Working Paper 93-03
Statistics and Econometrics Series 03
February 1993

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

INFERENCE ON SEMIPARAMETRIC MODELS WITH DISCRETE REGRESSORS

Miguel A. Delgado and Juan Mora*

Abstract

We study statistical properties of coefficient estimates of the partially linear regression model when some or all regressors, in the unknown part of the model, are discrete. The method does not require smoothing in the discrete variables. Unlike when there are continuous regressors, when all regressors are discrete independence between regressors and regression errors is not required. We also give some guidance on how to implement the estimate when there are both continuous and discrete regressors in the unknown part of the model. Weights employed in this paper seem straightforwardly applicable to other semiparametric problems.

Key Words

Semiparametric Partially Linear Model; Discrete Regressors; Empirical Conditional Expectation Estimate; Semiparametric Efficiency Bound; Higher Order Kernels.

*Delgado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid; Mora, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. Research partially supported by Spanish Dirección General de Investigación Científica y Técnica (DGICYT), reference number AP91-05165620.

1. INTRODUCTION

This paper is concerned with inference in the semiparametric partially linear regression model (SPLR model, hereafter) when some (or all) variables entering in the unknown part of the regression model are discrete.

The SPLR model has received considerable attention in Statistics and Econometrics; e.g. Green et al (1985), Denby (1986), Engle et al (1986), Rice (1986), Heckman (1986), Chen (1988), Speckman (1988) and Robinson (1988), to mention only a few. A common feature of the proposed parameter estimates of the SPLR model is that the unknown part of the regression model must be smooth and, therefore, regressors entering in this part must be nonstochastic variables or absolutely continuous random variables. In econometrics practice, few observable variables are continuous. Many of them are dummies, qualitative variables or counts; and other variables, though continuous in nature, are recorded at intervals and can be treated as discrete.

When regressors are discrete, a mere average of those observations of the dependent variable with the same regressor value will yield a consistent conditional expectation estimate. We show, in the following section, that sequences of weights constructed in this way are universally consistent in the sense of Stone (1977). This result suffices to obtain, in Section 3, a Central Limit Theorem (CLT) for the coefficient estimates of the SPLR model when all regressors in the unknown part of the model are discrete. This CLT does not require smoothing and, furthermore, it does not require either independence between regressors and regression errors, a feature typically present when regressors are continuous. In Section 4, we extend the methodology introduced in Section 3 to the case when there are both discrete and continuous regressors in the unknown part of the model. The proposed nonparametric weights are the product of the weights introduced in section 2, which apply to the discrete regressors, and higher order kernels weights, which apply to the continuous regressors. Proofs are confined to an appendix.

2. NONPARAMETRIC CONSISTENT WEIGHTS WITH DISCRETE REGRESSORS

Let Z be a discrete random variable. That is,

$$\exists\ \mathcal{D}\subset\mathbb{R}^q,\ \mathcal{D} \text{ countable set, such that } P(Z\subset\mathcal{D})=1. \qquad [2.1]$$

Let $(\zeta_1,Z_1),\ldots,(\zeta_n,Z_n)$ be independent and identically distributed (i.i.d.) random vectors. The vector of conditional expectations $m_\zeta(\mathfrak{z}) \equiv E[\zeta|Z=\mathfrak{z}]$ can then be estimated by,

$$\hat{m}_\zeta(\mathfrak{z}) = \textstyle\sum_j \zeta_j W_{nj}(\mathfrak{z}), \qquad [2.2]$$

where, hereafter, summations run from 1 to n unless otherwise specified, and

$$W_{nj}(\mathfrak{z}) = I(Z_j=\mathfrak{z})/(\textstyle\sum_k I(Z_k=\mathfrak{z})), \qquad [2.3]$$

where $I(A)$ is the indicator function of event A and, hereafter, we arbitrarily define $0/0$ to be 0. We prove that, under regularity conditions, this sequence of weights is consistent in the sense of Stone (1977)

THEOREM 1: If [2.1] *holds,* $E\|\zeta\|^r<\infty$ *and* (ζ,Z), $(\zeta_1,Z_1),\ldots,(\zeta_n,Z_n)$ *are i.i.d. random vectors, then:*

$$E\|\hat{m}_\zeta(Z)- m_\zeta(Z)\|^r = o(1).$$

PROOF: See appendix. ∎

In semiparametric problems, we need nonparametric estimates evaluated at each data point. Universal consistency results have been used by Robinson (1987) and Newey (1990) in other semiparametric problems. Such results require to estimate $E[\zeta_i|Z_i]$ without using (ζ_i,Z_i) as in Stone (1977). So, given $(\zeta_i,Z_i),(\zeta_1,Z_1),\ldots,(\zeta_{i-1},Z_{i-1}),(\zeta_{i+1},Z_{i+1}),\ldots,(\zeta_n,Z_n)$, i.i.d. random vectors, $m_{\zeta i}(\mathfrak{z}) \equiv E[\zeta_i|Z_i=\mathfrak{z}]$ is estimated by,

$$\hat{m}_{\zeta i}(\mathfrak{z}) = \textstyle\sum_{j\neq i} \zeta_j W_{nj}(\mathfrak{z}),$$

where now $W_{nj}(\mathfrak{z}) = I(Z_j=\mathfrak{z})/(\sum_{k\neq i} I(Z_k=\mathfrak{z}))$. The next corollary follows immediately from theorem 1.

COROLLARY 1.1: If [2.1] *holds,* $E\|\zeta\|^r<\infty$ *and* (ζ_1,Z_1), $\ldots,(\zeta_n,Z_n)$ *are i.i.d. random vectors, then*

$$E\|\hat{m}_{\zeta i}(Z_i)-m_{\zeta i}(Z_i)\|^r = o(1). \quad ∎$$

This corollary will be used in next section for proving asymptotic normality of a feasible estimate of the SPLR model coefficients, and it is also applicable to other semiparametric problems.

The sequence of weights introduced requires no smoothing value. However, it is easy to deduce from theorem 1 a similar property for weights which depend on a smoothing value. Specifically, let's define

$$\tilde{W}_{nj}(z) = \psi((z-Z_j)/h_n)/\sum_k \psi((z-Z_k)/h_n), \qquad [2.5]$$

where

ψ is a function from $\mathbb{R}^q$ to $\mathbb{R}$ with bounded support, [2.6]

$(h_n) \subset \mathbb{R}_+$ is a sequence which converges to 0 as $n \longrightarrow \infty$. [2.7]

We can construct

$$\tilde{m}_\zeta(z) = \sum_j \zeta_j W_{nj}(z). \qquad [2.8]$$

A similar result to theorem 1 can be proved with the additional assumption

$$\exists \ \mu>0 \ \text{ such that } \forall \ z_1, z_2 \in \mathcal{D} \ \|z_1 - z_2\| \geq \mu>0. \qquad [2.9]$$

As a corollary to theorem 1 we have

COROLLARY 1.2: If [2.1], [2.6], [2.7], [2.9] *hold,* $E\|\zeta\|^r<\infty$ *and* (ζ,Z), $(\zeta_1,Z_1),\ldots,(\zeta_n,Z_n)$ *are i.i.d. random vectors, then,*

$$E\|\tilde{m}_\zeta(Z) - m_\zeta(Z)\|^r = o(1).$$

PROOF: See appendix. ∎

A similar result has been obtained by Devroye and Wagner (1980) when regressors are continuous and by Bierens (1987) when there are both continuous and discrete regressors. The advantages and disadvantages of smoothing in semiparametric models with discrete regressors will be discussed in section 3.

