Western University
Scholarship@Western

Department of Economics Research Reports
Economics Working Papers Archive

1989

Nonparametric Estimation of the P-th Derivative of a Regression Function: Stochastic Regressors Case

I A. Ahmad

A Ullah

17784

ISSN:0318-725X
ISBN:0-7714-1100-6

RESEARCH REPORT 8903

NONPARAMETRIC ESTIMATION OF THE P-TH DERIVATIVE* OF A REGRESSION FUNCTION: STOCHASTIC REGRESSORS CASE

by

I. A. Ahmad and A. Ullah

University of Northern Illinois

and

University of Western Ontario

September 1988

ABSTRACT

In this paper we propose a simple nonparametric kernel method of estimating derivatives of the regression function. The regressors in the regression model are considered to be stochastic. The consistency and asymptotic normality results are established.

---

*The work on this paper was done when the first author visited The University of Western Ontario, whose generous hospitality is gratefully acknowledged. The second author acknowledges research support from SSHRC.

NONPARAMETRIC ESTIMATION OF THE P–TH DERIVATIVE*
OF A REGRESSION FUNCTION: STOCHASTIC REGRESSORS CASE

by

I. A. Ahmad and A. Ullah

University of Northern Illinois

and

University of Western Ontario

September 1988

ABSTRACT

In this paper we propose a simple nonparametric kernel method of estimating derivatives of the regression function. The regressors in the regression model are considered to be stochastic. The consistency and asymptotic normality results are established.

---

*The work on this paper was done when the first author visited The University of Western Ontario, whose generous hospitality is gratefully acknowledged. The second author acknowledges research support from SSHRC.

1. INTRODUCTION

Let (Y_i,X_i), $i = 1,\ldots,n$ be an i.i.d. sample such that $Y\epsilon R'$ and $X\epsilon R^d$, with joint density $f(\cdot,\cdot)$ and regression

$$(1.1) \qquad r(x) = E(Y|_{X=x}) = \frac{\int yf(y,x)dy}{f(x)} = \frac{g(x)}{f(x)}$$

provided $E|Y| < \infty$, where $f(x) = \int f(y,x)dy$ is the marginal X–density.

The Nadaraya (1964) and Watson (1964) type kernel estimator of r(x) is given by

$$(1.2) \qquad r_n(x) = \frac{g_n(x)}{f_n(x)} = r_n(x_1,\ldots,x_d)$$

where

$$(1.3) \qquad g_n(x) = \frac{1}{nh^d}\sum_{i=1}^{n} y_i K\left(\frac{x_i - x}{h}\right),\ f_n(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right);$$

$h = h_n$ is a sequence of positive real numbers which tends to 0 as $n \to \infty$ and $k(\cdot)$ is a Borel measurable function (called kernel). Throughout this paper we will treat 0/0 in (1.2) as 0. For detailed references on the nonparametric kernel regression see Collomb (1981), Hardle (1988) and Ullah (1988).

The conditional mean in (1.1) gives a formulation for the regression model as

$$(1.4) \qquad Y = r(x) + u = r(x_1,\ldots,x_d) + u$$

where, by construction, the disturbance term u is such that $E(u|x) = 0$; x and u are independent.

We are interested in estimating $r^{(p)}(x)$, the p–th partial derivative of r(x) with respect to, say the, j–th regressor x_j, $j = 1,\ldots,d$. For $p = 1$, $r^{(1)}(x)$ is the response or regression coefficient of Y with respect to changes in the regression x_j. For example if r(x) is linear the estimation of $r^{(1)}(x)$ is equivalent to estimating the regression coefficient. The estimation of higher order derivatives are also of interest in economics and other applied sciences. For example one may be interested in studying the curvature properties (concavity or convexity) of r(x).

The proposed estimation of $r^{(p)}(x)$ is

(1.5) $$r_n^{(p)}(x) = (\frac{1}{2h})^p \sum_{\ell=0}^{p} (-1)^\ell \binom{p}{\ell} r_n(x + (p-2\ell)h)$$

where $r_n(x + (p-2\ell)h) = r_n(x_1,\ldots,x_j + (p-2\ell)h,\ldots,x_d)$. For p = 1, the estimator $r_n^{(1)}(x)$ and its weak consistency (pointwise) has been studied in Rilstone and Ullah (1987). In this paper we consider $r_n^{(p)}(x)$, $p \geq 1$, and study its weak and strong consistencies (pointwise as well as uniform) results.

