 at w_1 given $\underset{\sim}{x}' = \underset{\sim}{w}'$ is

$$\hat{g}(w_1|\underset{\sim}{w}') = \frac{\hat{f}(\underset{\sim}{w}')}{\hat{\ell}_o(\underset{\sim}{w}')}$$

With the estimate of the conditional p.d.f. of x_1 given $\underset{\sim}{x}'$ in view, a nonparametric estimate of the conditional mean $M(\underset{\sim}{w}')$ of x_1 given $\underset{\sim}{x}' = \underset{\sim}{w}'$ is therefore

$$(2.3) \qquad \hat{M}(\underset{\sim}{w}') = \int w_1 \hat{g}(w_1|\underset{\sim}{w}')dw_1 = \frac{\hat{\ell}_1(\underset{\sim}{w}')}{\hat{\ell}_o(\underset{\sim}{w}')}$$

Finally, our proposed estimator of the conditional variance $V(\underset{\sim}{w}')$ of x_1 given $\underset{\sim}{x}' = \underset{\sim}{w}'$ is

$$(2.4) \qquad \hat{V}(\underset{\sim}{w}') = (\frac{\hat{\ell}_2(\underset{\sim}{w}')}{\hat{\ell}_o(\underset{\sim}{w}')} - \hat{M}^2(\underset{\sim}{w}'))$$

We will claim in the next section that the statistics $\hat{f}$, $\hat{M}$ and $\hat{V}$ are consistent estimators for f, M and V respectively.

3. Consistency and distributional properties of the estimates $\hat{f}$, $\hat{M}$ and $\hat{V}$ along with confidence intervals for f, M and V

Under certain regularity conditions on f we show in this section that the statistics $\hat{f}(\underset{\sim}{w})$, $\hat{M}(\underset{\sim}{w}')$ and $\hat{V}(\underset{\sim}{w}')$ are consistent estimators for $f(w)$, $M(w') = E(x_1|\underset{\sim}{x}' = \underset{\sim}{w}')$ and $V(\underset{\sim}{w}') = var(x_1|\underset{\sim}{x}' = \underset{\sim}{w}')$ respectively. We further obtain the variances and the estimates of the variances of $\hat{f}$ $\hat{M}$ and $\hat{V}$. We also prove the asymptotic normality of these estimators. Finally, using the estimates of the variances of $\hat{f}$, $\hat{M}$ and $\hat{V}$ and their distributional properties we obtain 100(1-α)% confidence intervals for f, M and V. Proof of the results will be presented in Section 4.

3.1 Consistencies and asymptotic distributions of $\hat{f}$, $\hat{M}$ and $\hat{V}$

Theorem 3.1. Let all the s^{th} order partial derivatives of f be continuous at $\underset{\sim}{w}$. Then taking

(3.1) $$h_i \propto n^{-1(2s+p)}$$

we have

(3.2) $$\hat{f}(\underset{\sim}{w}) = f(\underset{\sim}{w}) + O_p(n^{-s/(2s+p)}),$$

and with

(3.3) $$a_n = n(\prod_{i=1}^{p} h_i),$$

(3.4) $$a_n \operatorname{var}(\hat{f}(\underset{\sim}{w})) = \Lambda_o(\underset{\sim}{w}) + o(1)$$

where

(3.5) $$\Lambda_o(\underset{\sim}{w}) = f(\underset{\sim}{w}) \int K^2; \int K^2 = \prod_{i=1}^{p} (\int K_i^2(y)dy),$$

and

(3.6) $$a_n^{½}(\hat{f}(\underset{\sim}{w}) - E\hat{f}(\underset{\sim}{w})) \to N(0, \Lambda_o(\underset{\sim}{w})).$$

Remark 3.1 Since, as we will see in the proof of Theorem 3.1,

(3.7) $$E\hat{f}(\underset{\sim}{w}) = f(\underset{\sim}{w}) + O(\max\{h_1^s, \ldots, h_p^s\}),$$

if we take h_i's so that

(3.8) $$a_n^{½}(\max\{h_1^s, \ldots, h_p^s\}) = o(1),$$

for example take h_i's proportional to $n^{-1/(2s+p-\varepsilon)}$ for any $\varepsilon > 0$, then

$$(3.9)\qquad a_n^{1/2}(\hat{f}(\underset{\sim}{w}) - f(\underset{\sim}{w})) \rightarrow N(0,\Lambda_o(\underset{\sim}{w})).$$

Theorem 3.2 Let the s^{th} order partial derivatives of ℓ_o and ℓ_1 be continuous at $\underset{\sim}{w}'$. Further, let ℓ_2 be continuous at $\underset{\sim}{w}'$. Then taking

$$(3.1)'\qquad h_i \propto n^{-1/(2s+p-1)},$$

we have

$$(3.10)\qquad \hat{M}(\underset{\sim}{w}') = M(\underset{\sim}{w}') + O_p(n^{-s/(2s+p-1)})$$

and with

$$(3.3)'\qquad a'_n = n(\prod_{i=2}^{p} h_i),$$

$$(3.11)\qquad a'_n \operatorname{var}(\hat{M}(\underset{\sim}{w}') = \Lambda_1(\underset{\sim}{w}') + o(1)$$

where

$$(3.12)\qquad \Lambda_1(\underset{\sim}{w}') = \frac{(\ell_2(\underset{\sim}{w}')/\ell_o(\underset{\sim}{w}')) - M^2(\underset{\sim}{w}')}{\ell_o(\underset{\sim}{w}')} \cdot \int(K')^2$$

$$= [\operatorname{var}(x_1|\underset{\sim}{x}' = \underset{\sim}{w}')/\ell_o(\underset{\sim}{w}')]\int(K')^2,$$

with

$$\int(K')^2 = \prod_{i=2}^{p}(\int K_i^2(y)dy);$$

and

$$(3.13)\qquad (a'_n)^{1/2}(\hat{M}(\underset{\sim}{w}') - E(\hat{M}(\underset{\sim}{w}'))) \rightarrow N(0,\Lambda_1(\underset{\sim}{w}')).$$

Remark 3.2   From the proof of Theorem 3.2, it follows that

$$E\hat{M}(w') = M(w') + O(\max\{h_2^s,\ldots,h_p^s\}).$$

Therefore, if h_i's are chosen in such a way so that

$$(a_n')^{1/2}(\max\{h_2^s,\ldots,h_p^s\}) = o(1),$$

e.g. take h_i's proportional to $n^{-1/(2s+p-1-\epsilon)}$ for any $\epsilon > 0$, then

$$(3.14)\qquad a_n'(\hat{M}(\underset{\sim}{w}') - M(\underset{\sim}{w}')) \to N(0,\Lambda_1(\underset{\sim}{w}'))$$

Theorem 3.3   Let the s^{th} order partial derivatives of ℓ_o, ℓ_1 and ℓ_2 be continuous at $\underset{\sim}{w}'$. Further, let ℓ_4 be continuous at $\underset{\sim}{w}'$. Then taking h_i's as in (3.1)', we have

$$(3.15)\qquad \hat{V}(\underset{\sim}{w}') = V(\underset{\sim}{w}) + O_p(n^{-s/(2s+p-1)})$$

$$(3.16)\qquad a_n' \operatorname{var}(\hat{V}(\underset{\sim}{w}')) = \Lambda_2(\underset{\sim}{w}') + o(1)$$

where

$$(3.17)\qquad \Lambda_2(\underset{\sim}{w}') = \frac{(\ell_4(w')/\ell_o(w')) - (\ell_2(w')/\ell_o(w'))^2}{\ell_o(\underset{\sim}{w}')} \int (K')^2$$

$$= [\operatorname{var}(x_1^2|x' = w')/\ell_o(w')] \int (K')^2,$$

and

$$(3.18)\qquad (a_n')^{1/2}(\hat{V}(\underset{\sim}{w}') - E(\hat{V}(\underset{\sim}{w}'))) \to N(0,\Lambda_2(\underset{\sim}{w}'))$$

Remark 3.3   It will be seen in the proof of Theorem 3.3 that $E(\hat{V}(w')) = V(w') + O(\max\{h_2^s,..,h_p^s\})$. Thus if h_i satisfy the hypothesis of Remark 3.2, then

$$(3.19)\qquad (a'_n)^{1/2}(\hat{V}(\underset{\sim}{w}') - V(\underset{\sim}{w}')) \to N(0,\Lambda_2(\underset{\sim}{w}')).$$

The computable confidence intervals of f, M and V, respectively, can easily be written from (3.6), (3.14) and (3.18) after replacing f and ℓ_j, j=0,1,2 and 4 by their consistent estimates $\hat{f}$ and $\hat{\ell}_j$.

4. Proofs of the Theorems in Section 3

In this section we give proofs of our theorems in Section 3.