3. PARTIALLY LINEAR REGRESSION WITH DISCRETE REGRESSORS

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

$$E[Y|X,Z] = \beta'X + \theta(Z) \quad a.s. \qquad [3.1]$$

where β is an $\mathbb{R}^p$-valued unknown parameter vector and θ is an unknown real function. Given a random sample $\{(Y_i,X_i,Z_i),\ i=1,\ldots n\}$ from (Y,X,Z), if we define $\varepsilon_{\zeta i} \equiv \zeta_i - m_{\zeta i}$, where $m_{\zeta i} \equiv E[\zeta_i|Z_i]$, then,

$$\varepsilon_{Yi} = \beta'\varepsilon_{Xi} + U_i \qquad i=1,2,\ldots, \qquad [3.2]$$

where $U_i = Y_i - E[Y_i|X_i,Z_i]$. Let us assume that the following conditions hold,

$$E[U_i^2|X_i,Z_i] = E[U_i^2] = \sigma^2<\infty, \qquad [3.3]$$

$$\Phi \equiv E[\varepsilon_{Xi}\varepsilon'_{Xi}] \text{ is positive definite (p.d.).} \qquad [3.4]$$

Define $\bar{\Phi} = n^{-1}\sum_i \varepsilon_{Xi}\varepsilon'_{Xi}$. We can construct the following unfeasible estimate:

$$\bar{\beta} = \bar{\Phi}^{-1}n^{-1}\sum_i \varepsilon_{Xi}\varepsilon_{Yi}. \qquad [3.5]$$

Under [3.1], [3.3] and [3.4], $\bar{\beta}$ is asymptotically normal with asymptotic variance

$$\text{AsyVar}(n^{1/2}(\bar{\beta}-\beta)) = \sigma^2\Phi^{-1}. \qquad [3.6]$$

Chamberlain (1992) has shown that [3.6] is a semiparametric asymptotic bound for model [3.1]. Heckman (1986) and Engle et al. (1986) proposed feasible estimates of β using splines, but Rice (1986) proved that the rate of convergence for these estimates was slower than $n^{-1/2}$. Chen (1988) proposed an estimate of β based on a piecewise polynomial estimator of the unknown function θ, whereas Chen and Shiau (1991) proposed a two-stage spline smoothing estimate of β. They both proved that with those estimators the negative result reported in Rice (1986) disappears. Speckman (1988) and Robinson (1988) proposed feasible estimates of β by estimating the conditional expectations in ε_{Yi} and ε_{Xi} -the same approach which we have applied in this paper-. Robinson (1988) proved that it is possible to obtain asymptotically efficient semiparametric estimates of β using higher order kernels weights in a random-design model. Speckman (1988) obtained similar results in a fixed-design model, and his results also applied to the random-design model when considering bias and variance of the estimator conditional on X,Z. However, the asymptotic properties of their estimates are derived assuming that θ and the density function of Z are smooth. These assumptions can not be justified when regressors are discrete. In this case, we can employ the estimates defined in the previous section to construct a feasible estimate for the vector of parameters β.

Using the estimate defined in [2.4] we can obtain residuals

$$\hat{\varepsilon}_{\zeta i} = \zeta_i - \hat{m}_{\zeta i}, \qquad [3.7]$$

for any random variable ζ_i, where $\hat{m}_{\zeta i} = \hat{m}_{\zeta i}(Z_i)$. Using these estimated residuals for $\zeta_i = Y_i, X_i$ it is possible to construct feasible estimates for Φ, β and σ^2. However, it is necessary to make a previous trimming: according to [2.4], if i is an observation such that $\sum_{k\neq i} I(Z_k = Z_i) = 0$, then $\hat{m}_{Yi} = 0$, $\hat{m}_{Xi} = 0$. Therefore, those observations must not be taken into account in order to estimate the parameters of interest. So, let us define the random variable

$$I_i = I(\sum_{k\neq i} I(Z_k=Z_i)>0). \qquad [3.8]$$

We can now construct

$$\hat{\Phi} = n^{-1}\sum_i \hat{\varepsilon}_{Xi}\hat{\varepsilon}'_{Xi}I_i, \qquad [3.9]$$

$$\hat{\beta} = \hat{\Phi}^{-1}n^{-1}\sum_i \hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Yi}I_i, \qquad [3.10]$$

$$\hat{\sigma}^2 = n^{-1}\sum_i(\hat{\varepsilon}_{Yi}-\hat{\beta}'\hat{\varepsilon}_{Xi})^2 I_i. \qquad [3.11]$$

The estimate $\hat{\beta}$ achieves the semiparametric bound [3.6] under certain regularity conditions which are stated in the following theorem.

THEOREM 2: If [2.1], [3.1], [3.3], [3.4] *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 vectors, then*

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

PROOF: See appendix. ■

Note that, unlike Robinson (1988), it is not necessary to assume independence between regressors and regression errors. In addition, no smoothing is required to prove this theorem and the feasible estimate is conditionally unbiased: note that if $I_i = 1$ then,

$$\sum_{j\neq i} W_{nj}(Z_i)\theta(Z_j) = \theta(Z_i). \qquad [3.12]$$

Therefore,

$$\hat{\varepsilon}_{Yi} = \beta'X_i+\theta(Z_i)+U_i-\sum_{j\neq i}W_{nj}(Z_i)(\beta'X_j+\theta(Z_j)+U_j) = \beta'\hat{\varepsilon}_{Xi}+\hat{\varepsilon}_{Ui}. \qquad [3.13]$$

Hence,

$$\hat{\beta} = \beta + \hat{\Phi}^{-1}n^{-1}\sum_i \hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Ui}I_i, \qquad [3.14]$$

and

$$E[(\hat{\beta}-\beta)\,|\,(X_i,Z_i),\ i=1,\ldots n] = 0. \qquad [3.15]$$

Conditional unbiasedness does not hold when regressors are continuous and smoothers are used for computing conditional expectations (see Robinson 1988 and Speckman 1988). Consistent estimates of conditional expectations with discrete regressors can also be obtained using smoothers, as Bierens (1983, 1987) has proposed. Our approach avoids the choice of a smoothing value and, on the other hand, if smoothers are used, [3.12], and then [3.15], do not necessarily hold.