We note here that Gasser and Muller (1984) and Georgiev (1984) have earlier studied the estimation of p–th under partial derivations of r(x). However, their estimator is restricted to model (1.4) with fixed "design" variables whose domain can be transformed into the (0,1) interval, and they have considered the Priestley and Chao (1972) type kernel regression estimator. Also, our theoretical results do not follow from theirs.

The plan of the paper is as follows. In Section 2 we present our main results. Then in Section 3 we give proofs of our results in Section 2.

2. RESULTS

We study here the large sample properties of

(2.1) $$r^{(p)}(x) = (\frac{1}{2h})^p \sum_{\ell=0}^{p} (-1)^\ell \binom{p}{\ell} r_n(x + (p-2\ell)h),$$

where $r_n(x + (p-2\ell)h) = r_n(x_1,\ldots,x_j + (p-2\ell)h,\ldots,x_d)$ is given by (1.2) and (1.3). Throughout we consider K in (1.3) belonging to the class of all Borel measurable bounded functions K(x) such that

(i) $\int K(x)dx = 1$ (ii) $\int |K(x)|dx < \infty$

(iii) $\|x\|^d |K(x)| \to 0$ as $\|x\| \to \infty$ (iv) $\sup|K(x)| < \infty$

(v) $K^{(p)}(x)$ exists.

However, we note here that instead of (2.2) one can consider the kernels K as bounded density with compact support whose p–th derivatives exist. We keep (2.2) for the sake of exposition and uniformity in proving all our results.

The conditions we consider are

1. $h \to 0$ as $n \to \infty$

2. $nh^{d+2p} \to \infty$ as $n \to \infty$

3. (i) $f(x)$ is continuous at x

   (ii) $g(x)$ is continuous at x

4. The partial derivatives of $f(x)$ at x exist up to the $p + 1$ order, with the $p + 1$–th partial continuous at x.

5. The partial derivatives of $r(x)$ at x exist up to the $p + 1$ order, with the $p + 1$–th partial continuous at x.

6. $\int y^2 f(y,x)dy$ exists and is continuous at x.

We now state our results on pointwise consistency. The proofs of the following Theorems are in Section 3.

Theorem 1 (Weak Pointwise Consistency). If the conditions 1 to 5 hold, then $r_n^{(p)}(x) \longrightarrow r^{(p)}(x)$ in probability as $n \to \infty$ at every continuity point x of $r^{(p)}(x)$.

Theorem 2 (Strong Pointwise Consistency). Assume that conditions 1, and 3 to 6 hold, and in addition $|Y| \le C$ w.p.1, and for all $\tau > 0$

7. $\sum_{n=1}^{\infty} \exp(-\tau nh^{d+p}) < \infty$

then $r_n^{(p)}(x) \to r^{(p)}(x)$ w.p.1.

Note that $nh^{d+p}/\log n \to \infty$ implies 7.

Next we give results for uniform consistencies.

Theorem 3 (Uniform Weak Consistency). Let the characteristic function or $K(x)$ be absolutely integrable. If conditions 1 and 6 hold, and in addition

8. $nh^{2d+p} \to \infty$, as $n \to \infty$,

9. The p-th order derivatives of $r(x)$ is uniformly continuous.

Then, in probability,

$$\sup_{x \in B} |r_n^{(p)}(x) - r^{(p)}(x)| \to 0 \text{ as, } n \to \infty,$$

B is the support set of $r^{(p)}(x)$.

Theorem 4 (Uniform Strong Consistency) Let $K(x)$ be a function of bounded variation, $|Y| \le C < \infty$ w.p.1. and conditions 1, 5 and 8 hold. Further if, for any $\delta > 0$

10. $\sum_{n=1}^{\infty} \exp(-\delta n h^{2d+p}) < \infty$

then, w.p.1.

$$\sup_{x \in B} |r_n^{(p)}(x) - r^{(p)}(x)| \to 0 \text{ as } n \to \infty$$

Next we present the result on asymptotic normality. But before that we give the following additional conditions.

11. $\int x_i^t K(x-m)dx = m^t, \quad 1 \le t \le p$

12. $\sigma_u^{2+\eta}(x) = E(|u|^{2+\eta}|x)$ exists for some $\eta > 0$ and is continuous at x.

13. $\int |K(w)|^{2+\eta} du < \infty$ for some $\eta > 0$.

Further, we introduce the following notation:

$$\Lambda(x) = \frac{\sigma_u^2(x)}{f(x)} \int K^{(p)}(x)dx. \tag{2.3}$$

Theorem 5 (Asymptotic Normality) Suppose that the conditions 2, 4, 11, 12 and 13 hold. Further, if for any small $\varepsilon > 0$

14. $h \alpha n^{-(\varepsilon+1/(d+2p))}$.