Proof of Theorem 3.1. Since $(x_{t1},\ldots,x_{tp})$, t=1,...,n are i.i.d. with joint p.d.f. f, taking the expectation of $f(\hat{\underset{\sim}{w}})$ in (2.2) with respect to the joint distribution of $\underset{\sim}{x} = (x_1,\ldots,x_p)$ and using the transformation theorem, we get

$$(4.1)\qquad E\hat{f}(\underset{\sim}{w}) = \int K(\underset{\sim}{y})f(\underset{\sim}{w} + \underset{\sim\sim}{hy})d\underset{\sim}{y}$$

where $(\underset{\sim}{w} + \underset{\sim\sim}{hy}) = (w_1 + h_1y_1,\ldots,w_p + h_py_p)$. Now replacing $f(\underset{\sim}{w} + \underset{\sim\sim}{hy})$ by its Taylor-series expansion at $\underset{\sim}{w}$ with Lagrange's form of the remainder at the s^{th} stage, appling the properties of K_i, and then using the continuity of the s^{th} order partial derivatives of f at $\underset{\sim}{w}$, it follows from the arguments in the proof of Theorem 1 of Singh (1981) that

$$(4.2)\qquad E\hat{f}(\underset{\sim}{w}) = f(\underset{\sim}{w}) + O(\max\{h_1^s,\ldots,h_p^s\}) .$$

Further, since $(x_{t1},\ldots,x_{tp})$ for t=1,2,...,n are i.i.d. with joint p.d.f. f,

$$(4.3)\qquad \operatorname{var}(\hat f(\underset{\sim}{w})) = n^{-1}\operatorname{var}\Big(\prod_{i=1}^{p} h_i^{-1} K\Big(\frac{x_{1i} - w_i}{h_i}\Big)\Big)$$

$$= n^{-1}\Big[\Big(\prod_{i=1}^{p} h_i^{-1}\Big)\int K^2(\underset{\sim}{y}) f(\underset{\sim}{w} + \underset{\sim}{hy}) - (E\hat f(\underset{\sim}{w}))^2\Big]$$

$$= n^{-1}\Big(\prod_{i=1}^{p} h_i^{-1}\Big)[f(\underset{\sim}{w})\int K^2 + o(1)]$$

where $\int K^2 = \prod_{i=1}^{p}(\int K_i^2(y_i)dy_i)$, and the last equation follows by arguments used to prove Theorem 2.2 of Singh and Ullah (1984). Now (3.4) follows from (4.3).

Now (4.2) and (4.3) followed by (3.1) prove that

$$(4.4)\qquad \operatorname{MSE}(\hat f(\underset{\sim}{w})) = O(n^{-2s/(2s+p)})$$

which, with an application of Tchebysheff's inequality prove (3.2).

To prove (3.6), let

$$L_{nt} = n^{-1}\Big(\prod_{j=1}^{p} h_j^{-1}\Big)\Big[K\Big(\frac{\underset{\sim}{x}_t - \underset{\sim}{w}}{\underset{\sim}{h}}\Big) - EK\Big(\frac{\underset{\sim}{x}_t - \underset{\sim}{w}}{\underset{\sim}{h}}\Big)\Big]/(\operatorname{var}\hat f(\underset{\sim}{w}))^{1/2}$$

where

$$K\Big(\frac{\underset{\sim}{x}_t - \underset{\sim}{w}}{\underset{\sim}{h}}\Big) = K\Big(\frac{x_{t1} - w_1}{h_1},\ldots,\frac{x_{tp} - w_p}{h_p}\Big) .$$

Then $L_{n1},\ldots,L_{nn}$ are i.i.d. centered random variables with $S_n = \Sigma_{t=1}^{n} L_{nt} = (\hat f(\underset{\sim}{w}) - E\hat f(\underset{\sim}{w}))/(\operatorname{var}(\hat f(\underset{\sim}{w})))^{1/2}$ and $\operatorname{var}(S_n) = 1$. Temporarily, let $\Phi(\cdot)$

denote the distribution function of the standard normal r.v. Then by the Berry-Esseen Theorem (see Theorem 7.41 of Chung (1968))

$$\sup_{\xi \in R} |P[S_n \le \xi) - \Phi(\xi)| \le c \sum_{t=1}^{n} E|L_{nt}|^3 \tag{4.5}$$

where c is an absolute constant. But by c_r-inequality (Loeve (1963), p. 155),

$$E|L_{nt}|^3 \le 4(\operatorname{var} \hat{f}(\underset{\sim}{w}))^{-3/2} \left(n^{-1} \prod_{j=1}^{p} h_j^{-1}\right)^3 E|K(\frac{\underset{\sim}{x}_t - \underset{\sim}{w}}{\underset{\sim}{h}})|^3 .$$ Since

$$\left(\prod_{j=1}^{p} h_j^{-1}\right) E|K(\frac{\underset{\sim}{x}_t - \underset{\sim}{w}}{\underset{\sim}{h}})|^3 = \int |K|^3 f(\underset{\sim}{w} + \underset{\sim}{hy}) = f(\underset{\sim}{w}) \int |K|^3 + o(1),$$ and

$a_n \operatorname{var}(\hat{f}(\underset{\sim}{w})) = \Lambda_o(\underset{\sim}{w}) + o(1)$, we see that

$$\Sigma_{t=1}^{n} E|L_{nt}|^3 = O(a_n^{-1/2})$$

Thus, we conclude that

$$\sup_{\xi \in R} |P[\frac{\hat{f}(\underset{\sim}{w}) - E\hat{f}(\underset{\sim}{w})}{(\operatorname{var} \hat{f}(\underset{\sim}{w}))^{1/2}} \le t] - \Phi(t)| = O(a_n^{-1/2})$$

and (3.6) follows from (3.4). □

<u>Proof of Theorem 3.2.</u> Throughout this proof $\hat{\ell}_o$, $\hat{\ell}_1$, $\hat{\ell}_2$, ℓ_o, ℓ_1 and ℓ_2 are evaluated at $\underset{\sim}{w}' = (w_2, \ldots, w_p) \in R^{p-1}$ and therefore the argument $\underset{\sim}{w}'$ in these functions will not be displayed. From the proof of Theorem 3.1, and the hypothesis on ℓ_1, it follows that

(4.6) $E\hat{\ell}_o = \ell_o + O(\max\{h_2^s,\ldots,h_p^s\})$, and

$\mathrm{var}(\hat{\ell}_o) = (a_n')^{-1}[\ell_o\int(K')^2 + o(1)]$

Thus with the choice of h_i's in (3.1)', $MSE(\hat{\ell}_o) = O(n^{-2s/(2s+p-1)})$, and hence

(4.7) $\hat{\ell}_o = \ell_o + O_p(n^{-2/(2s+p-1)})$.

Similarly, in view of the hypothesis on ℓ_1 and ℓ_2, it follows that

(4.8) $E\hat{\ell}_1 = \ell_1 + O(\max\{h_2^s,\ldots,h_p^s\})$, and

$\mathrm{var}(\hat{\ell}_1) = (a_n')^{-1}[\ell_2\int(K')^2 + o(1)]$

and, hence with (3.1)'

(4.9) $\hat{\ell}_1 = \ell_1 + O_p(n^{-s/(2s+p-1)})$.

Now we evaluate the $\mathrm{cov}(\hat{\ell}_o,\hat{\ell}_1)$. Recall that $\underset{\sim}{x}_t' = (x_{t2},\ldots,x_{tp})$ and $K'(y_2,\ldots,y_p) = \prod_{i=2}^{p} K_i(y_i)$. Since summonds in $\hat{\ell}_o$ are i.i.d., and so are the summonds in $\hat{\ell}_1$,

(4.10) $a_n' \mathrm{cov}(\hat{\ell}_o,\hat{\ell}_1) = (\prod_{i=2}^{p} h_i^{-1})\mathrm{cov}(x_{11} K'(\frac{\underset{\sim}{x}_1' - \underset{\sim}{w}'}{\underset{\sim}{h}'}), K'(\frac{\underset{\sim}{x}_1' - \underset{\sim}{w}'}{\underset{\sim}{h}'}))$

where $K'(\frac{\underset{\sim}{x}_1' - \underset{\sim}{w}'}{\underset{\sim}{h}'}) = \prod_{i=2}^{p} (K_i(\frac{x_{1i} - w_i}{h_i}))$. But the r.h.s. of (4.10) is

$$\int y_1 (K'(\underset{\sim}{y}))^2 f(y_1, w_2 + h_2 y_2, \ldots, w_p + h_p y_p) d\underset{\sim}{y} - (\prod_{i=2}^{p} h_i)(E\hat{\ell}_o)(E\hat{\ell}_1)$$

$$= \ell_1 \int (K')^2 + o(\prod_{i=2}^{p} h_i)$$

by arguments used to prove the first part of Theorem 3.1. Thus, from (4.11)

$$(4.11) \qquad a'_n \operatorname{cov}(\hat{\ell}_o, \hat{\ell}_1) = \ell_1 \int (K')^2 + o(1)$$

Now, writing

$$(4.12) \qquad \hat{M} = \hat{\ell}_1 / \hat{\ell}_o$$

$$= \frac{E\hat{\ell}_1}{E\hat{\ell}_o} [1 + \frac{\hat{\ell}_1 - E\hat{\ell}_1}{E\hat{\ell}_1} - \frac{\hat{\ell}_o - E\hat{\ell}_o}{E\hat{\ell}_o} + O_p(a'_n)^{-1}]$$

and applying (4.6)-(4.9), we get (3.10).

Now (4.12) followed by (4.6)-(4.9) and (4.11) gives

$$a'_n \operatorname{var}(\hat{M}) = \frac{\ell_1^2}{\ell_o^2} [\frac{\ell_2 \int (K')^2}{\ell_1^2} - \frac{\int (K')^2}{\ell_o} + o(1)]$$

$$= \Lambda_1 + o(1)$$

from the definition of Λ_1 in (3.12). Thus the proof of (3.11) is complete.