As noted in section 2, when the support of **Z** contains many different points and the sample size is small, it may be convenient to smooth. For instance, variables like "age" take many values and, in small samples, many observations are likely to be thrown out when computing $\hat{\beta}$. Hence, the actual sample size will decrease dramatically. There are at least two ways to solve this problem in practice,

1) Redefine Z by grouping the observations into intervals (so that the number of possible values for **Z** is small) and construct $\hat{\beta}$ according to [3.10].

2) Estimate the parameters of interest using a nonparametric estimate which depends on a smoothing value, i.e., estimate Φ, β and σ^2 with $\tilde{\Phi}$, $\tilde{\beta}$ and $\tilde{\sigma}^2$ –estimates defined in the same way as $\hat{\Phi}$, $\hat{\beta}$, and $\hat{\sigma}^2$ but using the nonparametric estimate defined in [2.8], instead of the nonparametric estimate defined in [2.4].

This latter procedure produces estimates which have the same asymptotic behaviour as the estimates which do not use smoothers. Specifically, in the same way as corollary 1.2 was deduced from theorem 1, it is straightforward to obtain the following corollary from theorem 2.

COROLLARY 2.1: If [2.1], [2.6], [2.7], [2.9], [3.1], [3.3], [3.4] *hold,* $E[U^4]<\infty$, $E\|X\|^4<\infty$ *and* (Y_1,X_1,Z_1), ..., (Y_n,X_n,Z_n) *are i.i.d. random vectors, then,*

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

Therefore, when all regressors in the unknown part of the regression function are discrete, asymptotically there is no difference between the estimate which uses smoothers and the one which does not. With a small sample size, it seems reasonable to smooth if Z may take many different values, but otherwise smoothing is not an advantage.

The homoskedasticity assumption can be easily removed but the asymptotic variance will change in the usual way (see e.g. Eicker 1963 and White 1980).

4. PARTIALLY LINEAR REGRESSION WITH CONTINUOUS AND DISCRETE REGRESSORS

Now, suppose that [3.1] holds for a random vector **Z** such that,

$$\left.\begin{array}{l} Z = (Z^{(1)},Z^{(2)}),\ \text{where } Z^{(1)}\subset\mathbb{R}^s \text{ is discrete and} \\ Z^{(2)}\subset\mathbb{R}^q \text{ is absolutely continuous; } q+s = r,\ q\geq 1,\ s\geq 1. \end{array}\right\} \qquad [4.1]$$

Given (Y_1,X_1,Z_1), (Y_1,X_1,Z_1), ..., $(Y_{1-1},X_{1-1},Z_{1-1})$, $(Y_{1+1},X_{1+1},Z_{1+1})$, ..., (Y_n,X_n,Z_n) we need to define a nonparametric estimate of $m_{\zeta 1}$. Suppose we are given a sequence of weights based on the continuous regressors $\{V_{nj}(Z^{(2)},\mathfrak{z}^{(2)}),j=1,\dots,n\}$. For simplicity, in our semiparametric problem we use a "leave-one-out" estimator. So, if we define $Z^{(2)}_{-1} = (Z^{(2)}_1,\dots,Z^{(2)}_{1-1},Z^{(2)}_{1+1},\dots,Z^{(2)}_n)$, we can then construct, for $j\neq i$,

$$W_{nj}(Z_1)=V_{nj}(Z^{(2)}_{-1},Z^{(2)}_1)I(Z^{(1)}_j=Z^{(1)}_1)/\textstyle\sum_{k\neq 1}V_{nk}(Z^{(2)}_{-1},Z^{(2)}_1)I(Z^{(1)}_k=Z^{(1)}_1). \qquad [4.2]$$

Now, we estimate $m_{\zeta 1}$ by

$$\hat{m}_{\zeta 1} = \textstyle\sum_{j\neq 1}\zeta_j W_{nj}(Z^{(1)}_1,Z^{(2)}_1), \qquad [4.3]$$

for any random variable ζ. Using these estimates it is possible to construct estimated residuals $\hat{\varepsilon}_{\zeta 1}$, as in [3.7], and estimates of the parameters of interest $\hat{\Phi}$, $\hat{\beta}$ and $\hat{\sigma}^2$ as in [3.9], [3.10] and [3.11] respectively, where now,

$$I_1 = I(\textstyle\sum_k V_{nk}(Z^{(2)}_{-1},Z^{(2)}_1)I(Z^{(1)}_k=Z^{(1)}_1)>b_n), \qquad [4.4]$$

where b_n is a sequence of positive real numbers which we will refer to as "sequence of trimming values" for obvious reasons.

We want to obtain a similar result to theorem 2 for the model defined in this section. We will use Nadaraya-Watson kernel estimators (Nadaraya 1964, Watson 1964) for the continuous regressors (though the result probably holds for many other nonparametric weights as well). In this case, [4.2] becomes

$$W_{nj}(Z_1)=K((Z^{(2)}_1-Z^{(2)}_j)/a_n)I(Z^{(1)}_1=Z^{(1)}_j)/\textstyle\sum_{k\neq 1}K((Z^{(2)}_1-Z^{(2)}_k)/a_n)I(Z^{(1)}_1=Z^{(1)}_k). \quad [4.5]$$

where K is a function from $\mathbb{R}^q$ to $\mathbb{R}$ defined by $K(z) = k(z_1)k(z_2)\cdots k(z_q)$, k is a function from $\mathbb{R}$ to $\mathbb{R}$ which we will name "kernel function" and a_n is a sequence of positive real numbers which we will refer to as "sequence of smoothing values".

Some additional assumptions are required to prove that a similar result to theorem 2 holds when there are both continuous and discrete regressors in the unknown part of the model. Given $d\in\mathcal{D}$, we can consider the following functions from $\mathbb{R}^q$ to $\mathbb{R}$: $\theta_d(u) \equiv \theta(d,u)$, $\xi_d(u) \equiv E[X|Z^{(1)}=d, Z^{(2)}=u]$ and $f_d(u)$ is the conditional probability density function of $Z^{(2)}$ given $Z^{(1)}=d$. We will suppose that these functions verify certain differentiability conditions. Inparticular, we require the following assumptions,

$$\exists\ t\in\mathbb{N} : \theta_d \in \mathcal{G}^4_{tq},\ \xi_d \in \mathcal{G}^2_{tq},\ f_d \in \mathcal{G}^\infty_{tq} \text{ uniformly in } \mathcal{D}. \qquad [4.5]$$

The kernel function k verifies that $k \in \mathcal{K}_{2tq}$. [4.6]

The class of functions $\mathcal{G}_{\mu}^{\alpha}$ and $\mathcal{K}_{w}$ are defined in Robinson (1988), and "uniformly in $\mathcal{D}$" means that the constants which appear in the definition of the class of functions $\mathcal{G}_{\mu}^{\alpha}$ must not depend on the value d taken by $Z^{(1)}$. Basically, we have to assume that θ_d, ξ_d and f_d are all, at least, $(tq-1)$-times partially differentiable functions which can be expanded in a Taylor series with a local Lipschitz condition on the remainder, f_d are all bounded functions whose partial derivatives are also bounded, θ_d have all finite moments of order 4 and ζ_d have all finite moments of order 2. In addition, we assume that a kernel of order $2tq$ is used in the continuous nonparametric estimation.

