Econometrica, Vol. 63, No. 6 (November, 1995), 1477–1484

NOTES AND COMMENTS

NONPARAMETRIC AND SEMIPARAMETRIC ESTIMATION WITH DISCRETE REGRESSORS

BY MIGUEL A. DELGADO AND JUAN MORA[1]

1. INTRODUCTION AND SUMMARY

THIS NOTE IS CONCERNED with nonparametric and semiparametric inference in regression models where regressors are not continuous. In econometric practice, few explanatory variables are continuous. Many of them are dummies, qualitative variables, or counts; and others, though continuous in nature, are recorded at intervals and can be treated as discrete.

When regressors are discrete with finite support, a mere average of those observations of the dependent variable with the same regressor value will yield a root-n-consistent conditional expectation estimate. We show that sequences of weights constructed in this way are consistent in the sense of Stone (1977), even when the discrete regressors have infinite support, as in the Poisson distribution. This procedure does not require any smoothing.

These results are applied to the estimation of semiparametric models. Frequently, root-n-consistency of parameter estimates is not easy to achieve due to the problem of bias. When regressors are discrete, the bias term exactly equals zero when the nonsmoothing estimate is used. We exploit this fact to derive the asymptotic properties of semiparametric estimates under weaker conditions than those required when regressors are continuous. We discuss in detail the partially linear model (see, e.g., Robinson (1988)). The Central Limit Theorem (CLT) we derive does not require independence between regressors and regression errors, a feature typically present when regressors are continuous. This approach is proven useful in other semiparametric problems.

The relationship between the nonsmoothing estimate and other well-known nonparametric estimation techniques is also analyzed. We show that the nonsmoothing estimator is asymptotically equivalent to the k-nearest neighbors (k-NN) estimator when all regressors are discrete. Using this result it is easily shown that parameter estimates of semiparametric models based on nonsmoothing and k-NN weights are asymptotically equivalent up to the first order. We also discuss the equivalence between other nonparametric estimates and the nonsmoothing weights.

2. NONPARAMETRIC CONSISTENT WEIGHTS WITH DISCRETE REGRESSORS AND ITS APPLICATION IN SEMIPARAMETRIC ESTIMATION

Let (ζ, Z) be an $\mathbb{R}^s \times \mathbb{R}^q$-valued observable random variable such that $E\|\zeta\| < \infty$. We will assume that Z is discrete, that is,

$$(1) \qquad \exists\, \mathscr{D} \subset \mathbb{R}^q,\ \mathscr{D} \text{ countable set, such that } P(Z \subset \mathscr{D}) = 1 \text{ and } z \in \mathscr{D} \Rightarrow P(Z = z) > 0.$$

[1] We are grateful to a co-editor and two anonymous referees for their valuable comments on an earlier version of the paper. This article is based on research funded by Spanish Dirección General de Investigación Científica y Técnica (DGICYT), reference number PB92-0247.

Given a random sample $\{(\zeta_j, Z_j),\ j=1,\dots,n\}$ of (ζ, Z), the regression function $E[\zeta \mid Z=z] \equiv m_\zeta(z)$ is estimated by

$$m_\zeta^{(k)}(z) \equiv \sum_j \zeta_j W_{nj}^{(k)}(z),$$

where $\{W_{nj}^{(k)}(z),\ j=1,\dots,n\}$ is a sequence of weights (the superscript k is used in order to distinguish among different types of weights) and all summations run from 1 to n unless otherwise specified. When Z is discrete, a simple estimate, which does not require any smoothing, can be obtained considering the mean of those observations of the dependent variable ζ_j for which $Z_j = z$. The weights we obtain in this way are

$$W_{nj}^{(1)}(z) \equiv I(Z_j = z) \Big/ \Big(\sum_k I(Z_k = z) \Big),$$

where $I(\cdot)$ is the indicator function and we arbitrarily define $0/0$ to be 0. The corresponding nonparametric estimate $m_\zeta^{(1)}(z)$ is termed *nonsmoothing* estimate. These weights are globally consistent as we state in Theorem 1.

THEOREM 1: *If* (1) *holds,* $E\|\zeta\|^r < \infty$ $(r \geq 1)$ *and* $(\zeta, Z), (\zeta_1, Z_1), \dots, (\zeta_n, Z_n)$ *are independent and identically distributed* (*i.i.d.*) *random variables, then* $E\|m_\zeta^{(1)}(Z) - m_\zeta(Z)\|^r = o(1)$.