Now we prove (3.13). From the arguments used to prove the asymptotic normality of $\hat{f}$, it follows that

$$(4.13)\quad (a_n')^{1/2}(\hat{\ell}_o - E\hat{\ell}_o) \to N(0, \ell_o \int (K')^2)$$

and

$$(4.14)\quad (a_n')^{1/2}(\hat{\ell}_1 - E\hat{\ell}_1) \to N(0, \ell_2 \int (K')^2)$$

These results applied to (4.10) give (3.13). □

Proof of Theorem 3.3. As in the proof of Theorem 3.2, throughout the proof of this theorem too, $\hat{\ell}_j$, $\hat{V}$, $\hat{M}$, ℓ_j, V and M are evaluated at $\underset{\sim}{w}'$ the point displayed in Theorem 3.2.

By the arguments identical to those used in the proof of Theorem 3.2, it follows that

$$E\hat{\ell}_2 = \ell_2 + O(\max\{h_2^s, \ldots, h_p^s\})$$

$$(4.15)\quad \mathrm{var}(\hat{\ell}_2) = (a_n')^{-1}[\ell_4 \int (K')^2 + o(1)], \text{ and}$$

$$\hat{\ell}_2 = \ell_2 + O_p(n^{-s/(2s+p-1)})$$

with h_i's taken as in Theorem 3.1. Further, arguments applied to prove (4.11) can be used to show that

$$(4.16)\quad a_n' \,\mathrm{cov}(\hat{\ell}_o, \hat{\ell}_2)\int = \ell_2 \int (K')^2 + o(1),$$

and writing

$$(4.17)\quad \frac{\hat{\ell}_2}{\hat{\ell}_o} = \frac{E\hat{\ell}_2}{E\hat{\ell}_o}\left[1 + \frac{\hat{\ell}_2 - E\hat{\ell}_2}{E\hat{\ell}_2} - \frac{\hat{\ell}_o - E\hat{\ell}_o}{E\hat{\ell}_o} + O_p(a_n')^{-1}\right],$$

and using (4.6)-(4.9) and (4.15)-(4.16), it follows that

$$E(\hat{\ell}_2/\hat{\ell}_o) = (\ell_2/\ell_o) + O(\max\{h_2^s,\ldots,h_p^s\})$$

$$(4.18)\quad a_n' \operatorname{var}(\hat{\ell}_2/\hat{\ell}_o) = \frac{[\ell_4^2 - (\ell_2^2/\ell_o)]\int(K')^2}{\ell_o^2} + o(1)$$

$$= \Lambda_2 + o(1).$$

Hence by Tchebysheff's inequality

$$(4.19)\quad \frac{\hat{\ell}_2}{\hat{\ell}_o} = \frac{\ell_2}{\ell_o} + O_p(a_n')^{-1/2}$$

Moreover by the arguments used to prove the asymptotic normality of $\hat{f}$, it follows that

$$(4.20)\quad (a_n')^{1/2}(\hat{\ell}_2 - E\hat{\ell}_2) \to N(0,\ell_4\int(K')^2)$$

Thus (4.17) followed by (4.13) and (4.20) proves

$$(4.21)\quad (a_n')^{1/2}((\hat{\ell}_2/\hat{\ell}_o) - E(\hat{\ell}_2/\hat{\ell}_o)) \to N(0,\Lambda_2)$$

Now we obtain the results of Theorem 3.3 with regard to $\hat{V}$. From (4.12), (4.6)-(4.9) and (4.11) it follows that

$$(4.22) \quad E(\hat{M})^2 = \left(\frac{E\hat{\ell}_1}{E\hat{\ell}_o}\right)^2 \left[1 + \frac{var(\hat{\ell}_1)}{(E\hat{\ell}_1)^2} + \frac{var(\hat{\ell}_o)}{(E\hat{\ell}_o)^2} - \frac{2\,cov(\hat{\ell}_1,\hat{\ell}_o)}{(E\hat{\ell}_1)E\hat{\ell}_o} + O(a'_n)^{-1}\right]$$

$$= (\ell_1/\ell_o)^2[1 + O(a'_n)^{-1}]$$

$$= (\ell_1/\ell_o)^2 + O(a'_n)^{-1}$$

Hence from (4.18) it follows that

$$(4.23) \quad E(\hat{V}) = (\ell_2/\ell_o) - (\ell_1/\ell_o)^2 + O(\max\{h_2^s,\ldots,h_p^s\})$$

$$= V + O(\max\{h_2^s,\ldots,h_p^s\})$$

and from (4.19) and (3.10) it follows that

$$(4.24) \quad \hat{V} = V + O_p(a'_n)^{-½}$$

which proves (3.15). Now, from (4.12) and (4.6)-(4.9) it follows that

$$\hat{V} = (\hat{\ell}_2/\hat{\ell}_o) - \hat{M}^2 = (\ell_2/\ell_o) - \hat{M}^2 + O_p(a'_n)^{-1}$$

This result followed by (4.18) proves (3.16), and followed by (4.21) proves (3.18). □

5. Applications

In this section we consider the econometric applications of the estimation of densities and variances.

5.1 Estimating the Density Functions of Exact Sampling Distributions of Econometric Estimators

An important application of the kernel estimator is in estimating the density functions of the exact sampling distributions of econometric estimators and test statistics which are non linear functions of the endogenous data. Such estimated density functions are useful directly e.g. for estimating the true size of asymptotic tests, and indirectly as input to the extended rational approximants (ERA's) of Phillips (1983). The aim of this section is to illustrate the technique with an example which is simple but also has wide applicability and allows the production of new results.

Assume a data generating process (DGP) or joint p.d.f. of the form:

(5.1) $$y_{1t} = \beta x_{1t} + \gamma y_{2t} + u_{1t}.$$

(5.2) $$y_{2t} = \pi_{12} x_{1t} + \pi_{22} x_{2t} + v_{2t};\ t=1,\dots, T$$

Equation (5.1) is a structural equation containing the parameters of interest β and γ while (5.2) is a reduced form equation showing how the endogenous variable y_2 is generated as a linear combination of the two exogenous variables x_1 and x_2 plus an independent normal error v_2 which has zero mean and variance ω_{22}. The error u_1 is also assumed to be independent, normal with mean zero and variance σ_{11}. There is also a reduced form equation for y_1:

(5.3) $$y_{1t} = \pi_{11}x_{1t} + \pi_{21}x_{2t} + v_{1t}$$

where v_1 is independent $N(0, \omega_{11})$.

Because equation (5.1) is just-identified we can write the parameters of interest as functions of the reduced form parameters:

$$(5.4) \qquad \gamma = \frac{\pi_{21}}{\pi_{22}} \quad \text{and} \quad \beta = \pi_{11} - \frac{\pi_{12}\pi_{21}}{\pi_{22}}$$

The normality of the errors together with the exogeneity of the x's implies that least squares (LS) applied to (5.2) and (5.3) will produce maximum-likelihood (ML) estimators of the parameters π_{ij} from which ML estimators of β and γ can be obtained from (5.4). These estimators of β and γ are also indirect-least-squares (ILS) or instrumental-variable (IV) estimators where x_1 and x_2 are the instruments. They are consistent with asymptotically normal sampling distributions. Their exact sampling distributions were given by Basmann et al (1971) who pointed out that they possess no positive integral-order moments.

Two test statistics of natural interest are:

$$(5.5) \qquad t_\beta = \frac{\hat{\beta} - \beta}{as(\hat{\beta})} \quad \text{and} \quad t_\gamma = \frac{\hat{\gamma} - \gamma}{as(\hat{\gamma})}$$

where $as(\hat{\beta})$ and $as(\hat{\gamma})$ are the (estimated) asymptotic standard errors of $\hat{\beta}$ and $\hat{\gamma}$. Asymptotically t_β and t_γ follow the standard normal distribution but their exact distribution appears to be unknown. (Richardson and Rohr (1971) derive the exact distribution for similar test statistics in the over-identified case). Although the random denominators in these ratios are always positive, the fact that the numerators lack moments makes one reluctant to assume that the ratios have moments.

Monte Carlo simulation can be used to produce samples of values of $\hat{\beta}$, $\hat{\gamma}$, t_β and t_γ. However, measures of bias, mean-squared-error and other moments computed from such samples are of little relevance, given the non-existence of the population parameters. However, nonparametric methods are well suited to estimating the density functions of $\hat{\beta}$, $\hat{\gamma}$, t_β and t_γ. Indeed, such density function estimates may well be regarded as more complete and useful than moment estimates even if the population moments did exist.

For the purposes of the Monte Carlo experiment the parameters of the DGP were set to the following values:

$$\beta = .6, \ \gamma = .2$$

$$\pi_{11} = .545455, \ \pi_{21} = .163636, \ \pi_{12} = -.272727, \ \pi_{22} = .818182, \ T = 20$$

Values of x_{1t} and x_{2t} were generated such that:

$$\Sigma_t x_{1t} = \Sigma_t x_{2t} = 0.0, \quad \Sigma_t x_{1t}^2 = \Sigma_t x_{2t}^2 = 20.0$$

and $\Sigma_t x_{1t} x_{2t} = 0.0$. The set of x's was fixed over repeated samples. Two alternative reduced form error covariances were employed. Under the heading "loose fit" $\omega_{11} = .535537$ and $\omega_{22} = 53.5537$. When combined with the x's, these values give standard errors such that: $s(\hat{\pi}_{21}) = \pi_{21}$ and $s(\hat{\pi}_{22}) = 2\pi_{22}$, i.e. the probability of obtaining a negative value of $\hat{\pi}_{22}$ is .309. Under the heading "tight fit" $\omega_{11} = .0360331$ and $\omega_{22} = .0826446$. These values gave population goodness of fit measures of .90 for both reduced form equations.