Certain conditions on the rate of convergence of the sequences of trimming values and smoothing values are also required, specifically,

$$b_n \longrightarrow 0, \qquad Nb_n^{-4}a_n^{4tq} \longrightarrow 0, \qquad Nb_n^4a_n^2 \longrightarrow \infty \quad (as\ n \longrightarrow \infty). \quad [4.7]$$

The following theorem justify asymptotic inferences on β.

THEOREM 3: If [3.1], [3.3], [3.4], [4.1], [4.5], [4.6], *and* [4.7] *hold, U is independent of (X,Z),* $E\|X\|^4<\infty$ *and* $(Y_1,X_1,Z_1),\dots,(Y_n,X_n,Z_n)$ *are i.i.d. random vectors, then,*

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

PROOF: See appendix. ∎

Observe that when there are not continuous regressors (i.e. $q=0$), theorem 3 is a weaker result than theorem 2 (because in the latter one it is not required independence between regressors and regression errors), whereas when there are not discrete regressors (i.e. $s=0$), theorem 3 is a weaker result than Robinson's theorem (Robinson, 1988). In fact, it would be possible to establish theorem 3 in a way entirely similar to Robinson's theorem, but we have preferred this weaker version because its proof is somewhat simpler and our assumption [4.7] is more understandable than Robinson's assumption *ix*.

As noted in previous sections, when the support of the discrete variable $Z^{(1)}$ contains many different values, it may be necessary to smooth in the discrete part as well. Provided that the kernel used with the discrete regressors ψ and the smoothing values h_n used with the discrete regressors satisfy [2.6] and [2.7] a similar result to corollary 2.1 may be easily deduced from theorem 3.

In practice, assumption [4.5] is impossible to verify as the functions f_d, θ_d and ξ_d are not known. However, in most situations it will be reasonable to assume certain differentiability in these functions so that theorem 3 may apply. If we suppose that $a_n=O(n^{-c})$ and $b_n=O(n^{-d})$ –what is often the case–, then it is easy to verify that convergence conditions in assumption [4.7] can be interpreted in terms of the set of inequalities for c and d,

$$c>0,\ d>0,\ 1-2tqc+4d<0,\ 1-2qc-4d>0. \qquad [4.8]$$

This means that in a two-dimensional c/d graphic, the point (c,d) chosen must lie within the triangle whose vertices are $(1/q(1+2t),(2t-1)/4(2t+1))$, $(1/4tq,0)$, and $(1/2q,0)$. Table I contains admissible values (c,d) for different q and t. These values have been selected trying to maintain c as close as possible to $(q+4)^{-1}$, which is the optimal smoothing value in nonparametric problems. The choice of d is less important than the choice of c because we can trim as few observations as desired by choosing $b_n=Mn^{-d}$ for a suitable M.

TABLE I

	$t=1$	$t=2$	$t\geq 3$
$q=1$	$c=2/5$, $d=1/21$	$c=1/4$, $d=1/9$	$c=1/5$, $d=1/7$
$q=2$	$c=1/5$, $d=1/21$	$c=1/6$, $d=1/13$	
$q\geq 3$	$c=1/(2q+1)$, $d=1/8(q+1)$		

In semiparametric models, the choice of a_n is not as critical as in nonparametric ones. Hence, we think that in most empirical applications this table may be used as a reference, though further research on the choice of smoothing values in this model is of interest.

Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, 28903-GETAFE, Madrid, Spain.

APPENDIX

Proof of theorem 1: We must prove that the sequence of weights defined in [2.2] 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 1-3 below.

Proposition 1.1.- For every nonnegative Borel function f from $\mathbb{R}^d$ to $\mathbb{R}$ with $E[f(Z)] < \infty$,

$$E[\textstyle\sum_j W_{nj}(Z)f(Z_j)] \le 2E[f(Z)] \qquad \text{for all } n\ge 1$$

PROOF:
$$\begin{aligned} E[\textstyle\sum_j W_{nj}(Z)f(Z_j)] &\le E[\textstyle\sum_j 2f(Z_j)I(Z_j=Z)/(1+\sum_k I(Z_k=Z))] \\ &= nE[2f(Z_1)I(Z_1=Z)/(1+\textstyle\sum_k I(Z_k=Z))] \\ &= E\{2f(Z_1)I(Z=Z_1)E[n/(2+\textstyle\sum_{k=2}^n I(Z_k=Z))\,|\,Z,Z_1]\} \end{aligned}$$

If we define $B_n^* \equiv \sum_{k=2}^n I(Z_k=z))$, and $p_z \equiv P(Z=z)$, then

$$\begin{aligned} E[n/(2+B_n^*)] &= \textstyle\sum_{s=0}^{n-1}\binom{n-1}{s} p_z^s(1-p_z)^{n-1-s} n/(2+s) \\ &\le p_z^{-1}\textstyle\sum_{s=0}^{n-1}\binom{n}{s+1} p_z^{s+1}(1-p_z)^{n-s-1} \\ &\le p_z^{-1}\textstyle\sum_{s=0}^{n-1}\binom{n}{s+1} p_z^{s+1}(1-p_z)^{n-s-1} \\ &= p_z^{-1}[1-(1-p_z)^n] \le p_z^{-1}. \end{aligned}$$

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

$$\begin{aligned} E[\textstyle\sum_j W_{nj}(Z)f(Z_j)] &\le E\{2f(Z_1)I(Z=Z_1)E[n/(2+\textstyle\sum_{k=2}^n I(Z_k=Z))\,|\,Z,Z_1]\} \\ &\le E[2f(Z_1)I(Z=Z_1)P(Z)^{-1}] \\ &= E\{2f(Z_1)E[I(Z=Z_1)P(Z)^{-1}\,|\,Z_1]\} = E[2f(Z_1)]. \end{aligned}$$ ■