Then, as $n \to \infty$

$$(nh^{d+2p})^{\frac{1}{2}}(r_n^{(p)}(x) - r^{(p)}(x)) \sim N(0, \Lambda(x)). \tag{2.4}$$

3. PROOFS

Here we give the proofs of Theorems 1 to 5 in Section 2. For this we first note that

$$\tilde{r}^{(p)}(x) = (\frac{1}{2h})^p \sum_{\ell=0}^{p} (-1)^\ell \binom{p}{\ell} r(x + (p-2\ell)h) \tag{3.1}$$

converges to $r^{(p)}(x)$ as $n \to \infty$. Thus for the asymptotic properties of $r_n^{(p)}(x) - r^{(p)}(x)$, it suffices for our purposes to discuss the differences

$$(\frac{1}{2h})^p\{r_n(x_\ell) - r(x_\ell)\}, \quad x_\ell = x_{\ell n} = x + (p-2\ell)h \tag{3.2}$$

Such that $x_\ell \to x$ as $n \to \infty$; $r_n^{(p)}(x)$ and $r_n(x_\ell)$ are as defined in (1.5).

Next we state the following results which either follow from the well known results in the literature or can be derived easily, see e.g., Parzen (1962).

Lemma 1 (Asymptotic Mean and Variance). Suppose the conditions 1 to 3 hold. Then, as $n \to \infty$

$$\begin{aligned} &Ef_n(x_\ell) \to f(x)\int K(w+m)dw = f(x) \\ &nh^d V(f_n(x_\ell) \to f(x)\int K^2(w+m)dw; \end{aligned} \tag{3.3}$$

$m = p - 2\ell$. Further, if conditions 1, 2 and 5 hold then

$$\begin{aligned} &Eg_n(x_\ell) \to g(x)\int K(w+m)dw \\ &nh^d V(g_n(x_\ell) \to w^2(x)f(x)\int K^2(w+m)dw \end{aligned} \tag{3.4}$$

where $w^2(x) = \int y^2 f(y,x)dy$.

Lemma 2 (Approximate Bias) Suppose the conditions 4 and 11 hold. Then

(3.5) $$E(f_n(x_\ell) - f(x_\ell)) \simeq c_0 h^{p+1} f^{(p+1)}(x)$$

Further, if 5 and 11 hold, then

(3.6) $$E(g_n(x_\ell) - g(x_\ell)) \simeq c_1 h^{p+1} g^{(p+1)}(x),$$

where c_0 and c_1 are some constants.

Now we can prove our results.

Proof of Theorem 1 For any $\varepsilon > 0$, we note from (3.2) (dropping $(\frac{1}{2})^p$ for the simplicity) that

(3.7) $$P[|r_n(x_\ell) - r(x_\ell)| > h^p\varepsilon] \le P[|g_n(x_\ell) - g(x_\ell)| > \frac{h^p\varepsilon}{2} f_n(x_\ell)]$$

$$+ P[|f_n(x_\ell) - f(x_\ell)| > \frac{h^p\varepsilon}{2} f_n(x_\ell)(r(x_\ell))^{-1}]$$

$$\le P[|g_n(x_\ell) - g(x_\ell)| > \frac{h^p\varepsilon}{2}(f(x_\ell) - \delta)]$$

$$+ P[|f_n(x_\ell) - f(x_\ell)| > \frac{h^p\varepsilon}{2}(f(x_\ell) - \delta)(r(x_\ell))^{-1}]$$

$$+ 2P[|f_n(x_\ell) - f(x_\ell)| > \delta, \quad \text{for any } \delta > 0$$

$$\le \frac{4}{\varepsilon^2}(f(x_\ell) - \delta)^{-2}[h^{-2p}E(g_n(x_\ell) - g(x_\ell)^2$$

$$+ (r(x_\ell))^2 h^{-2p} E(f_n(x_\ell) - f(x_\ell))^2$$

$$+ 2\delta^{-2} E(f_n(x_\ell) - f(x_\ell))^2.$$

Now using the result in lemma 2 and condition 2

(3.8) $$h^{-2p}E(g_n(x_\ell) - g(x_\ell))^2 = h^{-2p}[V(g_n(x_\ell)] + h^{-2p}(\text{Bias } g_n(x_\ell))^2 \to 0 \text{ as } n \to \infty.$$

Similarly, both $h^{-2p}E(f_n(x_\ell) - f(x_\ell))^2$ and $E(f_n(x_\ell) - f(x_\ell))^2 \to 0$ as $n \to \infty$. This establishes the Theorem 1.