While this may be a realistic specification it implies that the probability of obtaining a negative $\hat{\pi}_{22}$ is less than 1.7765×10^{-33}. In both cases the covariance between v_{1t} and v_{2t}, ω_{12}, was set to 0.0. Two experiments, each of 100 replications, were run; one loose fit and one tight fit.

The random number generator used was a version of Marsaglia's Super Duper generator as implemented by McLeod (1982).

Each experiment resulted in a frequency distribution for $\hat{\beta}$, $\hat{\gamma}$, t_β and t_γ. In addition frequency distributions were formed for $\hat{\pi}_{21}$ and $t_\pi = (\hat{\pi}_{21} - \pi_{21})/s(\hat{\pi}_{21})$, where $s(\hat{\pi}_{21})$ is the (estimated) standard error of $\hat{\pi}_{21}$. While these frequency distributions convey some information about the underlying sampling distributions, the modest number of replications used means they are lumpy, with several empty classes.

Non-parametric estimates of the densities of $\hat{\beta}$, $\hat{\gamma}$, t_β, and t_γ were obtained from

$$(5.6) \qquad \hat{f}(z) = \frac{1}{100h} \sum_{i=1}^{100} K\left(\frac{z_i - z}{h}\right)$$

where:

(i) z_i is a value of $\hat{\beta}$, $\hat{\gamma}$, t_β or t_γ, obtained by Monte Carlo Simulation, standardized by subtracting its average over the 100 replications and dividing by its standard deviation over the 100 replications. This standardizing transformation alters the location and scale of the density but not its shape.

(ii) z is a value at which the density is to be estimated. Values of z were set between -5 and 5 in increments of .1.

(iii) h is the window width, $100^{-1/5} = .398$.

(iv) The normal kernel is used;

$$K\left[\left(\frac{z_i - z}{h}\right)\right] = (2\Pi)^{-1/2} \exp\left\{-\frac{1}{2}\left[\frac{z_i - z}{h}\right]^2\right\}$$

Of course, the estimates obtained by this procedure embody some sampling error. Therefore, when (5.6) was evaluated with z_i formed from $\hat{\pi}_{21}$, the resulting density function was not exactly standard normal: compare the standard normal density in Figure 1 to the estimated densities for $\hat{\pi}_{21}$ from the tight fit experiment in Figure 2. (The estimated density of $\hat{\pi}_{21}$ from the loose fit experiment and the estimated densities of t_π for both experiments were nearly identical to Figure 2.) Although the density function in Figure 2 has its peak slightly too far right and is slightly skewed left, it is still a very good estimate of the standard normal density function in Figure 1, even though it is based on only 100 points. This gives us confidence that the estimated densities for $\hat{\beta}$, $\hat{\gamma}$, t_β and t_γ will also be close to their population counterparts.

The estimated density for $\hat{\beta}$ from the loose fit experiment is plotted in Figure 3. The analogous plot for $\hat{\gamma}$ is extremely close to that shown in Figure 3. Both densities have very high peaks and long, thin tails. The estimated density of t_β from the loose fit experiment, see Figure 5, looks very similar to Figure 2, but the estimated density of t_γ is strongly skewed to the left, see Figure 6.

The estimated density of $\hat{\beta}$ from the right fit experiment is plotted in Figure 4. (The plot of $\hat{\gamma}$ was very similar to that for $\hat{\beta}$.) It contrasts sharply with the earlier results; the high peaks and long tails are absent. Indeed Figure 4 looks very much like Figure 2. Figures 7 and 8 show the estimated densities for t_β and t_γ when the fit was tight. Now the t_γ distribution closely resembles the t_β distribution, in contrast to the skewed distribution obtained when the fit was loose.

The estimated density functions presented in Figures 3 to 8 suffer from the disadvantage that they are point estimates. One might reasonably ask for measures of their precision or, better still, interval estimates. Asymptotic 95% confident intervals for f(z) are given by

$$\hat{f}(z) \pm 1.96 \left[\frac{\hat{f}(z)}{2nh\sqrt{\pi}}\right]^{1/2} .$$

with n = 100, the number of replications used in the simulation. These confidence intervals are plotted for t_β and t_γ for the loose fit experiment in Figures 9 and 10. Both sets of confidence intervals from the right fit experiment closely resembled Figure 9. The Standard normal density function (Figure 1) lies entirely within those confidence limits for t_β from both experiments and for t_γ from the tight-fit experiment. However, it lies outside these limits for t_γ from the loose-fit experiment.

The nonparametric density estimates presented in this section suggest several conclusions. First, the shape of the exact, small sample distributions of ILS/IV estimators of the structural parameters of just identified models depends crucially upon the probability that reduced form coefficient estimates, which appear in the denominators of ratios entering the expression for structural coefficient estimates, change in sign. This probability will be high if the goodness of fit of the reduced form is low and/or when small samples are employed. When this probability is high the small-sample distributions of the structual coefficient estimators have high peaks and long thin tails, i.e.: they are much different from their large-sample asymptotic distributions. The difference between the small and large sample distribution grows less as the probability of sign change becomes smaller.

The second conclusion, which is of much greater operational significance, is that the small-sample distribution of t ratios depends not only on the probability of reduced form coefficient estimates changing sign, but also upon which structural coefficients enters the t ratio. Those formed from the coefficients of exogenous variables appear to have small-sample distributions which always resemble their large sample distributions. However, t ratios formed from the coefficients of endogenous variables have small-sample distributions resembling the standard normal only if the probability of sign change noted above is small, e.g., if the reduced form fits tightly. In other cases their shape is distinctly non-normal so that the use of the standard normal may yield poor inferences.

5.2 Estimation of unknown variances (heteroskedasticity)

Here we analyze the conditional variance of earning (y) with respect to experience (z). For simplicity in illustration, we have assumed schooling to be constant. Our main interest is to look into the specification of the variability in earnings. For this purpose we considered Canadian data (1971 Canadian Census Public Use Tapes) on 205 individuals' ages (for experience) and their earnings. These individuals were educated to grade 13. The conditional variance, $V(y|z)$, in (2.4) was estimated by using the kernel function:

$$K\left(\frac{z_t - z}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z_t - z}{h}\right)^2}, \quad h = n^{-1/5} s$$

where $s^2 = \Sigma(z_t - \bar{z})^2/n$ is the sample variance of z.

It is clear from the estimate of conditional variances in Figure 11 that the true form of the variability in y with respect to z, $\hat{V}(y|z)$, is a second degree polynomial convex to the z axis. This is consistent with the result of Mincer (1974, p. 101). The important point to note, however, is that the variability of earnings here has been examined without using the grouped data unlike in Mincer (1974).

In view of the above finding we may conclude that the $V(y|z)$ is negatively related with the nonparametric estimate of $E(y|z)$ which is, as indicated in Ullah (1985), a second degree polynomial concave to the z axis. To see that this is actually the case we estimated the regression of y on z, z^2 and $\hat{V}(y|z)$. The result was as follows:

$$\underset{}{y} = \underset{(.987)}{11.649} + \underset{(.039)}{.1152} - \underset{(.005)}{.001z^2} - \underset{(.602)}{1.103\hat{V}(y|z)}$$

Note that the coefficient of $\hat{V}(y|z)$ is negative and significant indicating the negative relationship between $E(y|z)$ and $\hat{V}(y|z)$. The above result provides a possible alternative specification of the earnings equation with variability as an additional variable.

The nonparametric estimates of $V(y|z)$ can also be utilized to perform the generalized least squares (GLS) estimation technique in the earnings equation

$$y = \alpha + \beta z + \gamma z^2 + u = X\delta + u$$

where $X = [1\ z\ z^2]$ and $\delta = [\alpha\ \beta\ \gamma]'$. The GLS estimator is

$$\hat{\delta} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y$$

where $\hat{\Sigma} = \text{Diag.}[\hat{V}(y|z_1),\ldots,\hat{V}(y|z_n)]$.

The least squares (LS) and the GLS estimates obtained are:

$$\text{LS:} \quad y = \underset{(.518)}{10.041} + \underset{(.027)}{.173z} - \underset{(.0003)}{.002z^2}$$

$$\text{GLS:} \quad y = \underset{(.498)}{10.274} + \underset{(.025)}{.165z} - \underset{(.0003)}{.002z^2}$$

where the numbers in parentheses are standard errors. It is clear that the GLS outperforms the LS estimates. The important point to note here is that the GLS estimates have been obtained without using any assumption about the form of heteroskedasticity. To our knowledge, this has not been done in the literature on heteroskedasticity.

References

Basmann, R. L., F. L. Brown, W. S. Dawes and G. L. Schoepfee (1971), "Exact Finite Sample Density Functions of GCL Estimators of Structural Coefficients in a Leading Exactly Identifiable Case," Journal of the American Statistical Association 66, 122-126.

Collomb, G. (1979), "Conditions necessaires et suffisantes de convergence uniforme d'un estimateur de la regression, estimation des derivaees de la regression," C.R. Acad. Sei. Paris 288, A, 161-164.

_______ (1981), "Estimation nonparametrique de la regression," Revue Bibliographique I.S.R., 75-93.

Cacoullos, T. (1966), "Estimation of a Multivariate Density," Annals of the Institute of Statistical Mathematics 18, 179-189.

Loeve, M. (1963), Probability Theory, 3rd ed., Van Nostrand, Princeton, New Jersey.

McLeod, A. I. (1982), "Efficiency Fortran Coding of a Random Number Generator," Department of Statistical and Actuarial Sciences, University of Western Ontario, TR-82-08.