Discrete regressors with possibly infinite support are not a problem in some semiparametric models in which the focus of interest is to improve efficiency of the estimates (see, e.g., Robinson (1987) or Newey (1990)). However, in many semiparametric inference problems, a bias term, which increases with the dimension of the regressors set, makes it difficult to achieve root-n-consistency results. Robinson (1988) introduced higher order kernels as a bias reduction technique in semiparametric problems. This approach has been also applied to other semiparametric procedures, like the average derivative method (Powell, Stock, and Stoker (1989)) and shape-invariant modelling (Pinkse and Robinson (1995)), among others.

When regressors are discrete and nonsmoothing weights are used, the bias term is exactly equal to 0 and, hence, no bias reduction techniques are required. In this section we show how this fact can be exploited to obtain asymptotic properties in the semiparametric partly linear regression model.

Suppose (Y, X, Z) is an $\mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$-valued observable random vector such that

$$E[Y \mid X, Z] = \beta' X + \theta(Z) \quad \text{a.s.,} \tag{2}$$

where β is an $\mathbb{R}^p$-valued unknown parameter vector and $\theta(\cdot)$ is an unknown real Borel function. Given a random sample $\{(Y_i, X_i, Z_i),\ i=1,\dots n\}$ from (Y, X, Z), if we define $\varepsilon_{\zeta i} \equiv \zeta_i - E[\zeta_i \mid Z_i]$ (for $\zeta = X, Y$) then $\varepsilon_{Yi} = \beta' \varepsilon_{Xi} + U_i$, where $U_i \equiv Y_i - E[Y_i \mid X_i, Z_i]$ $(1 \leq i \leq n)$. Assume that the following condition holds:

$$E[U_i^2 \mid X_i, Z_i] = E[U_1^2] = \sigma^2 < \infty \text{ and } \Phi \equiv E[\varepsilon_{Xi} \varepsilon_{Xi}'] \text{ is p.d.} \tag{3}$$

Under (2) and (3), the least squares estimate $\bar{\beta} = \{\Sigma_j \varepsilon_{Xj} \varepsilon_{Xj}'\}^{-1} \Sigma_j \varepsilon_{Xj} \varepsilon_{Yj}$ is asymptotically normal with covariance matrix $\sigma^2 \Phi^{-1}$. But this estimate is infeasible because $E[Y_i \mid Z_i]$ and $E[X_i \mid Z_i]$ are unknown. When Z is an absolutely continuous random variable, Robinson (1988) proposed asymptotically efficient estimates of β by estimating the conditional expectations in ε_{Yi} and ε_{Xi}. We follow here this approach.

We shall use "leave-one-out" estimates. Given (ζ_i, Z_i) and $\{(\zeta_j, Z_j),\ j=1,\dots,n,\ j \neq i\}$ i.i.d. random vectors, $m_{\zeta i} \equiv E[\zeta_i \mid Z_i]$ is estimated by $m_{\zeta i}^{(1)} \equiv \Sigma_{j \neq i} \zeta_j W_{nj(-i)}^{(1)}(Z_i)$, where

$W_{nj(-i)}^{(1)}(\cdot)$ are weights as defined above, but which do not employ the ith observation. As stated in Theorem 1, if Z is discrete then $E\|m_{\zeta_1}^{(1)} - m_{\zeta_1}\|^r = o(1)$, whenever $E\|\zeta\|^r < \infty$.

The proposed estimate of β in (1) is obtained as follows. Define $\varepsilon_{\zeta i}^{(1)} \equiv \zeta_i - m_{\zeta i}^{(1)}$, for any random variable ζ; using these estimated residuals for $\zeta_i = Y_i, X_i$ it is possible to estimate Φ, β, and σ^2 by $\Phi^{(1)} \equiv n^{-1}\Sigma_i \varepsilon_{Xi}^{(1)}\varepsilon_{Xi}^{(1)\prime} I_i$, $\beta^{(1)} \equiv \Phi^{(1)-1} n^{-1}\Sigma_i \varepsilon_{Xi}^{(1)}\varepsilon_{Yi}^{(1)} I_i$, and $\sigma^{2(1)} \equiv n^{-1}\Sigma_i(\varepsilon_{Yi}^{(1)} - \beta^{(1)\prime}\varepsilon_{Xi}^{(1)})^2 I_i$, where the function I_i is defined as $I_i \equiv I(\Sigma_{j \neq i} I(Z_j = Z_i) > 0)$. The estimate $\beta^{(1)}$ is asymptotically as efficient as the infeasible estimate $\bar{\beta}$.

