Working Paper 94-11
Statistics and Econometrics Series 07
May 1994

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

NONPARAMETRIC AND SEMIPARAMETRIC ESTIMATION WITH DISCRETE REGRESSORS

Miguel A. Delgado and Juan Mora*

Abstract

This paper presents and discusses procedures for estimating regression curves when regressors are discrete and applies them to semiparametric inference problems. We show that pointwise root-*n*-consistency and global consistency of regression curve estimates are achieved without employing any smoothing, even for discrete regressors with unbounded support. These results still hold when smoothers are used, under much weaker conditions than those required with continuous regressors. Such estimates are useful in semiparametric inference problems. We discuss in detail the partially linear regression model and shape-invariant modelling. We also provide some guidance on estimation in semiparametric models where continuous and discrete regressors are present. The paper also includes a Monte Carlo study.

Key Words

Nonparametric Regression; Semiparametric Inference; Discrete Regressors; Empirical Conditional Expectation Estimate; Regressograms; Kernels; Nearest Neighbours; Partially Linear Model; Shape-Invariant Modelling.

*Delgado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid; Mora, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. This article is based on research funded by Spanish Dirección General de Investigación Científica y Técnica (DGICYT), reference number PB92-0247.

NONPARAMETRIC AND SEMIPARAMETRIC ESTIMATION

WITH DISCRETE REGRESSORS[1]

Miguel A. Delgado and Juan Mora

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid

January 1994

ABSTRACT

This paper presents and discusses procedures for estimating regression curves when regressors are discrete and applies them to semiparametric inference problems. We show that pointwise root-*n*-consistency and global consistency of regression curve estimates are achieved without employing any smoothing, even for discrete regressors with unbounded support. These results still hold when smoothers are used, under much weaker conditions than those required with continuous regressors. Such estimates are useful in semiparametric inference problems. We discuss in detail the partially linear regression model and shape-invariant modelling. We also provide some guidance on estimation in semiparametric models where continuous and discrete regressors are present. The paper also includes a Monte Carlo study.

Keywords and Phrases: Nonparametric regression; semiparametric inference; discrete regressors; empirical conditional expectation estimate; regressograms; kernels; nearest neighbours; partially linear model; shape-invariant modelling.

[1] This article is based on research funded by Spanish Dirección General de Investigación Científica y Técnica (DGICYT), reference number PB92-0247.

1. INTRODUCTION

In econometrics practice, few explanatory variables in regression models 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. This chapter is concerned with nonparametric and semiparametric inference in regression models where regressors are not continuous.

When regressors are discrete with bounded 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. In section 2 we show that sequences of weights constructed in this way are also consistent in the sense of Stone (1977), even when the discrete regressors have unbounded support, like the Poisson distribution. This procedure does not require any smoothing. As a corollary we show that, when regressors are discrete, commonly used nonparametric sequences of weights -like regressograms, kernels or *k*-nearest neighbours- are also consistent under weaker conditions than those required in the presence of continuous regressors.

The weights introduced in section 2 are applied, in section 3, to estimation of semiparametric models where root-*n*-consistency of parameter estimates is not easy to achieve due to the problem of bias, which enforces to application of bias reduction techniques like higher order kernels. When all regressors are discrete, there is no bias problem in the estimation of these models.

We consider first the partially linear regression model, see 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. We prove a Central Limit Theorem (CLT) for the coefficient estimates of the partially linear regression model when all regressors in the unknown part of the model are discrete. This CLT does not require independence between regressors and regression errors -a feature typically present when regressors are continuous. Then, heteroskedasticity is allowed. We also provide some guidance on how to deal with discrete and continuous regressors in the unknown part of this regression model. Secondly, we analyse shape-invariant modelling, as suggested by Härdle and Marron (1990) and Pinkse and Robinson (1993). Section 3 also discusses applications in other semiparametric models and concludes with a Monte Carlo experiment. Proofs are confined to an appendix.