Mincer, J., Schooling, Experience and Earnings, N.B.E.R., New York, Columbia University Press.

Nadaraya, E. A. (1964) "On Estimating Regression," Theory of Probability and Applications 9, 141-142.

Noda, K. (1976), "Estimation of a Regression Function by the Parzen Kernel Type Density Estimators," Ann. Inst. Statist. Math. 128, 221-234.

Parzen, E. (1962), "On the Estimation of Probability Density and Model," Ann. Math. Statist. 33, 1065-1076.

Phillips, P. C. B. (1983), "ERA's: A New Approach to Small Sample Theory," Econometrica 51, 1505-1526.

Richardson, D. H. and R. J. Rohr (1971), "Distribution of a Structural t-Statistic for the Case of Two Included Endogenous Variables," Journal of the American Statistical Association 66, 375-382.

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

_______ (1969), "Conditional Probability Density and Regression Estimators," in P. R. Krishnaiah (ed.), Multivariate Analysis II, 25-31, New York, Academic Press.

Singh, R. S. (1979), "On Necessary and Sufficient Conditions for Uniform Strong Consistency of Estimators of a Density and its Derivatives," J. of Mult. Anal., Vol. 9, 157-164.

_______ (1981), "Speed of Convergence in Nonparametric Estimation of a Multivariate μ-Density and its Mixed Partial Derivatives," Journal of Statistical Planning and Inference 5, 287-298.

Singh, R. S. and A. Ullah, "Estimation of Multivariate DGP," University of Western Ontario Manuscript.

_______ (1984), "Estimation of Density and its Applications in Econometrics," University of Western Ontario Manuscript.

Ullah A., "Specification Analysis of Econometric Models," J. of Quantitative Economics 1, 187-209.

Watson, G. S., "Smooth Regression Analysis," Sankhya, Series A 26, 359-372.