THEOREM 2: *Assume that* (1), (2), (3) *hold*, $E[U^4] < \infty$, $E\|X\|^4 < \infty$, *and* $(Y_1, X_1, Z_1), \ldots, (Y_n, X_n, Z_n)$ *are i.i.d. random variables. Then*

$$n^{1/2}(\sigma^{2(1)}\Phi^{(1)-1})^{-1/2}(\beta^{(1)} - \beta) \xrightarrow{d} N(0, I_p).$$

The crucial aspect of Theorem 2 is that the feasible estimate $\beta^{(1)}$ is unbiased. Note that, if $I_i = 1$, then, $\Sigma_{j \neq i} W_{nj}^{(1)}(Z_i)\theta(Z_j) = \theta(Z_i) \Rightarrow \varepsilon_{Yi}^{(1)} = \beta' \varepsilon_{Xi}^{(1)} + \varepsilon_{Ui}^{(1)} \Rightarrow \beta^{(1)} = \beta + \Phi^{(1)-1} n^{-1}\Sigma_i \varepsilon_{Xi}^{(1)}\varepsilon_{Ui}^{(1)} I_i$. So, $E[\beta^{(1)}] = \beta$. Assumption (3) can be easily relaxed, allowing for conditional heteroskedasticity, i.e. $E[U^2 \mid X, Z] = \sigma^2(X, Z)$. In this case, the asymptotic variance of $n^{1/2}(\beta^{(1)} - \beta)$ will be $\Phi^{-1}E[\sigma^2(X_i, Z_i)\varepsilon_{Xi}\varepsilon_{Xi}']\Phi^{-1}$.

When the sample size is small and there are many different values of Z in the sample, it may be convenient to smooth. One of the most popular smoothing procedures is the one which uses k-NN weights, defined as

$$W_{nj}^{(2)}(z) \equiv \left(\sum_{i=1}^{e(j,n,z)} c_{n,d(j,n,z)+i}\right) \Big/ e(j,n,z),$$

where $e(j,n,z) \equiv \#\{i\colon 1 \le i \le n,\ \rho_n(Z_i, z) = \rho_n(Z_j, z)\}$, $d(j,n,z) \equiv \#\{i\colon 1 \le i \le n,\ \rho_n(Z_i, z) < \rho_n(Z_j, z)\}$ (hereafter we suppose that $E\|Z\|^2 < \infty$ and $\rho_n(u,v)$ denotes the Euclidean distance between vectors u and v after scaling their components by the sample standard deviations of the corresponding component of Z; see Stone (1977)), and $c_{n,i}$ $(1 \le i \le n)$ are constants such that

$$\sum_i c_{n,i} = 1, \qquad c_{n,1} \ge \cdots \ge c_{n,n} \ge 0, \quad \text{and} \quad c_{n,i} = 0 \quad \forall\, i > k_n,$$

for a given sequence k_n satisfying $1 \le k_n \le n$. These weights are well motivated when all regressors are discrete. If the number of observations of Z which are equal to z is greater than or equal to k_n, then the k-NN weights are identical to the nonsmoothing weights. Otherwise, some observations of the dependent variable (those whose corresponding regressor value is nearest z) enter in the weighted average. There are different possible k-NN estimates, according to various choices of $c_{n,i}$ (see Stone (1977)). The most popular ones are *uniform* k-NN estimates, for which $c_{n,i} = I(i \le k_n)/k_n$. Applying Stone's (1977) results, we know that some k-NN estimates (e.g., uniform, quadratic, and triangular ones) satisfy a similar result to Theorem 1 if

$$(4) \qquad 1/k_n + k_n/n \to 0 \quad (\text{as } n \to \infty).$$

In fact, in the discrete case any k-NN estimate and the nonsmoothing estimate behave asymptotically in a very similar way under (4), as follows from the following theorem.