2. NONPARAMETRIC CONSISTENT WEIGHTS WITH DISCRETE REGRESSORS

Let Z be an $\mathbb{R}^q$-valued discrete random variable. That is,

$$\exists\ \mathcal{D}\subset\mathbb{R}^q,\ \mathcal{D} \text{ countable set, with } P(Z\subset\mathcal{D})=1 \text{ and } z_1\in\mathcal{D} \Rightarrow P(Z=z_1)>0. \qquad (2.1)$$

Let (ζ_1,Z_1), ..., (ζ_n,Z_n) be independent and identically distributed (i.i.d.) random vectors. In this section we present asymptotic properties of alternative nonparametric estimates of the conditional expectation (or *regression function*) $m_\zeta(z) \equiv E[\zeta|Z=z]$.

2.1. Nonparametric regression estimates for discrete variables

When regressors are discrete, $m_\zeta(z)$ can be estimated by

$$\hat{m}_\zeta(z) = \textstyle\sum_j \zeta_j W_{nj}(z)$$

where, hereafter, summations run from 1 to n unless otherwise stated, and the nonparametric weights are defined as

$$W_{nj}(z) = I(Z_j=z)/(\textstyle\sum_k I(Z_k=z)),$$

where $I(A)$ is the indicator function of event A and, hereafter, we arbitrarily define $0/0$ to be 0. Observe that these nonparametric weights do not require any smoothing value and, hence, we will refer to $\hat{m}_{\zeta}(\mathfrak{z})$ as the *non-smoothing estimate*. When the sample size is small and there are many different values of Z in the sample, it may be convenient to smooth. We will consider three popular nonparametric smoothing estimates of the regression function, the regressogram, kernels and k-nearest neighbours.

Regressogram weights are defined as

$$\mathring{W}_{nj}(\mathfrak{z}) = I(\mathfrak{B}_{nj}(\mathfrak{z}))/(\textstyle\sum_k I(\mathfrak{B}_{nk}(\mathfrak{z}))),$$

where $\mathfrak{B}_{nj}(\mathfrak{z}) = \{\exists\ i,\ 1\le i\le k(n)\ :\ \mathfrak{z}\in\mathfrak{I}_i,\ Z_j\in\mathfrak{I}_i\}$ and $\mathfrak{I}_1, \ldots, \mathfrak{I}_{k(n)}$ are pairwise disjoint subsets such that $\bigcup_{j=1}^{k(n)}\mathfrak{I}_j = \mathbb{R}^q$. The corresponding *regressogram estimate* of $m_{\zeta}(\mathfrak{z})$ is

$$\mathring{m}_{\zeta}(\mathfrak{z}) = \textstyle\sum_j \zeta_j \mathring{W}_{nj}(\mathfrak{z}).$$

When studying its asymptotic properties, we have to assume that

$$V_n \equiv \max_{1\le i\le k(n)} V(\mathfrak{I}_i) \longrightarrow 0, \text{ (as } n \longrightarrow \infty), \tag{2.2}$$

where $V(S)$ denotes the volume of the set S. The main advantage of these weights is that they are easy to compute.

Kernel weights are defined as

$$\tilde{W}_{nj}(\mathfrak{z}) = \psi((\mathfrak{z}-Z_j)/h_n)/\textstyle\sum_k \psi((\mathfrak{z}-Z_k)/h_n),$$

where ψ is a function from $\mathbb{R}^q$ to $\mathbb{R}$ and h_n is a sequence of positive real numbers. The *kernel estimate* of $m_{\zeta}(\mathfrak{z})$ is

$$\tilde{m}_{\zeta}(\mathfrak{z}) = \textstyle\sum_j \zeta_j \tilde{W}_{nj}(\mathfrak{z}).$$

This estimate, which was first defined by Nadaraya (1964) and Watson (1964), is the most popular one in the nonparametric literature. We will assume that

$$\psi \text{ has bounded support and } h_n \longrightarrow 0 \qquad (\text{as } n \longrightarrow \infty). \tag{2.3}$$