Lemma 1.- Let Z be a discrete random variable with support $\mathcal{D}$ and probability function $P(Z=i) = p_i \ \forall\ i\in\mathcal{D}$, let $Z,Z_1,\ldots,Z_n$ be i.i.d. random variables and $m\in\mathbb{Z}$, $m\ge 0$ (m fixed). Then

$$\lim_{n\to\infty} P(\textstyle\sum_k I(Z_k=Z)=m) = 0$$

PROOF: $P(\sum_k I(Z_k=Z)=m) = \sum_{z\in\mathcal{D}} P(Z=z)P(\sum_k I(Z_k=Z)=m\,|\,Z=z)$. But $\sum_k I(Z_k=Z)$ conditional on Z=z has binomial distribution $B(n,p_z)$, where $p_z\equiv P(Z=z)$. Hence,

$$P(\textstyle\sum_k I(Z_k=Z)=m) = \sum_{z\in\mathcal{D}} p_z\binom{n}{m} p_z^m(1-p_z)^{n-m}$$

$$= \sum_{z\in\mathcal{D}} p_z^{m+1}\binom{n}{m}\left(\sum_{s=0}^{n-m}(-1)^s\binom{n-m}{s}p_z^s\right)$$

$$= \sum_{s=0}^{n-m}\binom{n}{m}\binom{n-m}{s}(-1)^s\left(\sum_{z\in\mathcal{D}} p_z^{s+m+1}\right).$$

Define $p_0 = \max_{z\in\mathcal{D}} p_z < 1$ and $q \in (p_0,1)$. (If $p_0 = 1$, Z is degenerate and lemma 1 is straightforward). Then, $\forall\ k\geq 1$ and $\forall\ z\in\mathcal{D}$, $(p_z/(1-q))^k \leq p_z/(1-q) < p_z/p_0$. Hence,

$$\sum_{z\in\mathcal{D}}(p_z/(1-q))^k < \sum_{z\in\mathcal{D}} p_z/p_0 = 1/p_0 \Rightarrow \sum_{z\in\mathcal{D}} p_z^k < (1-q)^k/p_0 < (1-q)^{k-1}/p_0 \Rightarrow$$

$$P(\textstyle\sum_k I(Z_k=Z)=m) = \sum_{s=0}^{n-m}\binom{n}{m}\binom{n-m}{s}(-1)^s\left(\sum_{z\in\mathcal{D}} p_z^{s+m+1}\right)$$

$$< \binom{n}{m}\sum_{s=0}^{n-m}\binom{n-m}{s}(-1)^s(1-q)^{s+m} = \binom{n}{m}(1-q)^m q^{n-m} = o(1). \blacksquare$$

Proposition 1.2.- $\sum_k W_{nk}(Z) \xrightarrow{p} 1.$

PROOF: $\sum_k W_{nk}(Z) = I(\sum_k I(Z_k=Z)\neq 0)$. Then, for $\varepsilon>0$, $P(|\sum_k W_{nk}(Z)-1|>\varepsilon) \leq P(\sum_k W_{nk}(Z)=0) = P(\sum_k I(Z_k=Z)=0) = o(1)$ (by lemma 1). ■

Proposition 1.3.- $V_n(Z,Z_1,\dots,Z_n) \equiv \max_j W_{nj}(Z) \xrightarrow{p} 0.$

PROOF: Given $\varepsilon>0$,

$$P(|V_n|>\varepsilon) = P(\textstyle\sum_k I(Z_k=Z)\neq 0,\ (\sum_k I(Z_k=Z))^{-1}>\varepsilon) = P(0<\sum_k I(Z_k=Z)<1/\varepsilon).$$

Define $\mathfrak{Z}(\varepsilon) = \mathbb{N} \cap (0,1/\varepsilon)$, which is a finite subset of $\mathbb{N}$. Then,

$$P(0<\textstyle\sum_k I(Z_k=Z)<1/\varepsilon) = \sum_{m\in\mathfrak{Z}(\varepsilon)} P(\sum_k I(Z_k=Z)=m) = o(1),$$

(because the summatory has a finite number of terms, each converging to 0 by Lemma 1). ■

Proof of Corollary 1.2: By [2.5] we know that $\exists\ M : \|x\|\geq M \Rightarrow \psi(x) = 0$. By [2.7] there exists n such that $n\geq n_0 \Rightarrow \mu/h_n \geq M$ and then $(\mathfrak{z}-Z_j)/h_n \geq MI(\mathfrak{z}\neq Z_j)$. Hence if $n\geq n_0$, then $\tilde{W}_{nj}(\mathfrak{z}) \equiv W_{nj}(\mathfrak{z})$ and $\tilde{m}_\zeta(\mathfrak{z}) \equiv \hat{m}_\zeta(\mathfrak{z})$. ■

Proof of Theorem 2: From equation [3.14], it suffices to prove that

$$n^{-1/2}\textstyle\sum_i \hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Ui}I_i = n^{-1/2}\sum_i (X_i-\hat{m}_{Xi})(U_i-\hat{m}_{Ui})I_i \xrightarrow{d} N(0,\sigma^2\Phi), \qquad [A.1]$$

$$\hat{\phi} \xrightarrow{p} \Phi, \qquad [A.2]$$

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2. \qquad [A.3]$$

Propositions 2.1-2.4 below prove [A.1]; [A.2] and [A.3] may be easily proved employing similar arguments.

Proposition 2.1.- $E\|n^{-1/2}\sum_i(m_{Xi}-\hat{m}_{Xi})\hat{m}_{Ui}I_i\|^2 = o(1)$.

PROOF: $E\|n^{-1/2}\sum_i(m_{Xi}-\hat{m}_{Xi})\hat{m}_{Ui}I_i\|^2 =$

$$n^{-1}\sum_{i=1}^{n}E[\|m_{Xi}-\hat{m}_{Xi}\|^2\hat{m}_{Ui}^2I_i]+ n^{-1}\sum_i\sum_{j,\,j\neq i}E[I_i\hat{m}_{Ui}(m_{Xi}-\hat{m}_{Xi})'(m_{Xj}-\hat{m}_{Xj})\hat{m}_{Uj}I_j] =$$

$$E[\|m_{X1}-\hat{m}_{X1}\|^2\hat{m}_{U1}^2I_1] + (n-1)E[I_1\hat{m}_{U1}(m_{X1}-\hat{m}_{X1})'(m_{X2}-\hat{m}_{X2})\hat{m}_{U2}I_2].$$

We will prove that the first term converges to 0 and the second one is 0.

For the first term, applying Cauchy-Schwartz inequality,

$$E[\|m_{X1}-\hat{m}_{X1}\|^2\hat{m}_{U1}^2I_1] \leq E[\|m_{X1}-\hat{m}_{X1}\|^2\hat{m}_{U1}^2] \leq \{E\|m_{X1}-\hat{m}_{X1}\|^4E[\hat{m}_{U1}^4]\}^{1/2},$$

$E\|m_{X1}-\hat{m}_{X1}\|^4$ converges to 0 (applying corollary 1) and $\hat{m}_{U1}$ is an estimate of $m_{U1} \equiv E[U_1|Z_1] = 0$, and hence $E[\hat{m}_{U1}^4]$ converges to 0 applying also corollary 1.