THEOREM 3: *If* (1), (4) *hold*, $E\|\zeta\| < \infty$, *and* $(\zeta, Z), (\zeta_1, Z_1), \ldots, (\zeta_n, Z_n)$ *are i.i.d. random variables, then there exists* $q_0 \in (0,1)$ *such that* $P\{m_\zeta^{(1)}(Z) \neq m_\zeta^{(2)}(Z)\} = o(q_0^n)$, *where* $m_\zeta^{(2)}(\cdot)$ *is any k-NN estimate.*

Note that, as a result from Theorem 2, $\forall\ t \geq 0$, $P\{m_{\zeta}^{(1)}(Z) \neq m_{\zeta}^{(2)}(Z)\} = o(n^{-t})$.

Using Theorem 3, it is possible to obtain a similar result to Theorem 2 when k-NN estimates are used in the nonparametric estimation. Specifically, we may define $\beta^{(2)}$, $\sigma^{2(2)}$, and $\Phi^{(2)}$ in the same way as $\beta^{(1)}$, $\sigma^{2(1)}$, and $\Phi^{(1)}$, but replacing $W_{nj(-i)}^{(1)}$ by $W_{nj(-i)}^{(2)}$. Then we have the following corollary.

COROLLARY: *Assume that* (1), (2), (3) *hold*, $E[\theta(Z)^2] < \infty$, $E[U^4] < \infty$, $E\|X\|^4 < \infty$, *and* $(Y_1, X_1, Z_1), \ldots, (Y_n, X_n, Z_n)$ *are i.i.d. random variables. If* (4) *holds and* $W_{nj(-i)}^{(2)}$ *are uniform, quadratic, or triangular k-NN weights, then*

$$n^{1/2}(\sigma^{2(2)}\Phi^{(2)-1})^{-1/2}(\beta^{(2)} - \beta) \xrightarrow{d} N(0, I_p).$$

Note that $\beta^{(2)}$ is no longer unbiased, but those terms which reflect bias may be easily handled thanks to Theorem 3.

A similar result to Theorem 2 may be deduced when kernel weights are used if the support $\mathscr{D}$ contains no accumulation points and the kernel function is bounded. In fact, if these assumptions are satisfied it is straightforward to check that the nonsmoothing and the kernel estimates will coincide for n large enough. If these assumptions are not satisfied, then it is necessary to restrict the probability mass which can be contained in the neighborhoods of any accumulation point; otherwise, it is possible to construct examples in which $P\{m_{\zeta}^{(1)}(Z) \neq m_{\zeta}^{(3)}(Z)\}$ does not converge to 0, where $m_{\zeta}^{(3)}(Z)$ denotes the kernel estimate.

We have performed some Monte Carlo experiments in order to compare the finite-sample behavior of the nonsmoothing, kernel and k-NN estimates in the partly linear regression model. The conclusions of this simulation are by no means a surprise: the nonsmoothing estimate behaves better than the others when the support of $\mathscr{D}$ contains only a few points or when the underlying regression functions $E[Y \mid Z = \varkappa]$ and $E[X \mid Z = \varkappa]$ exhibit high volatility. (We refer the interested reader to Delgado and Mora (1995) for details.)

In many other semiparametric estimation problems implementation of discrete regressors using our methods is straightforward (see Delgado and Mora (1995)).

Departamento de Estadística y Econometría, Universidad Carlos III, 28903 Getafe, Madrid, Spain

and

Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, 03071 San Vicente, Alicante, Spain.

Manuscript received August, 1993; final revision received December, 1994.

APPENDIX: PROOFS

PROOF OF THEOREM 1: We must prove that the sequence $W_{nj}^{(1)}(\cdot)$ satisfies conditions 1–5 of Theorem 1 in Stone (1977). It is straightforward to see that Stone's conditions 2 and 3 hold. The other conditions also hold as it is proved in Propositions 1.1–1.3 below.

PROPOSITION 1.1: *For every nonnegative Borel function* f: $\mathbb{R}^q \to \mathbb{R}$,

$$E[f(Z)] < \infty \Rightarrow E\left[\sum_j W_{nj}^{(1)}(Z) f(Z_j)\right] \leq 2E[f(Z)] \quad \forall n \geq 1.$$