The *k-nearest neighbour estimate* of $m_\zeta(\mathfrak{z})$ (hereafter referred to as k-NN estimate) is defined as follows: let $Z^{(j)}$ be the jth coordinate of Z $(1 \le j \le q)$, and s_{nj} the sample standard deviation of $Z_1^{(j)}, \ldots, Z_n^{(j)}$. First of all, we define for $u, v \in \mathbb{R}^q$

$$\rho_n(u,v) = (\textstyle\sum_j ((u^{(j)}-v^{(j)})/s_{nj})^2)^{1/2},$$

where the sum extends over all j, $1 \le j \le q$, such that $s_{nj} > 0$. Let c_{in} $(1 \le i \le n)$ be constants satisfying

$$\textstyle\sum_i c_{in} = 1, \; c_{1n} \ge \ldots \ge c_{nn} \ge 0.$$

Define now for a given i $(1 \le i \le n)$

$$e(i,n,\mathfrak{z}) \equiv \#\{j : 1 \le j \le n, \; \rho_n(Z_j,\mathfrak{z}) = \rho_n(Z_i,\mathfrak{z})\},$$

$$d(i,n,\mathfrak{z}) \equiv \#\{j : 1 \le j \le n, \; \rho_n(Z_j,\mathfrak{z}) < \rho_n(Z_i,\mathfrak{z})\}.$$

A sequence of nonparametric weights can then be defined as

$$\omega_{ni}(\mathfrak{z}) = (\textstyle\sum_{k=1}^{e(i,n,\mathfrak{z})} c_{d(i,n,\mathfrak{z})+k})/e(i,n,\mathfrak{z}).$$

And the corresponding nonparametric estimate of $m_\zeta(\mathfrak{z})$ is

$$\check{m}_\zeta(\mathfrak{z}) = \textstyle\sum_j \zeta_j \omega_{nj}(\mathfrak{z}).$$

Given a sequence k_n, the nonparametric estimate $\check{m}_\zeta(\mathfrak{z})$ is said to be a k-NN estimate if the following condition holds,

$$i > k_n \Rightarrow c_{in} = 0.$$

There are different possible k-NN estimates, according to various choices of the sequence c_{in}. Some possible c_{in} are defined in Stone (1977) (see also Devroye 1978). The uniform k-NN estimate ($c_{in} = I(1 \le i \le k_n)/k_n$) is, possibly, the most popular one. In this case

$$\omega_{ni}(\mathfrak{z}) = \left\{ \begin{array}{ll} 1/k_n & \text{if } \rho_n(Z_i,\mathfrak{z}) < \rho_{nk}(\mathfrak{z}) \\ (k_n - d_{nk}(\mathfrak{z}))/(k_n e_{nk}(\mathfrak{z})) & \text{if } \rho_n(Z_i,\mathfrak{z}) = \rho_{nk}(\mathfrak{z}) \\ 0 & \text{if } \rho_n(Z\ ,\mathfrak{z}) > \rho_{nk}(\mathfrak{z}) \end{array} \right\},$$

where now $\rho_{nk}(\mathfrak{z})$ is the k-th value obtained after sorting the sequence of values $\rho_n(Z_1,\mathfrak{z})$, ..., $\rho_n(Z_n,\mathfrak{z})$, and $d_{nk}(\mathfrak{z})$, $e_{nk}(\mathfrak{z})$ are

$$e_{nk}(\mathfrak{z}) \equiv \#\{j : 1 \le j \le n,\ \rho_n(Z_j,\mathfrak{z}) = \rho_{nk}(\mathfrak{z})\},$$

$$d_{nk}(\mathfrak{z}) \equiv \#\ \{j : 1 \le j \le n,\ \rho_n(Z_j,\mathfrak{z}) < \rho_{nk}(\mathfrak{z})\}.$$

The k-NN weights are intuitively appealing. All nonparametric regression estimates can be viewed as local averages around the point at which regression is evaluated; with the k-NN estimates, one decides how many points are used in these local averages.

2.2. Global consistency

The non-smoothing weights satisfy the following property of global consistency.

THEOREM 1: If (2.1) *holds,* $E\|\zeta\|^r < \infty$ *and* (ζ,Z), (ζ_1,Z_1), ..., (ζ_n,Z_n) *are i.i.d. random vectors then* $E\|\hat{m}_\zeta(Z) - m_\zeta(Z)\|^r = o(1)$. ■

Observe that the non-smoothing weights are not "universally consistent", as defined by Stone (1977), since we must assume that Z is discrete. Global consistency of other weights is proved as a corollary. For regressogram and kernel weights we have to assume that

$$\exists\ \mu>0 : \forall\ z_1, z_2 \in \mathcal{D},\ z_1 \neq z_2 \Rightarrow \|z_1-z_2\| \geq \mu > 0. \tag{2.4}$$