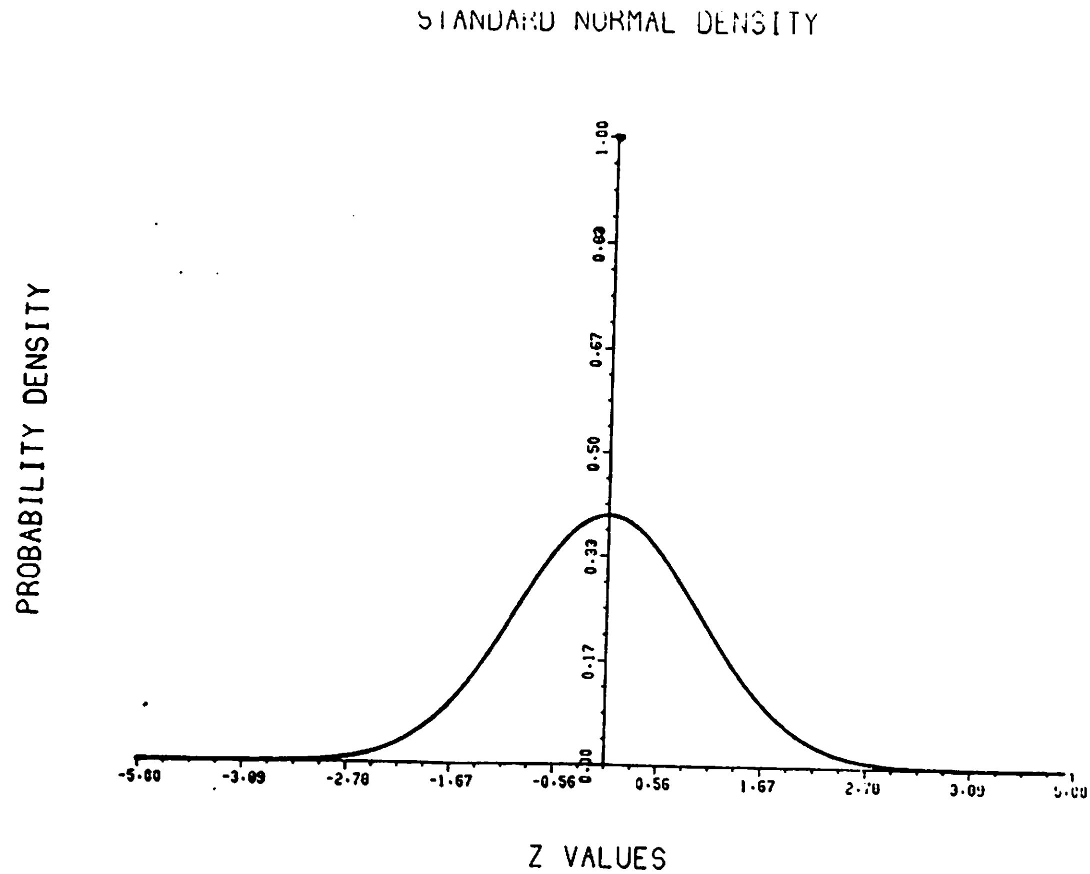

Figure 1

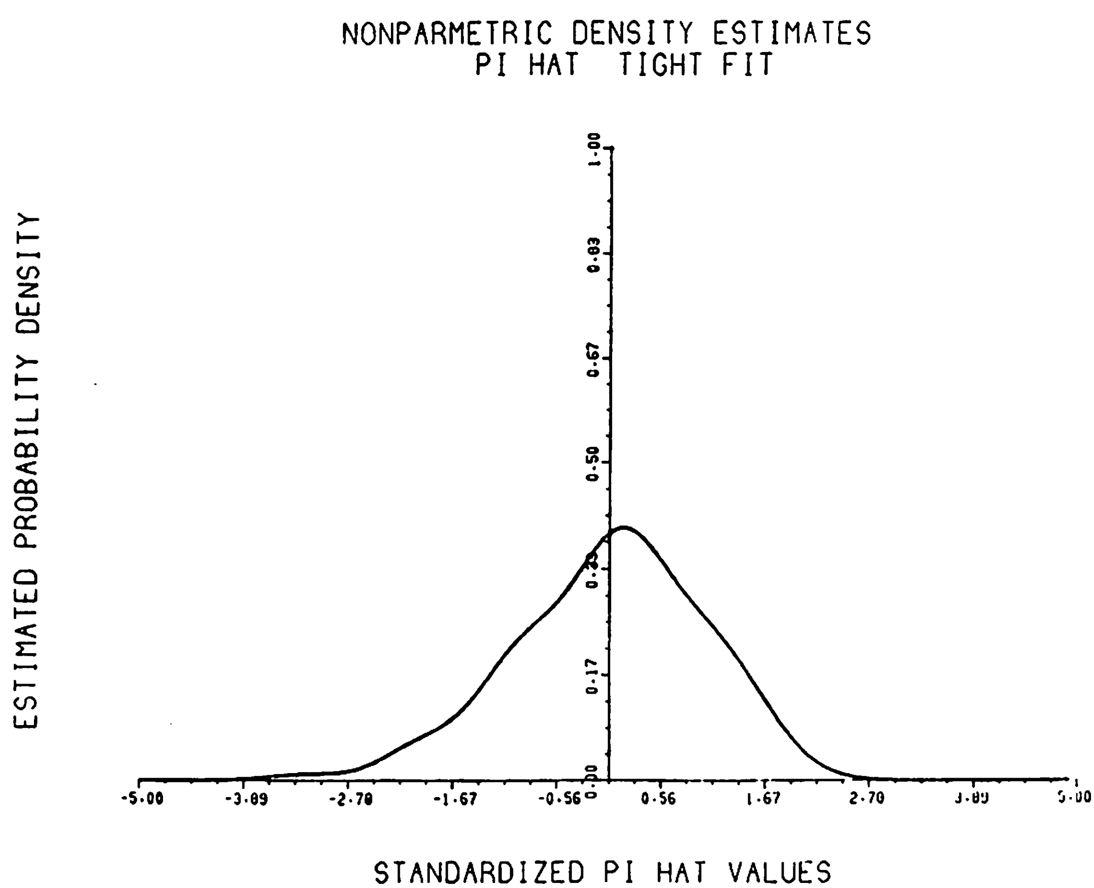

Figure 2

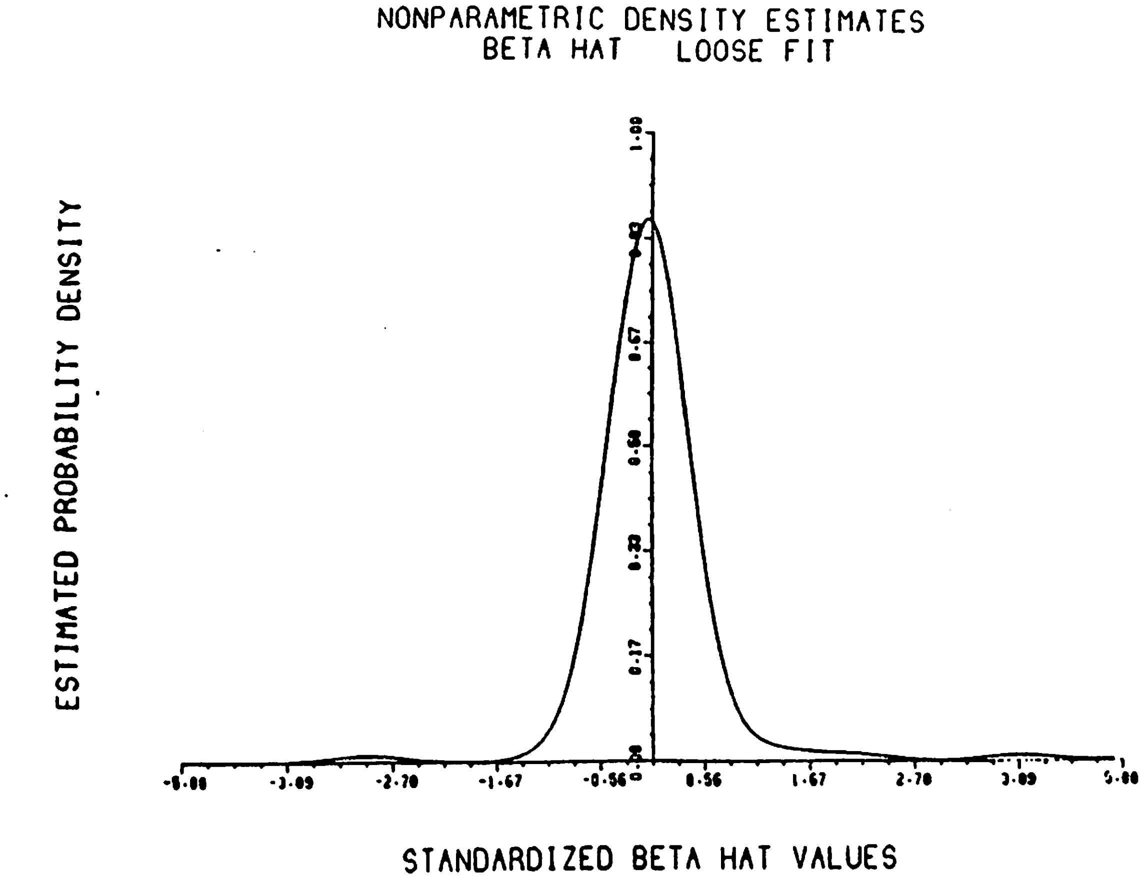

Figure 3

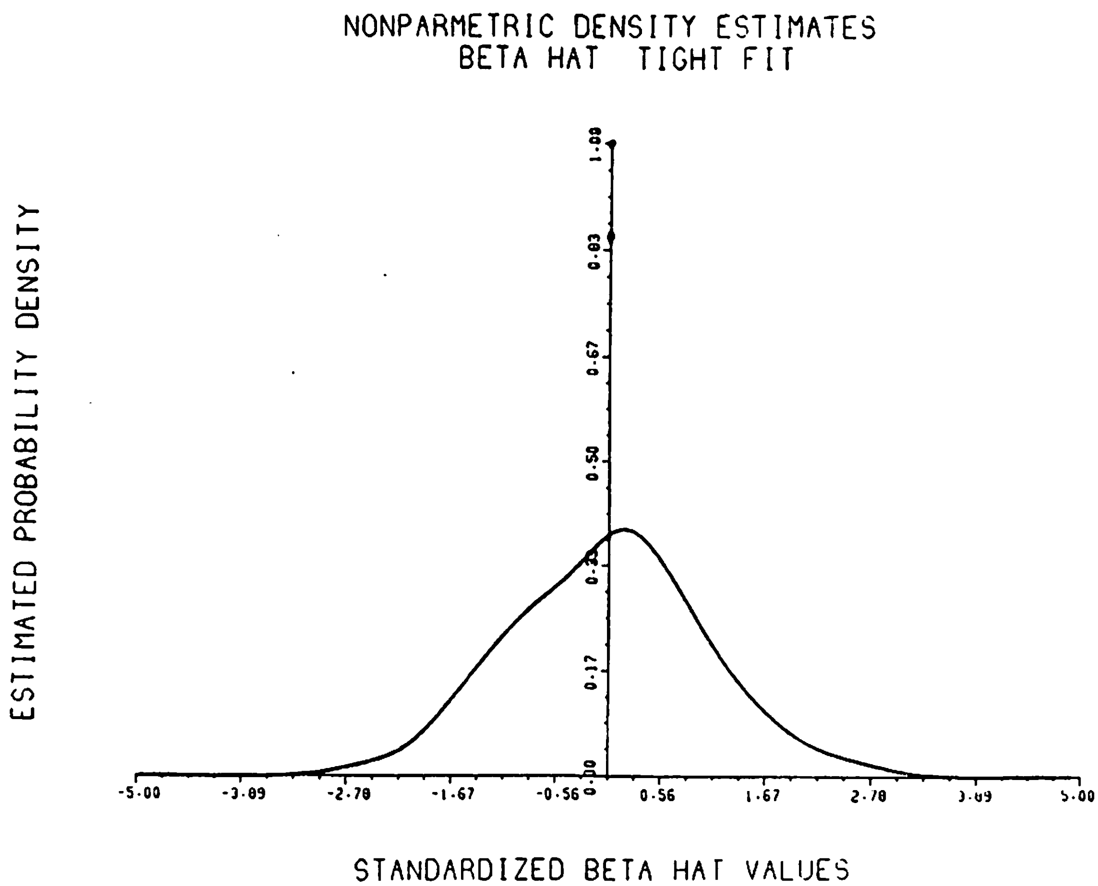

Figure 4

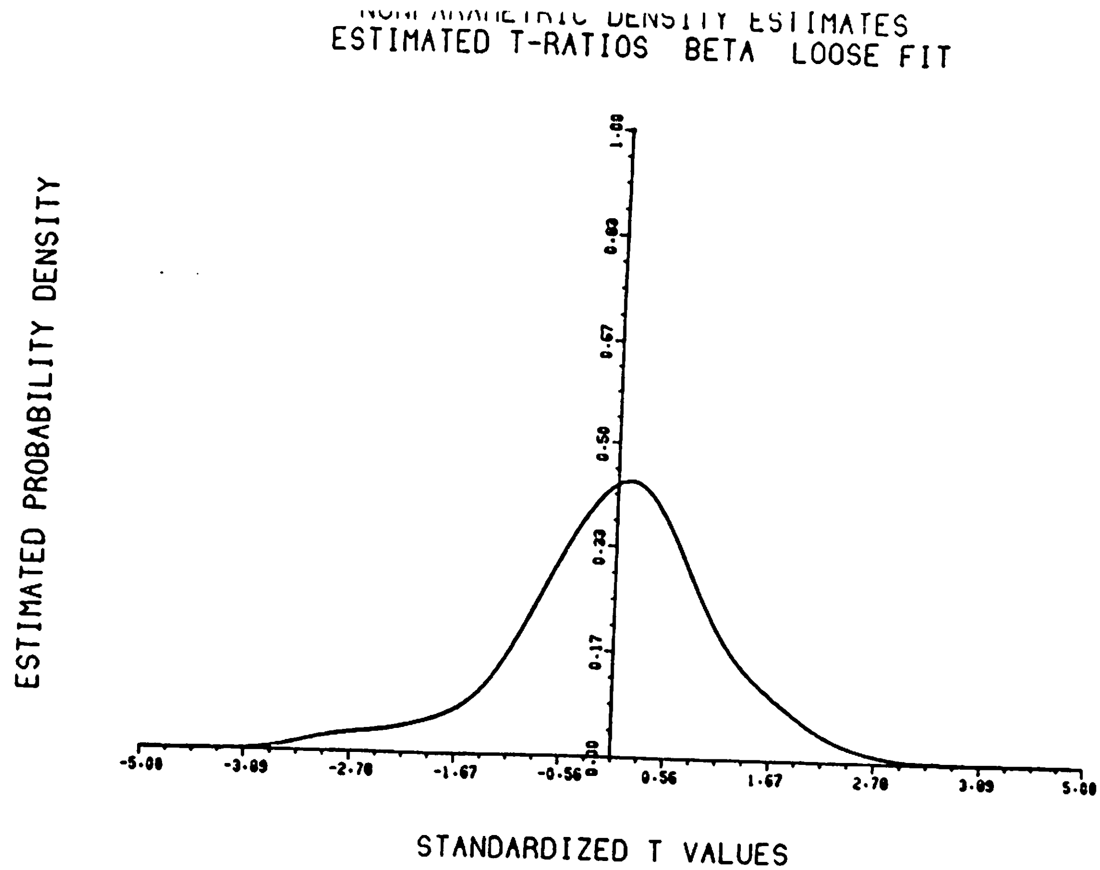