PROOF: $E[\Sigma_j W_{nj}^{(1)}(Z)f(Z_j)] \leq 2E[\Sigma_j f(Z_j)I(Z_j = Z)/(1+\Sigma_k I(Z_k = Z))] = 2nE[f(Z_1)I(Z_1 = Z)/(1+\Sigma_k I(Z_k = Z))] = 2E\{f(Z_1)I(Z=Z_1)E[n/(2+\Sigma_{k=2}^n I(Z_k = Z)) \mid Z, Z_1]\}$. Given $x \in \mathscr{D}$, if we define $B_n^* \equiv \Sigma_{k=2}^n I(Z_k = x)$ and $p_x \equiv P(Z = x)$, then

$$E[n/(2+B_n^*)] = \sum_{s=0}^{n-1} \binom{n-1}{s} p_x^s (1-p_x)^{n-1-s} n/(2+s)$$

$$\leq p_x^{-1} \sum_{s=0}^{n-1} \binom{n}{s+1} p_x^{s+1}(1-p_x)^{n-s-1}$$

$$= p_x^{-1}\{[p_x + (1-p_x)]^n - (1-p_x)^n\} \leq p_x^{-1}.$$

Therefore, if $P(Z)$ is the positive discrete random variable with support $\mathscr{B} = \{p_x : x \in \mathscr{D}\}$ and probability function $P(P(Z) = p_x) = p_x \ \forall p_x \in \mathscr{B}$,

$$E\left[\sum_j W_{nj}^{(1)}(Z)f(Z_j)\right] \leq 2E\left\{f(Z_1)I(Z=Z_1)E\left[n\Big/\left(2+\sum_{k=2}^n I(Z_k = Z)\right)\Bigg| Z, Z_1\right]\right\}$$

$$\leq 2E[f(Z_1)I(Z=Z_1)P(Z)^{-1}]$$

$$= 2E\{f(Z_1)E[I(Z=Z_1)P(Z)^{-1}|Z_1]\}.$$

Given $x \in \mathscr{D}$, the random variable $H(x,Z) = I(Z = x)P(Z)^{-1}$ is discrete and its support contains two values: $P\{H(x,Z)=0\} = 1-p_x$, $P\{H(x,Z) = p_x^{-1}\} = p_x$. Thus, $\forall x \in \mathscr{D}$ $E[H(x,Z)] = 1$ and hence

$$E\left[\sum_j W_{nj}^{(1)}(Z)f(Z_j)\right] \leq 2E\{f(Z_1)E[I(Z=Z_1)P(Z)^{-1}|Z_1]\} = 2E[f(Z_1)]. \qquad Q.E.D.$$

LEMMA 1: *Let Z be a discrete random variable with support $\mathscr{D}$ and probability function $P(Z = x) = p_x$ $\forall x \in \mathscr{D}$; let $Z, Z_1, \ldots, Z_n$ be i.i.d. random variables and $m \in \mathbb{Z}$, $m \geq 0$ (m fixed). Then*

$$\lim_{n\to\infty} nP\left\{\sum_k I(Z_k = Z) = m\right\} = 0.$$

PROOF: $P\{\Sigma_k I(Z_k = Z) = m\} = \Sigma_{x \in \mathscr{D}} P(Z = x)P\{\Sigma_k I(Z_k = Z) = m \mid Z = x\}$. But $\Sigma_k I(Z_k = Z)$ conditional on $Z = x$ has binomial distribution $B(n, p_x)$, where $p_x \equiv P(Z = x)$. Hence,

$$\text{(A.1)} \qquad nP\left\{\sum_k I(Z_k = Z) = m\right\} = n\sum_{x \in \mathscr{D}} p_x \binom{n}{m} p_x^m (1-p_x)^{n-m}$$

$$= n\sum_{x \in \mathscr{D}} p_x^{m+1}\binom{n}{m}\left(\sum_{s=0}^{n-m}(-1)^s\binom{n-m}{s}p_x^s\right)$$

$$= n\sum_{s=0}^{n-m}\binom{n}{m}\binom{n-m}{s}(-1)^s\left(\sum_{x \in \mathscr{D}} p_x^{s+m+1}\right).$$

Define $p_0 = \sup_{x \in \mathscr{D}} p_x < 1$ and $q \in (0, 1-p_0)$. (If $p_0 = 1$, Z is degenerate and Lemma 1 is straightforward.) Then, $\forall k \geq 1$ and $\forall x \in \mathscr{D}$, $(p_x/(1-q))^k \leq p_x/(1-q) < p_x/p_0$. So,

$$\sum_{x \in \mathscr{D}} (p_x/(1-q))^k < \sum_{x \in \mathscr{D}} p_x/p_0 = 1/p_0 \Rightarrow \sum_{x \in \mathscr{D}} p_x^k < (1-q)^k/p_0 < (1-q)^{k-1}/p_0.$$

Hence, (A.1) $\leq np_0^{-1}\binom{n}{m}\Sigma_{s=0}^{n-m}\binom{n-m}{s}(-1)^s(1-q)^{s+m} = p_0^{-1}\binom{n}{m}n(1-q)^m q^{n-m} = o(1)$. *Q.E.D.*

PROPOSITION 1.2: $\Sigma_k W_{nk}^{(1)}(Z) \xrightarrow{p} 1$.