As for the second term, defining $\mathfrak{F} = \{X_1,...X_n,Z_1,...,Z_n\}$, then,

$$E[I_1\hat{m}_{U1}(m_{X1}-\hat{m}_{X1})'(m_{X2}-\hat{m}_{X2})\hat{m}_{U2}I_2] =$$

$$E[\sum_{j=3}^{n}I_1(m_{X1}-\hat{m}_{X1})'(m_{X2}-\hat{m}_{X2})U_j^2W_{nj}(Z_1)W_{nj}(Z_2)I_2] =$$

$$(n-2)E\{I_1(m_{X1}-\hat{m}_{X1})'(m_{X2}-\hat{m}_{X2})W_{n3}(Z_1)W_{n3}(Z_2)I_2E[U_3^2|\mathfrak{F}]\} =$$

$$\sigma^2(n-2)\sum_{j,\,j\neq 1}\sum_{i,\,i\neq 2}E[I_1(m_{X1}-X_j)'(m_{X2}-X_i)W_{nj}(Z_1)W_{ni}(Z_2)W_{n3}(Z_1)W_{n3}(Z_2)I_2].$$

All terms in this last expression are 0 because if $W_{ji}^*(Z_1,Z_2,...,Z_n) = W_{nj}(Z_1)W_{ni}(Z_2)W_{n3}(Z_1)W_{n3}(Z_2)$, then,

$$W_{ji}^*(Z_1,Z_2,...,Z_n) = \begin{cases} 1/\sum_{k=2}^{n}I(Z_k=Z_1) & if \quad Z_1=Z_2=Z_3=Z_j=Z_i \\ 0 & otherwise. \end{cases}$$

Therefore:

$$E[I_1(m_{X1}-X_j)'(m_{X2}-X_i)W_{nj}(Z_1)W_{ni}(Z_2)W_{n3}(Z_1)W_{n3}(Z_2)I_2] =$$

$$E\{I_1W_{ji}^*(Z_1,Z_2,...,Z_n)I_2E[(m_{X1}-X_j)'(m_{X2}-X_i)|Z_1,...,Z_n]\} =$$

$$E[I_1W_{ji}^*(Z_1,Z_2,...,Z_n)I_2(m_{X1}-m_{Xj})'(m_{X2}-m_{Xi})] = 0.$$

(The last equality holds because the variable whose expectation is taken is 0; note that, if $W_{ji}^*(Z_1,Z_2,...,Z_n) \neq 0$, then $Z_1=Z_j$ and $Z_2=Z_i$, therefore $(m_{X1}-m_{Xj})'(m_{X2}-m_{Xi}) \equiv (E[X|Z_1]-E[X|Z_j])'(E[X|Z_2]-E[X|Z_i]) = 0)$. ■

Proposition 2.2 $E\|n^{-1/2}\sum_i(X_i-m_{Xi})\hat{m}_{Ui}I_i\|^2 = o(1)$.

PROOF: $E[\|n^{-1/2}\sum_i(X_i-m_{Xi})\hat{m}_{Ui}I_i\|^2] =$

$$n^{-1}\sum_iE[\|X_i-m_{Xi}\|^2\hat{m}_{Ui}^2I_i] + n^{-1}\sum_i\sum_{j,\,j\neq i}E[I_i(X_i-m_{Xi})'\hat{m}_{Ui}\hat{m}_{Uj}(X_j-m_{Xj})I_j] =$$

$$E[\|X_1-m_{X1}\|^2\hat{m}_{U1}^2I_1] + (n-1)E[I_1(X_1-m_{X1})'\hat{m}_{U1}\hat{m}_{U2}(X_2-m_{X2})I_2].$$

$E[\|X_1-m_{X1}\|^2\hat{m}_{U1}^2 I_1] + (n-1)E[I_1(X_1-m_{X1})'\hat{m}_{U1}\hat{m}_{U2}(X_2-m_{X2})I_2].$

The first term converges to 0 as in prop. 2.1. As for the second one,

$$E[I_1(X_1-m_{X1})'\hat{m}_{U1}\hat{m}_{U2}(X_2-m_{X2})I_2] =$$

$$= E[\textstyle\sum_{j=3}^{n} I_1(X_1-m_{X1})'(X_2-m_{X2})U_j^2 W_{nj}(Z_1)W_{nj}(Z_2)I_2]$$

$$= \sigma^2(n-2)E\{I_1W_{n3}(Z_1)W_{n3}(Z_2)I_2E[(X_1-m_{X1})'(X_2-m_{X2})|Z_1,\dots,Z_n]\} = 0. \qquad \blacksquare$$

Proposition 2.3.- $E\|n^{-1/2}\sum_i(m_{Xi}-\hat{m}_{Xi})U_iI_i\|^2 = o(1).$

PROOF: $E\|n^{-1/2}\sum_i(m_{Xi}-\hat{m}_{Xi})U_iI_i\|^2=$

$E[\|m_{X1}-\hat{m}_{X1}\|^2U_1^2I_1] + (n-1)E[I_1U_1(m_{X1}-\hat{m}_{X1})'(m_{X2}-\hat{m}_{X2})U_2I_2].$

The first term converges to 0 (applying CS inequality as in previous propositions) and the second one is 0 (because U_1 and U_2, conditional on $\mathfrak{F}$, are independent random variables whose expectation is 0). ■

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

PROOF: By Central Limit Theorem it follows that,

$$n^{-1/2}\textstyle\sum_i(X_i-m_{Xi})U_i \xrightarrow{d} N(0, \sigma^2\Phi),$$

(because $E[(X-m_X)U]=E\{(X-m_X)E[U|X,Z]\} = 0$ and $E[(X-m_X)U^2(X-m_X)'] = \sigma^2\Phi$).

On the other hand:

$$E\|n^{-1/2}\textstyle\sum_i(X_i-m_{Xi})U_i(1-I_i)\|^2 = n^{-1}(\sum_i E[\|(X_i-m_{Xi})U_i\|^2(1-I_i)] +$$

$$+ \textstyle\sum_i\sum_{j,j\neq i}E[U_i(X_i-m_{Xi})'(X_j-m_{Xj})U_j(1-I_i)(1-I_j)] =$$

$$= E[\|(X_1-m_{X1})U_1\|^2(1-I_1)] = \sigma^2E[\|X_1-m_{X1}\|^2(1-I_1)].$$

(The term with double summation is 0 because (U_i,X_i,Z_i) and (U_j,X_j,Z_j) are independent when $i\neq j$). Applying now Cauchy-Schwartz inequality and lemma 1 we conclude that this final term converges to 0. ■

Proof of Theorem 3: The following lemmas will be used in the proof. They are versions of Robinson's (1988) lemmas adapted to the mixed case. Throughout this section, Robinson will mean Robinson (1988).