Figure 5

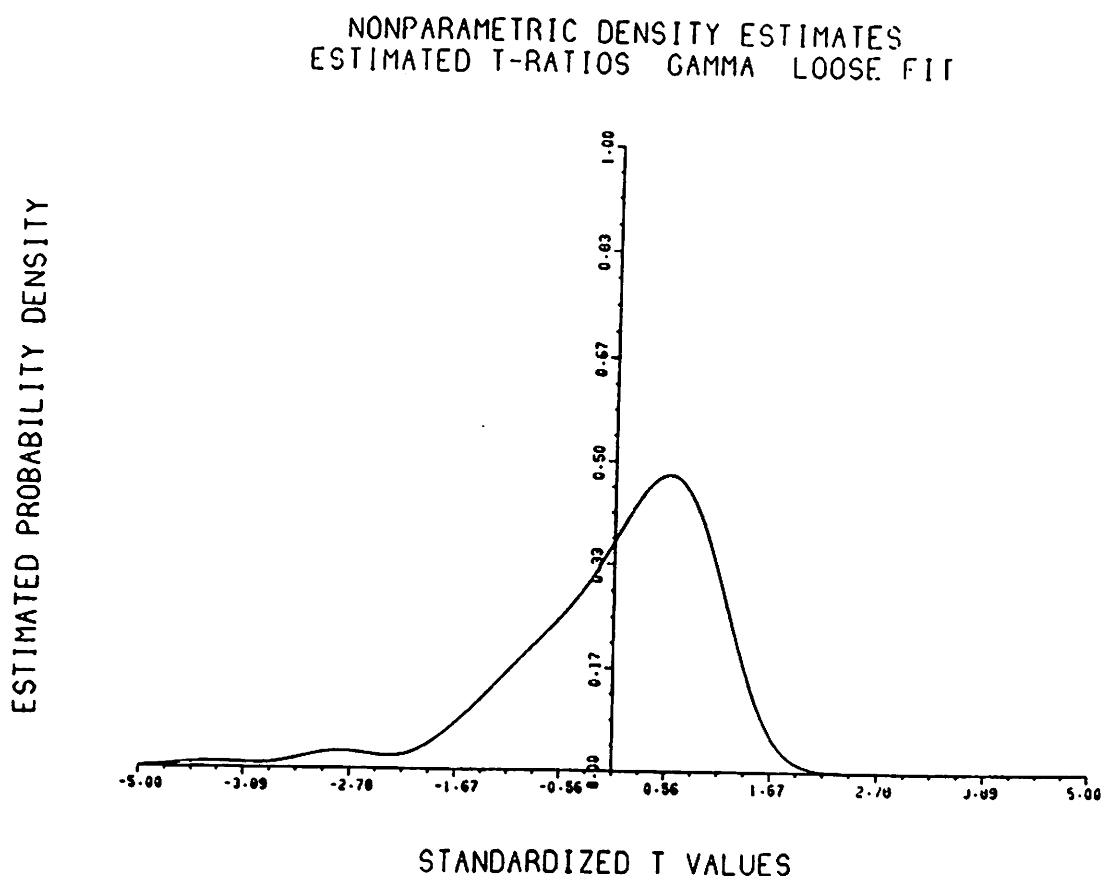

Figure 6

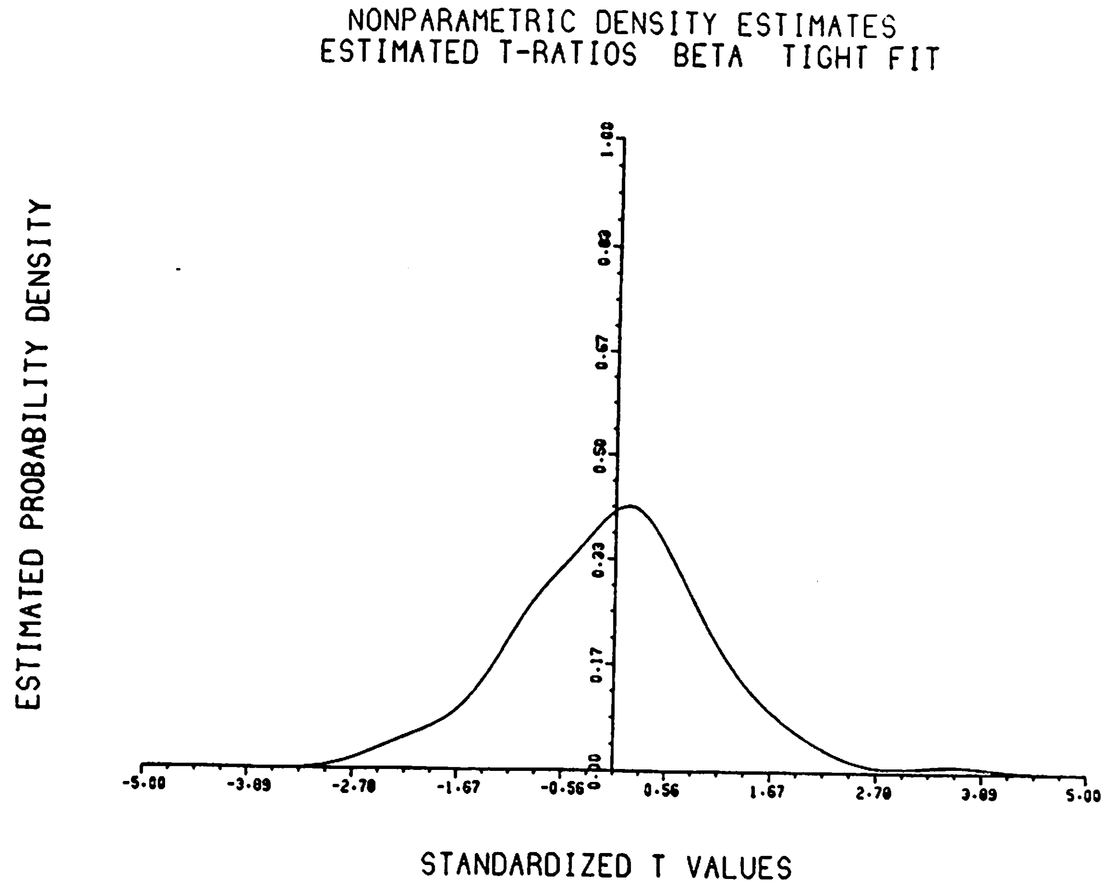

Figure 7

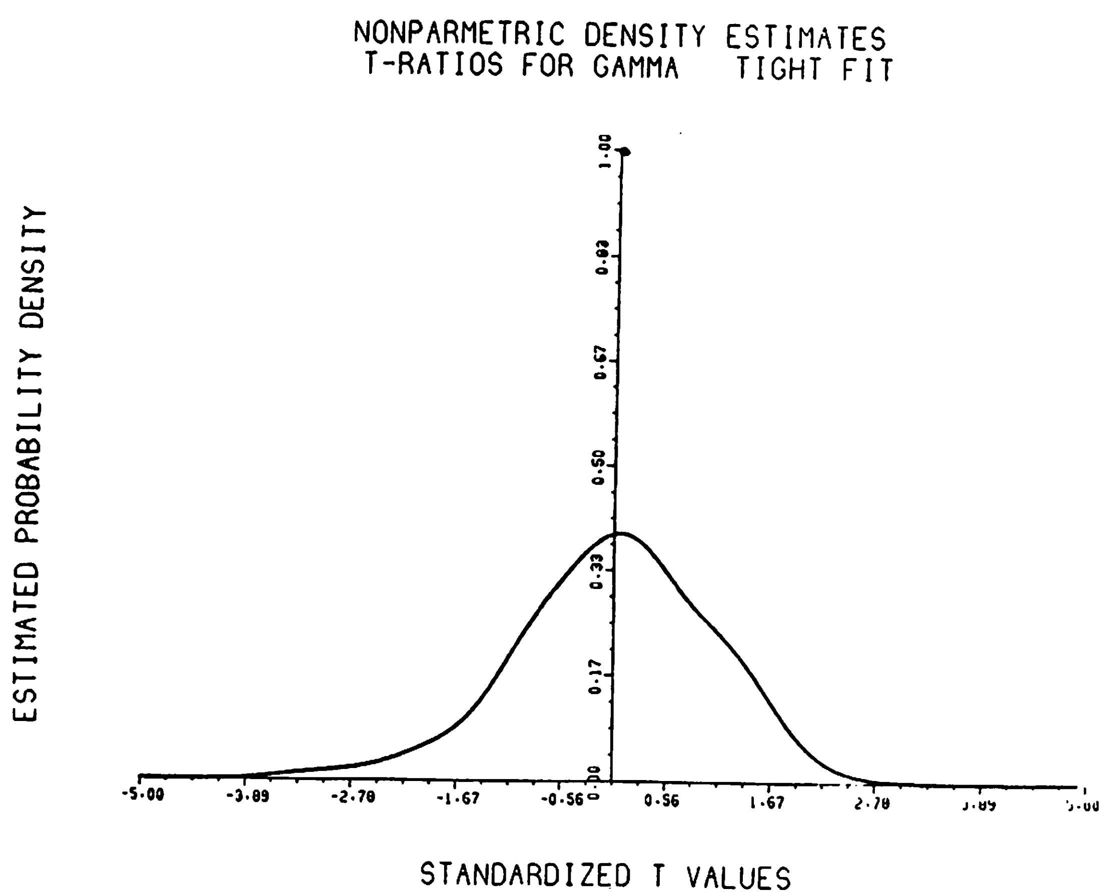

Figure 8

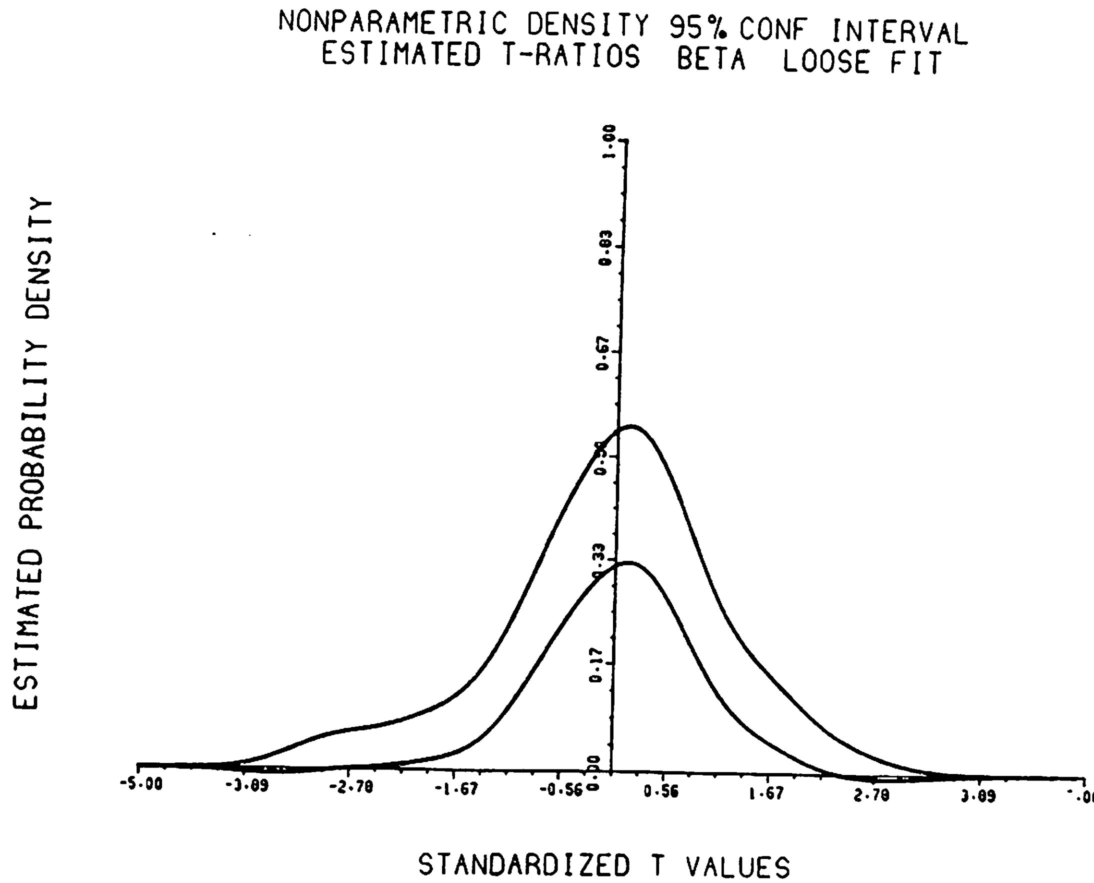

Figure 9

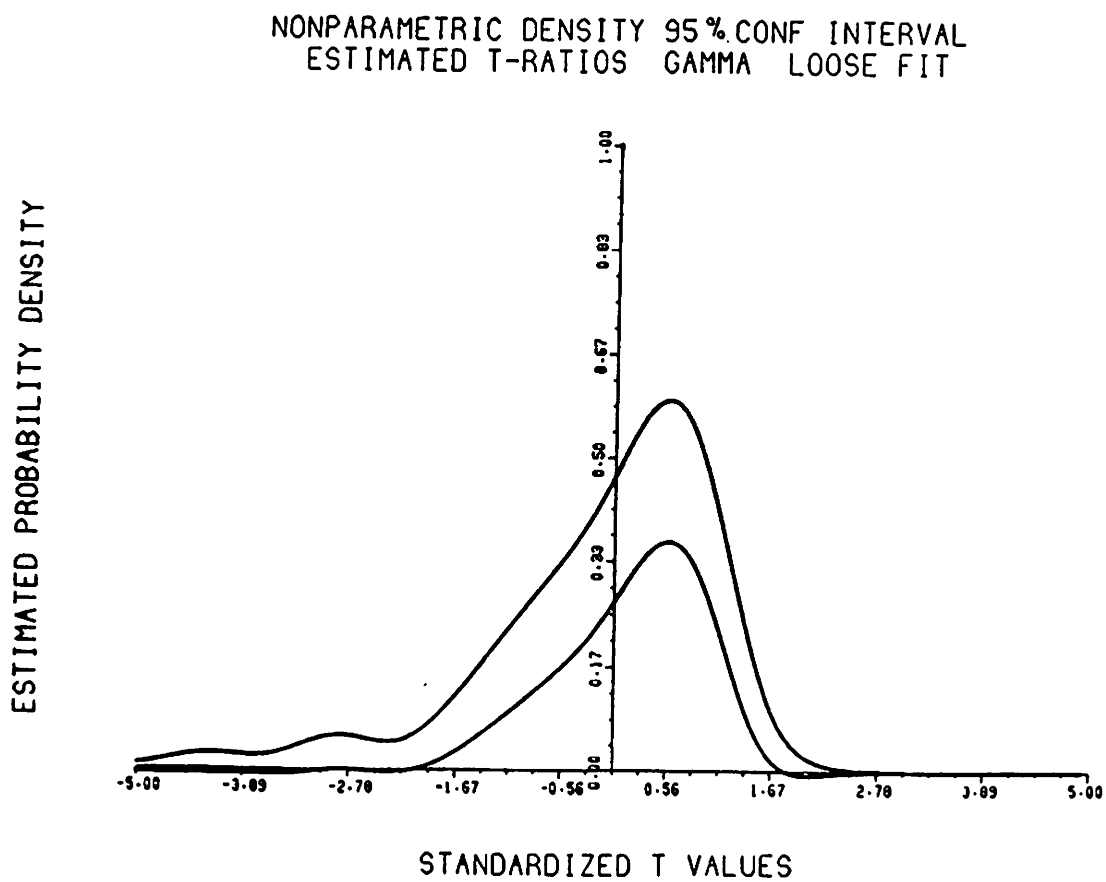