PROOF: $\Sigma_k W_{nk}^{(1)}(Z) = I(\Sigma_k I(Z_k = Z) \neq 0)$. Then, for $\varepsilon > 0$, $P\{|\Sigma_k W_{nk}^{(1)}(Z) - 1| > \varepsilon\} \leq P\{\Sigma_k W_{nk}^{(1)}(Z) = 0\} = P\{\Sigma_k I(Z_k = Z) = 0\} = o(1)$ (by Lemma 1). *Q.E.D.*

PROPOSITION 1.3: $Max_j W_{nj}^{(1)}(Z) \xrightarrow{p} 0$.

PROOF: Given $\varepsilon > 0$, $P\{|\max_j W_{nj}^{(1)}(Z)| > \varepsilon\} = P\{\Sigma_k I(Z_k = Z) \neq 0,\ (\Sigma_k I(Z_k = Z))^{-1} > \varepsilon\} = P\{0 < \Sigma_k I(Z_k = Z) < 1/\varepsilon\}$. Define $\Im(\varepsilon) = \mathbb{N} \cap (0, 1/\varepsilon)$, which is a finite subset of $\mathbb{N}$. Then, $P\{0 < \Sigma_k I(Z_k = Z) < 1/\varepsilon\} = \Sigma_{m \in \Im(\varepsilon)} P\{\Sigma_k I(Z_k = Z) = m\} = o(1)$, since the summation contains a finite number of terms, all of them converging to 0 by Lemma 1. *Q.E.D.*

PROOF OF THEOREM 2: We have

$$\beta^{(1)} = \beta + \Phi^{(1)-1} n^{-1} \sum_i \varepsilon_{Xi}^{(1)} \varepsilon_{Ui}^{(1)} I_i. \tag{A.2}$$

Thus, it suffices to prove that

$$n^{-1/2} \sum_i \varepsilon_{Xi}^{(1)} \varepsilon_{Ui}^{(1)} I_i^{(1)} = n^{-1/2} \sum_i (X_i - m_{Xi}^{(1)})(U_i - m_{Ui}^{(1)}) I_i \xrightarrow{d} N(0, \sigma^2 \Phi), \tag{A.3}$$

$$\Phi^{(1)} \xrightarrow{p} \Phi, \quad \sigma^{2(1)} \xrightarrow{p} \sigma^2. \tag{A.4}$$

Propositions 2.1–2.4 below prove (A.3); (A.4) follows similarly.

PROPOSITION 2.1: $E\|n^{-1/2}\Sigma_i (m_{Xi} - m_{Xi}^{(1)}) m_{Ui}^{(1)} I_i\|^2 = o(1)$.

PROOF: $E\|n^{-1/2}\Sigma_i (m_{Xi} - m_{Xi}^{(1)}) m_{Ui}^{(1)} I_i\|^2$

$$= E[\|m_{X1} - m_{X1}^{(1)}\|^2 m_{U1}^{(1)2} I_1] + (n-1) E[I_1 m_{U1}^{(1)} (m_{X1} - m_{X1}^{(1)})'(m_{X2} - m_{X2}^{(1)}) m_{U2}^{(1)} I_2]. \tag{A.5}$$

The first term in (A.5) converges to 0: applying Cauchy-Schwarz inequality,

$$E[\|m_{X1} - m_{X1}^{(1)}\|^2 m_{U1}^{(1)2} I_1] \leq E[\|m_{X1} - m_{X1}^{(1)}\|^2 m_{U1}^{(1)2}] \leq \{E\|m_{X1} - m_{X1}^{(1)}\|^4 E[m_{U1}^{(1)4}]\}^{1/2}$$