Proof of Theorem 2 From lemma 2, $Eg_n(x_\ell) - g(x_\ell) \to 0$ as $n \to \infty$ and $Ef_n(x_\ell) - f(x_\ell) \to 0$ as $n \to \infty$. Thus we need only to show that

$$(3.9) \qquad \sum_{n=1}^{\infty} P[|g_n(x_\ell) - Eg_n(x_\ell)| > \gamma_n] < \infty, \; \gamma_n = \frac{\varepsilon h^p}{2}(f(x_\ell) - \delta)$$

and

$$(3.10) \qquad \sum_{n=1}^{\infty} P[|f_n(x_\ell) - Ef_n(x_\ell)| > \beta_n] < \infty, \; \beta_n = \gamma_n (r(x_\ell))^{-1}.$$

Now observe that

$$g_n(x_\ell) - Eg_n(x_\ell) = \frac{1}{n}\sum_{i=1}^{n} w_{in}$$

where

$$(3.11) \qquad w_{in} = \frac{1}{h^d}\left(Y_i K\left(\frac{x_i - x_\ell}{h}\right) - EY_i K\left(\frac{x_i - x_\ell}{h}\right)\right).$$

If $|Y_i| \le C$ w.p.1. for all i. Then w_{in} are bounded by a constant times h^{-d}. Hence using Bennett's (1962) inequality.

$$(3.12) \qquad P[|g_n(x_\ell) - Eg_n(x_\ell)| > \gamma_n] = P[|\frac{1}{n}\sum_{i=1}^{n} w_{in}| > \gamma_n]$$

$$\le 2\exp\left[-\frac{3n\gamma_n^2}{2(3\sigma_n^2 + c\gamma_n)}\right].$$

where $\gamma_n = 0(h^p)$ and $\sigma_n^2 = V(w_{in}) = 0(h^{-d})$.

Thus using condition 7 we get

$$(3.13) \qquad \sum_{n=1}^{\infty} P[|g_n(x_\ell) - Eg_n(x_\ell)| > \gamma_n] \le B \sum_{n=1}^{\infty} \exp(-nh^{d+p}\tau) < \infty.$$

Similarly, substituting $Y_i = 1$ in w_{in} the result in (3.10) is also true. Now, using the Borel–Cantelli lemma along with (3.7) Theorem 2 follows.

Proof of Theorem 3 Let B denote the set where $r(x) > 0$. Thus to prove the uniform weak consistency of $r_n^{(p)}(x)$ it is enough to show that

$$(3.14) \qquad P[\sup_{x \in B} |r_n(x_\ell) - r(x_\ell)| > h^d \varepsilon] \to 0 \quad \text{as } h \to \infty.$$

But as in the proof of Theorem 2

$$(3.15) \qquad P[\sup_{x \in B} |r_n(x_\ell) - r(x_\ell)| > h^d \varepsilon] \le P[\sup_x |g_n(x_\ell) - g(x_\ell)| > \gamma_n^*]$$

$$+ P[\sup_x |f_n(x_\ell) - f(x_\ell)| > \beta_n^*]$$

where $\gamma_n^* = \varepsilon h^p(\mu_0 - \delta)/2 = 0(h^p)$ and $\beta_n^* = \gamma_n^* \mu = 0(h^p)$; $\mu_0 = \inf f(x)$ and $\mu = \sup r(x)$. By a direct application of Parzen (1962) it is straightforward to show that, under condition (8) and lemma 2, the right hand side of (3.5) tends to zero as $n \to \infty$ in probability. Hence the result in Theorem 3 follows.

Proof of Theorem 4 As in the proof of theorem 2 we need to show here

$$(3.16) \qquad \sum_{n=1}^{\infty} P[\sup_x |g_n(x_\ell) - Eg_n(x_\ell)| > \gamma_n^*] < \infty$$

and

$$(3.17) \qquad \sum_{n=1}^{\infty} P[\sup_x |f_n(x_\ell) - Ef_n(x_\ell)| > \beta_n^*] < \infty$$

where γ_n^* and β_n^* are as in (3.15). But under condition 10, (3.16) and (3.17) follow from the proof of theorem 1 in Nadaraya (1970). Hence the result in theorem 4.