Figure 10

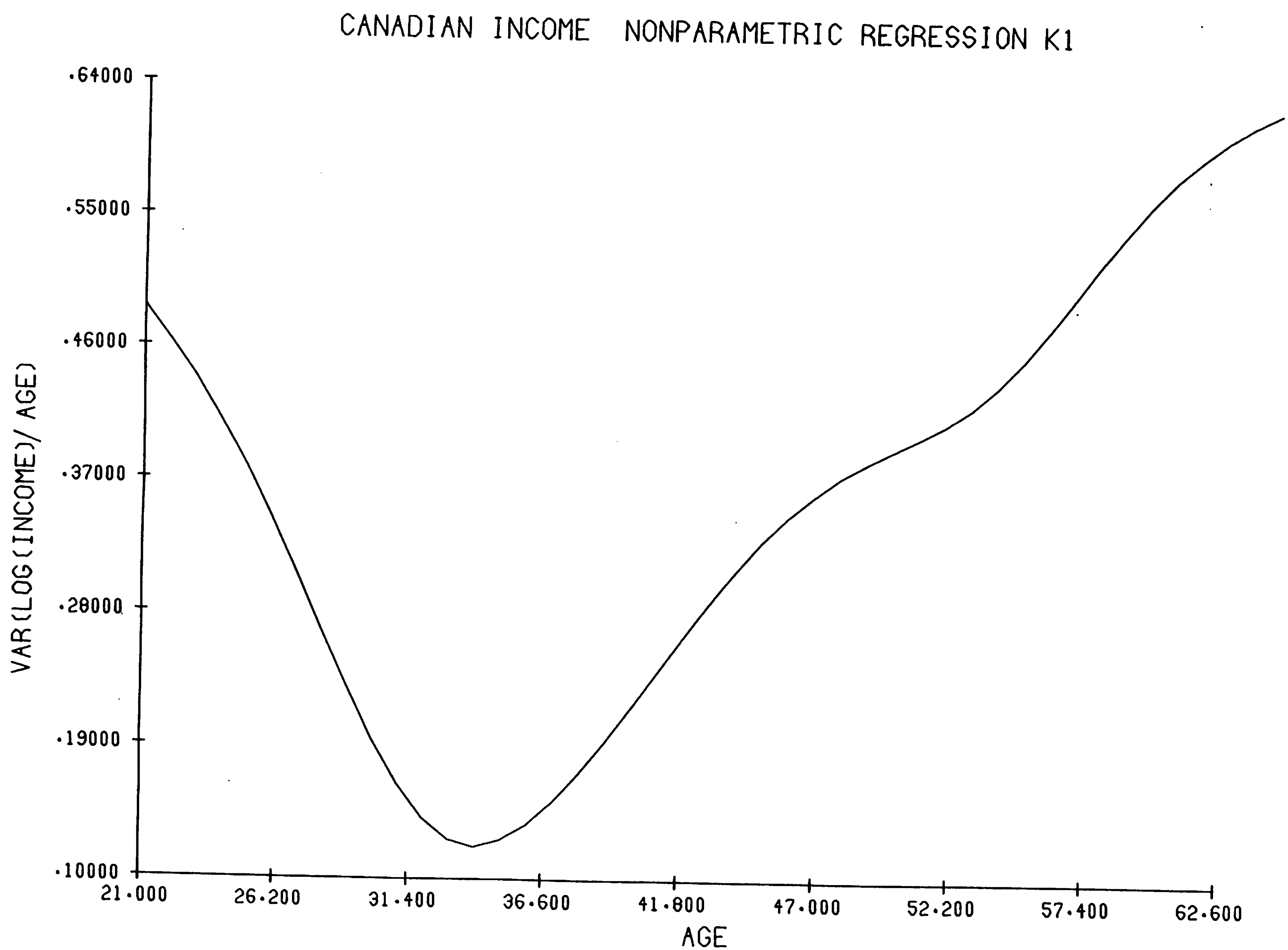
CANADIAN INCOME NONPARAMETRIC REGRESSION K1
VAR(LOG(INCOME)/ AGE)
.64000
.55000
.46000
.37000
.28000
.19000
.10000
21.000
26.200
31.400
36.600
41.800
47.000
52.200
57.400
62.600
AGE

Figure 11