Now, $E\|m_{X1} - m_{X1}^{(1)}\|^4$ converges to 0 (Theorem 1) and $m_{U1}^{(1)}$ is an estimate of $m_{U1} \equiv E[U_1 \mid Z_1] = 0$, and hence $E[m_{U1}^{(1)4}]$ converges to 0 applying also Theorem 1. The second term in (A.5) is exactly equal to 0:

$$\begin{aligned} E[I_1 m_{U1}^{(1)} &(m_{X1} - m_{X1}^{(1)})'(m_{X2} - m_{X2}^{(1)}) m_{U2}^{(1)} I_2] \\ &= \sigma^2 (n-2) \sum_{j \neq 1} \sum_{i \neq 2} E[I_1 (m_{X1} - X_j)'(m_{X2} - X_i) \\ &\quad \times W_{nj}^{(1)}(Z_1) W_{ni}^{(1)}(Z_2) W_{n3}^{(1)}(Z_1) W_{n3}^{(1)}(Z_2) I_2]. \end{aligned}$$

All terms in this expression are 0 because

$$W_{nj}^{(1)}(Z_1) W_{ni}^{(1)}(Z_2) W_{n3}^{(1)}(Z_1) W_{n3}^{(1)}(Z_2) = I(Z_1 = Z_2 = Z_3 = Z_j = Z_i) \Big/ \left(\sum_{k=2}^{n} I(Z_k = Z_1) \right)^4.$$

Q.E.D.

PROPOSITION 2.2: $E\|n^{-1/2}\Sigma_i(X_i - m_{Xi})m_{Ui}^{(1)}I_i\|^2 = o(1)$.

PROOF:

$$E[\|n^{-1/2}\Sigma_i(X_i - m_{Xi})m_{Ui}^{(1)}I_i\|]^2 = E[\|X_1 - m_{Xi}\|^2 m_{U1}^{(1)2} I_1] + \sigma^2(n-2)E\{I_1 W_{n3}^{(1)}(Z_1)W_{n3}^{(1)}(Z_2)I_2 \times E[(X_1 - m_{X1})'(X_2 - m_{X2})|Z_1,\dots,Z_n]\}.$$

Using similar arguments as in Proposition 2.1, the first term converges to 0 and the second one is 0. *Q.E.D.*

PROPOSITION 2.3: $E\|n^{-1/2}\Sigma_i(m_{Xi} - m_{Xi}^{(1)})U_iI_i\|^2 = o(1)$.

PROOF: Use similar arguments as in Propositions 2.1 and 2.2. *Q.E.D.*

PROPOSITION 2.4: $n^{-1/2}\Sigma_i(X_i - m_{Xi})U_iI_i \xrightarrow{d} N(0, \sigma^2\Phi)$.

PROOF: By Central Limit Theorem $n^{-1/2}\Sigma_i(X_i - m_{Xi})U_i \xrightarrow{d} N(0,\sigma^2\Phi)$, since $E[(X - m_x)U] = E\{(X - m_X)E[U \mid X, Z]\} = 0$, $E[(X - m_X)U^2(X - m_X)'] = \sigma^2\Phi$. On the other hand,

$$E\left\|n^{-1/2}\sum_i (X_i - m_{Xi})U_i(1 - I_i)\right\|^2 = \sigma^2 E[\|X_1 - m_{X1}\|^2(1 - I_i)].$$

Applying now Cauchy-Schwarz inequality and Lemma 1 we conclude that this term converges to 0. *Q.E.D.*

PROOF OF THEOREM 3: Denote $k \equiv k_n$. First observe that if $\Sigma_j I(Z_j = Z) \geq k$, then

$$I(Z_i = Z) = 1 \Rightarrow e(i,n,Z) = \sum_j I(Z_j = Z) \quad \text{and} \quad d(i,n,Z) = 0 \Rightarrow W_{ni}^{(2)}(Z) = W_{ni}^{(1)}(Z).$$