Proof of Theorem 5 First, from (2.1) and (3.1) we write

$$(3.18) \qquad (nh^{d+2p})^{\frac{1}{2}}(r_n^{(p)}(x) - \bar{r}^{(p)}(x)) = \sum_0^p a_\ell (nh^d)^{\frac{1}{2}}(r_n(x_\ell) - r(x_\ell))$$

$$= \sum_0^p a_p (nh^d)^{\frac{1}{2}}\left(r_n(x_\ell) - \frac{Eg_n(x_\ell)}{Ef_n(x_\ell)}\right)$$

$$+ \frac{Eg_n(x_\ell)}{Ef_n(x_\ell)} - r(x_\ell))$$

$$= \sum_0^p a_p (nh^d)^{\frac{1}{2}} (r_n(x_\ell) - \frac{Eg_n(x_\ell)}{Ef_n(x_\ell)} + o(1)$$

where

(3.19) $\quad a_\ell = (-1)^\ell \binom{p}{\ell},$

and we use the result in lemma 2 and condition 14.

Now, using lemmas 1 and 2 and $f(x_\ell) = f(x) + 0(h)$

(3.20) $$r_n(x_\ell) - \frac{Eg_n(x_\ell)}{Ef_n(x_\ell)} = \left[\frac{g_n(x_\ell)}{f(x_\ell)} - \frac{f_n(x_\ell)}{f(x_\ell)} \frac{Eg_n(x_\ell)}{Ef_n(x_\ell)}\right] \frac{f(x_\ell)}{f_n(x_\ell)}$$

has the same asymptotic distribution as that of

(3.21) $$\frac{1}{f(x)}[(g_n(x_\ell) - Eg_n(x_\ell)) - (f_n(x_\ell) - Ef_n(x_\ell))r(x)].$$

Thus the asymptotic distribution of $(nh^{d+2p})^{\frac{1}{2}} (r_n^{(p)}(x) - r^{(p)}(x))$ is the same as that of (using (1.4))

(3.22) $$T_n = (nh^d)^{-\frac{1}{2}} \Sigma(d_i K_i^* - Ed_i K_i^*) + (nh^d)^{-\frac{1}{2}} \Sigma u_i K_i^*$$

$$= T_{1n} + T_{2n},$$

where $d_i = r(x_i) - r(x)$ and

(3.23) $$K_i^* = (f(x))^{-1} \sum_{\ell=0}^{p} a_\ell K(\frac{x_i - x_\ell}{h}).$$

But $T_{1n} \to 0$ as $n \to \infty$, in probability. Further, noting that,

(3.24) $$\int (K^{(p)}(w))^2 dw = \int (\sum_0^p (-1)^\ell K(w - (p-2\ell)/2)^2 dw$$

and using Liapounov's central limit theorem along with the conditions 12 and 13 it follows that $T_{2n} \sim N(0,\Lambda(x))$ as $n \to \infty$, where $\Lambda(x)$ is as in (2.3). This establishes the result in (2.5).

REFERENCES

Bennett, G. (1962), Probability inequalities for the sum of independent random variables, *J. Amer. Statist. Assoc.*, 57, 33–45.

Collomb, A. (1981), Estimation nonparametrique de la regression, revue bibliographique, *International Statistical Review*, 49, 73–93.

Gasser, T. and H. G. Muller (1984), Estimating regression functions and their derivatives by the kernel method, *Scandinavian Journal of Statistics* 11, 171–185.

Georgiev, A.A. (1984), Speed of convergence in nonparametric kernel estimation of a regression function and its derivatives, *Annals of Institute of Statistical Mathematics*, 455–462.

Hardle, W. (1988), Applied Nonparametric Regression, manuscript, University of Bonn.

Nadaraya, E. (1964), On regression estimators, *Theory of Probability and Applications*, 9, 157–159.

Nadaraya, E. (1970), Remarks on nonparametric estimates of density functions and regression curves, *Theory of Probability and Applications* 15, 139–142.

Parzen, E. (1962), On estimation of a probability density function and mode, *Annals of Mathematical Statistics* 33, 1065–1076.

Priestly, M.B. and M. T. Chao (1972), Nonparametric function fitting, *Journal of Royal Statistical Society*, B34, 385–392.

Rilstone, P. and A. Ullah (1987), Nonparametric estimation of response coefficients, manuscript, University of Western Ontario.

Ullah, A. (1988), Nonparametric Estimation of Econometric Functionals, *Canadian Journal of Economics* 21, 625–658.

Watson, G.S. (1966), Smooth regression analysis, *Sankhya*, Ser. A 26: 15, 175–184.