Lemma 2.- Let Z be a random variable which satisfies [4.1], f_d the conditional probability density function of $Z^{(2)}$ given $Z^{(1)}=d$, k a function from $\mathbb{R}$ to $\mathbb{R}$ such that $\int|uk(u)|du<\infty$, K *a function defined by* $K(u_1,\dots,u_q)=k(u_1)\cdots k(u_q)$ and a_n a sequence of positive real numbers. If

$u\in\mathbb{R}$) then,

$$h(d,u) \equiv E[\,|K((Z^{(2)}-u)/a_N)|\,I(Z^{(1)}=d)] = O(a_N^q).$$

PROOF: $h(d,u)=P(Z^{(1)}=d)E[\,|K((Z^{(2)}-u)/a_n)|\,|Z^{(1)}=d]=P(Z^{(1)}=d)\int|K((v-c)/a_N)|f_d(v)dv$ $\le MP(Z^{(1)}=d)\left(\int|k(u)|du\right)^q a_N^q\le Ca_N^q$, where $C=M\left(\int|k(u)|du\right)^q<\infty$. ∎

Lemma 3.- Under the same conditions as in lemma 2, if $g(d,u)$ is a function from $\mathbb{R}^r$ to $\mathbb{R}$ such that $E[|g(Z^{(1)},Z^{(2)})|] < \infty$ and Z_1, Z_2 are i.i.d. random vectors, then,

$$E[\,|g(Z_1^{(1)},Z_1^{(2)})K((Z_2^{(2)}-Z_1^{(2)})/a_n)|\,I(Z_2^{(1)}=Z_1^{(1)})] = O(a_N^q).$$

PROOF: If $h(Z^{(1)},Z^{(2)})$ is as defined in lemma 2, then

$$E[\,|g(Z_1^{(1)},Z_1^{(2)})K((Z_2^{(2)}-Z_1^{(2)})/a_n)|\,I(Z_2^{(1)}=Z_1^{(1)})]$$
$$= E\{|g(Z_1^{(1)},Z_1^{(2)})|E[K((Z_2^{(2)}-Z_1^{(2)})/a_n)|I(Z_2^{(1)}=Z_1^{(1)})|Z_1^{(1)},Z_1^{(2)}]\}$$
$$= E[\,|g(Z_1^{(1)},Z_1^{(2)})|h(Z^{(1)},Z^{(2)})] \le Ca_n^qE[\,|g(Z_1^{(1)},Z_1^{(2)})|\,] = C'a_N^q$$

(the last inequality holds by lemma 2) where $C'= CE[\,|g(Z_1^{(1)},Z_1^{(2)})|\,]<\infty$. ∎

Lemma 4.- If the assumptions of theorem 2 are satisfied then,

$$E[(1-I_1)] = o(1).$$

PROOF: First we prove that if $d\in\mathcal{D}$, then $E[(1-I_1)|Z^{(1)}=d]$ converges to 0. Let $Z_j^{(2)}(d)$ be the conditional r. v. $Z_j^{(2)}$ given $Z_j^{(1)}= d$. For any $m\in\mathbb{N}$, we have,

$$E[(1-I_1)|Z^{(1)}=d] = P((1/na_n^q)\textstyle\sum_{j=2}^n I(Z_j^{(1)}=d)K((Z_j^{(2)}(d)-Z_1^{(2)}(d))/a_n) < b_n) \le$$

$$P((1/na_n^q)\textstyle\sum_{j=2}^n I(Z_j^{(1)}=d)K((Z_j^{(2)}(d)-Z_1^{(2)}(d))/a_n)<b_n,\ \sum_{j=2}^n I(Z_j^{(1)}=d)>m) +$$

$$P(\textstyle\sum_{j=2}^n I(Z_j^{(1)}=d)\le m) \le P(\sum_{j=2}^n I(Z_j^{(1)}=d)\le m) + P(\sum_{j=2}^n I(Z_j^{(1)}=d)/n<b_n) +$$

$$P((1/\textstyle\sum_{j=2}^n I(Z_j^{(1)}=d)a_n^q)\sum_{j=2}^n I(Z_j^{(1)}=d)K((Z_j^{(2)}(d)-Z_1^{(2)}(d))/a_n<b_n,\ \sum_{j=2}^n I(Z_j^{(1)}=d)>m).$$

Note that the last term is the nonparametric kernel estimator of the conditional density f_d when there are $\sum_{j=2}^n I(Z_j^{(1)}=d)$ observations. Therefore, applying Robinson's prop. 4, we know that this term converges to 0 when the number of observations converges to ∞. This decomposition of $E[(1-I_1)|Z^{(1)}=d]$ proves that it converges to 0: given $\varepsilon>0$, there exists m_0 such that if $m\ge m_0$ then the last term is less or equal than $\varepsilon/3$; as a result from lemma 1, then there exists $n_0\ge m_0$ such that if $n\ge n_0$ then $P(\sum_{j=2}^n I(Z_j^{(1)}=d)\le m_0)<\varepsilon/3$. And as $\sum_{j=2}^n I(Z_j^{(1)}=d)/n$ converges in probability to $P(Z^{(1)}=d)$, then there exists n_1 such that if $n\ge n_1$ then $P(\sum_{j=2}^n I(Z_j^{(1)}=d)/n<b_n) <\varepsilon/3$. Hence, if $n\ge max\{n_1,n_0\}$ then such that if $n \geq n_1$ then $P(\sum_{j=2}^{n} I(Z_j^{(1)}=d)/n < b_n) < \varepsilon/3$. Hence, if $n \geq max(n_1, n_0)$ then $E[(1-I_1)|Z^{(1)}=d] < \varepsilon$.