Therefore,

$$P\{m_\zeta^{(1)}(Z) \neq m^{(2)}(Z)\} \leq P\left\{\sum_j I(Z_j = Z) < k\right\} = \sum_{m=0}^{k-1} P\left(\sum_j I(Z_j = Z) = m\right) \leq p_0^{-1}\sum_{m=0}^{k-1}\binom{n}{m}(1-q)^m q^{n-m},$$

where in the last equality p_0 and $q \in (0,1)$ are as in Lemma 1. By (4) there exists n_0 such that $n \geq n_0 \Rightarrow k \leq n/2$. So, if $n \geq n_0$

$$P\{m_\zeta^{(1)}(Z) \neq m_\zeta^{(2)}(Z)\} \leq p_0^{-1}\sum_{m=0}^{k-1}\binom{n}{m}(1-p)^m q^{n-m} \leq p_0^{-1}k\binom{n}{k}q^{n-k},$$

where the second inequality holds because the summation contains k terms which are all less or equal than $q^{n-k}n!/[k!(n-k)!]$. Denote $q_0 \equiv q^{1/4} < 1$; then, by Stirling's formula,

$$q_0^{-n}P\{m_\zeta^{(1)}(Z) \neq m_\zeta^{(2)}(Z)\} \sim p_0^{-1}q_0^{n-2k} \times (2\pi)^{-1/2}(q_0 n/(n-k))^{n-k} \times q_0^{n-k}(n/k)^k k^{1/2},$$

where $a_n \sim b_n$ means that $a_n/b_n \to 1$; all terms in this product converge to 0 by (4) (the third term is equal to $\exp\{n \times [(n-k)\log(q_0)/n - (k/n)\log(k/n) + \log(k)/(2n)]\}$). *Q.E.D.*

PROOF OF COROLLARY: With k-NN weights (A.2) no longer holds, but

$$n^{1/2}(\beta^{(2)}-\beta)=\Phi^{(2)-1}\left(n^{-1/2}\sum_i \varepsilon_{Xi}^{(2)}\varepsilon_{\theta i}^{(2)}I_i+n^{-1/2}\sum_i \varepsilon_{Xi}^{(2)}\varepsilon_{Ui}^{(2)}I_i\right).$$

The first term converges to 0 because if A is the event $\{m_{\theta(Z1)}^{(1)}=m_{\theta(Z1)}^{(2)},\ I_1=1\}$ then $E[n\|\varepsilon_{\theta 1}^{(2)}\|^2]=E[\|\theta(Z_1)-m_{\theta(Z1)}^{(2)}\|^2 \mid A^c]\times nP(A^c)$ (here the first factor is bounded because $E[\theta(Z_1)^2]<\infty$ and the second factor converges to 0 by Theorem 2). As for the second term, a similar proof to Theorem 2 applies, but references to Theorem 1 must be replaced by references to Corollary 3 in Stone (1977), where it is proven that these k-NN weights are universally consistent. For example, in Proposition 2.1, (A.5) also holds when k-NN weights are used; but now the first term converges to 0 by universal consistency and the second term converges to 0 because it is equal to

$$\begin{aligned}(\text{A.6})\qquad & E[I_1 m_{U1}^{(2)}(m_{X_1}-m_{X1}^{(2)})'(m_{X_2}-m_{X2}^{(2)})m_{U2}^{(2)}I_2 \mid A_1\cap A_2]\times P(A_1\cap A_2)\\ &+E\left[I_1 m_{U1}^{(2)}(m_{X_1}-m_{X1}^{(2)})'(m_{X_2}-m_{X2}^{(2)})m_{U2}^{(2)}I_2 | (A_1\cap A_2)^c\right]\times P\{(A_1\cap A_2)^c\},\end{aligned}$$

where A_i denotes the event $\{m_{\theta(Zi)}^{(1)}=m_{\theta(Zi)}^{(2)}\}$, for $i=1,2$. (Note that in (A.6) the first term is 0 as in Proposition 2.1 and the second one converges to 0 by Theorem 2). *Q.E.D.*

REFERENCES

DELGADO, M. A., AND J. MORA (1995): "Asymptotic Inferences in Nonparametric and Semiparametric Models with Discrete and Mixed Regressors," mimeo.

NEWEY, W. K. (1990): "Efficient Instrumental Variable Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

PINKSE, C. A. P., AND P. M. ROBINSON (1995): "Pooling Nonparametric Estimates of Regression Functions with a Similar Shape," *Advances in Econometrics and Quantitative Economics*, Vol. in honor of C. R. Rao, ed. by G. S. Maddala et al., 172–197.

POWELL, J. L., J. L. STOCK, AND T. M. STOKER (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

ROBINSON, P. M. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 531–548.

——— (1988): "Root-n-consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

STONE, C. J. (1977): "Consistent Nonparametric Regression," *Annals of Statistics*, 4, 595–645.