Finally, given $\delta > 0$, there exists $\mathcal{P} \subset \mathcal{D}$, $\mathcal{P}$ finite, such that $P(Z^{(1)} \subset \mathcal{D}-\mathcal{P}) < \delta/2$; then, $E[(1-I_1)] \leq \sum_{d\in\mathcal{P}} E[(1-I_1)|Z^{(1)}=d] + P(Z^{(1)} \subset \mathcal{D}-\mathcal{P})$ and the first term can be made arbitrarily small because it is a finite sum of terms each one converging to 0. ∎

Lemma 5.- Under the same conditions as in lemma 2, let $g(d,u)$ be a function from $\mathbb{R}^r$ to $\mathbb{R}$ and define $g_d(u) = g(d,u)$. If there exist positive real numbers λ, α, μ such that $\forall\ d\in\mathcal{D}$ (and uniformly in d) $f_d \in \mathcal{G}_\lambda^\infty$, $g_d \in \mathcal{G}_\mu^\alpha$ and $k \in \mathcal{K}_{l+m-1}$ (where $l-1<\lambda\leq l$, $m-1<\mu\leq m$ and $\eta=min(\mu,\lambda+1)$) then,

$$E|E[(g(Z_1^{(1)},Z_1^{(2)})-g(Z_2^{(1)},Z_2^{(2)}))K((Z_1^{(2)}-Z_2^{(2)})/a_n)I(Z_1^{(1)}=Z_2^{(1)}|Z_1^{(1)},Z_1^{(2)}]|^\alpha = O(a_n^{\alpha(q+\eta)})$$

PROOF: Similar to lemma 3's proof and applying Robinson's lemma 5 to $E\{|E[(g_d(Z_1^{(2)}(d))-g_d(Z_2^{(2)}(d)))K((Z_1^{(2)}(d)-Z_2^{(2)}(d))/a_n)|^\alpha\}$, where $Z_1^{(2)}(d)$ is the conditional random variable $Z_1^{(2)}$ given $Z_1^{(1)}=d$. ∎

We can now prove theorem 3. It will suffice to prove that

$$N^{-1/2}\sum_i (X_i-\hat{m}_{Xi})(U_i-\hat{m}_{Ui})I_i \xrightarrow{d} N(0,\sigma^2\Phi), \quad [A.4]$$

$$N^{-1}\sum_i (X_i-\hat{m}_{Xi})(X_i-\hat{m}_{Xi})'I_i \xrightarrow{p} \Phi, \quad [A.5]$$

$$N^{-1/2}\sum_i (X_i-\hat{m}_{Xi})(\theta_i-\hat{m}_{\theta i})I_i \xrightarrow{p} 0, \quad [A.6]$$

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2. \quad [A.7]$$

All of these results can be proved in a similar way to Robinson's propositions 1-15 though under our assumptions some of his propositions may be omitted and apply Cauchy-Schwartz inequality may be used instead. (Also observe that Robinson uses an ordinary nonparametric estimator, whereas we use a "leave-one-out" one, so the summations in our proof only run from 2 to n). The lemmas in Robinson's appendix B do not apply any more; instead, the lemmas specified above must be employed. For instance, [A.4] follows from

$N^{-1/2}\sum_i (X_i-m_{Xi})U_iI_i \xrightarrow{d} N(0,\sigma^2\Phi)$ (applying CLT, Cauchy-Schwartz inequality and our previous lemma 4 in a similar way to Robinson's prop. 15);

$N^{-1}E[\sum_i \|m_{Xi}-\hat{m}_{Xi}\|^2 U_i^2 I_i] = O(N^{-1}a_n^{-q}b_n^{-2})$ (as in Robinson's prop. 10);

$N^{-1}E[\sum_i \|X_i-m_{Xi}\|^2 \hat{m}_{Ui}^2 I_i] = O(N^{-1}a_n^{-q}b_n^{-2})$ (as in Robinson's prop. 11);

$N^{-1}E[\sum_i \|\xi_i-\hat{m}_{\xi i}\|^2 U_i^2 I_i] = O(N^{-1}a_n^{-q}b_n^{-2}+ a_n^{2tq}b_n^{-2})$ (as in Robinson's prop. 8);

$N^{-1/2}\|\sum_i (m_{Xi}-\hat{m}_{Xi})\hat{m}_{Ui} I_i\|^2 \xrightarrow{p} 0$ (applying Cauchy-Schwartz inequality and similar results to Robinson's prop. 5 and 13);

$N^{-1/2}\|\sum_i (\xi_i-\hat{m}_{\xi i})\hat{m}_{Ui} I_i\|^2 \xrightarrow{p} 0$ (applying Cauchy-Schwartz inequality and similar results to Robinson's prop. 2 and 13).

[A.5], [A.6] and [A.7] follow in a similar way. ∎

REFERENCES

Bierens, H.J. (1983), "Uniform consistency of kernel estimators of a regression function under generalized conditions", *Journal of the American Statistical Association*, 78, 699-707.

Bierens, H.J. (1987), "Kernel estimators of regression functions", *Advances in Econometrics, Fifth World Congress , Vol. I,* (T.F. Bewley ed.), Cambridge: Cambridge University Press.

Chamberlain, G. (1992), "Efficiency bounds for semiparametric regression", *Econometrica*, 60, 567-596.

Chen, H. (1988), "Convergence rates for parametric components in a partly linear model", *Annals of Statistics*, 16, 136-146.

Chen, H. and Shiau, J.H. (1991), "A two-stage spline smoothing method for partially linear models", *Journal of Statistics, Planning and Inference*, 27, 187-201.

Denby, L. (1986), "Smooth regression function", *Statistical Report* n. 26, AT&T Bell Laboratories.

Devroye, L. and Wagner, T.J. (1980), "Distribution-free consistency results in nonparametric discrimination and regression function estimation", *Annals of Statistics*, 8, 231-239.

Eicker, F. (1963), "Asymptotic normality and consistency of the least squares estimator for families of linear regreessions", *Annals of Mathematical Statistics*, 34, 447-456.

Engle, R.F., Granger, W.J., Rice, J.A. and Weiss, A. (1986), "Semiparametric estimates of the relationship between weather and electricity sales", *Journal of the American Statistical Association*, 81, 310-320.

Green, P., Jennison, C. and Seheult, A. (1985), "Analysis of field experiments by least squares smoothing", *Journal of the Royal Statistical Society, B*, 47, 299-315.

Heckman, N.E. (1986), "Spline smoothing in a partly linear model", *Journal of the Royal Statistical Society, B*, 48, 244-248.

Nadaraya, E.A. (1964), "On estimating regression", *Theory of Probability and its Applications*, 9, 141-142.

Newey, W.K. (1990), "Efficient instrumental variable estimation of nonlinear models", *Econometrica*, 58, 809-837.

Rice, J.A. (1984), "Convergence rates for partially splined models", *Statistics and Probability Letters*, 4, 203-208.

Robinson, P.M. (1987), "Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form", *Econometrica*, 55, 531-548.

Robinson, P.M. (1988), "Root-n-consistent semiparametric regression", *Econometrica*, 56, 931-954.

Speckman, P. (1988), "Kernel smoothing in partially linear models", *Journal of the Royal Statistical Society, B,* 50, 413-446.

Stone, C.J. (1977), "Consistent nonparametric regression", *Annals of Statistics*, 4, 595-645.

Watson, G.S. (1964), "Smooth regression analysis, *Sankhya A*, 26, 359-372.

White, H. (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, 48, 817-838.