COROLLARY 1: Assume that (2.1), (2.4) hold, $E\|\zeta\|^r<\infty$ *and* (ζ,Z), (ζ_1,Z_1), ..., (ζ_n,Z_n) *are i.i.d. random vectors.*

a) If (2.2) *holds, then* $E\|\mathring{m}_\zeta(Z)- m_\zeta(Z)\|^r = o(1)$.

b) If (2.3) *holds, then* $E\|\tilde{m}_\zeta(Z)- m_\zeta(Z)\|^r = o(1)$. ■

Corollary 1.b has been proved by Devroye and Wagner (1980) considering jointly discrete and continuous regressors and under somewhat stronger conditions than (2.3). Devroye and Wagner need conditions on the kernel function which exclude, among others, Epanechnikov kernel and higher order kernels. They also need conditions on nh_n^q.

As for k-NN weights, applying Stone's (1977) results, we know that if the following condition holds,

$$1/k_n + k_n/n \longrightarrow 0 \quad (\text{as } n \longrightarrow \infty), \tag{2.5}$$

then the k-NN estimates satisfy a similar result to theorem 1. In fact, in the discrete case, the non-smoothing estimates and the k-NN ones are asymptotically equivalent when (2.5) holds.

THEOREM 2: If (2.1), (2.5) *hold,* $E\|Z\|^2<\infty$, (ζ,Z), (ζ_1,Z_1), ..., (ζ_n,Z_n) *are i.i.d. random vectors and* $\check{m}_\zeta(\mathfrak{z})$ *is a* k-*NN estimate then there exists* $q_0\in(0,1)$ *such that* $P(\hat{m}_\zeta(Z)\neq\check{m}_\zeta(Z))= o(q_0^n)$ *-thus,* $P(\hat{m}_\zeta(Z)\neq\check{m}_\zeta(Z))= o(n^{-t})\ \forall\ t\in\mathbb{R}$ *(t fixed).* ■

This result will be used in section 3 for proving root-n-consistency of various semiparametric estimates which utilise k-NN weights.

2.3. Pointwise root-n-consistency

When regressors are discrete, all nonparametric estimates defined above are root-n-consistent. We assume that ζ is an $\mathbb{R}^s$-valued random variable satisfying

$$E[\zeta\zeta'] < \infty. \tag{2.6}$$

Given $\mathfrak{z}\in\mathcal{D}$, denote $p(\mathfrak{z})\equiv P(Z=\mathfrak{z})$, $\Sigma(\mathfrak{z})\equiv Var(\zeta|Z=\mathfrak{z})$ and $\Gamma(\mathfrak{z})\equiv p(\mathfrak{z})^{-1}\Sigma(\mathfrak{z})$, which can be estimated by $\hat{p}(\mathfrak{z})$, $\hat{\Sigma}(\mathfrak{z})$ and $\hat{\Gamma}(\mathfrak{z})$ respectively, defined as

$$\hat{p}(\mathfrak{z}) = n^{-1}\textstyle\sum_j I(Z_j=\mathfrak{z}),$$

$$\hat{\Sigma}(\mathfrak{z}) = \textstyle\sum_j \zeta_j\zeta_j' W_{nj}(\mathfrak{z}) - \hat{m}_\zeta(\mathfrak{z})\hat{m}_\zeta(\mathfrak{z})'$$

$$\hat{\Gamma}(\mathfrak{z}) = \hat{p}(\mathfrak{z})^{-1}\hat{\Sigma}(\mathfrak{z}).$$

Given $\mathfrak{z}_1$, ..., $\mathfrak{z}_f$, let $\Gamma(\mathfrak{z}_1,..,\mathfrak{z}_f)$ and $\hat{\Gamma}(\mathfrak{z}_1,..,\mathfrak{z}_f)$ be block diagonal $sf \times sf$ matrices with components $\Gamma(\mathfrak{z}_j)$ and $\hat{\Gamma}(\mathfrak{z}_j)$ $(1\le j\le f)$ respectively. Then, we have

THEOREM 3: If (2.1), (2.6) *hold,* (ζ_1,Z_1), ..., (ζ_n,Z_n) *are i.i.d. random vectors and* $\mathfrak{z}_j\in\mathcal{D}$ $(j=1,..,f)$ *then*

$$n^{1/2}\begin{pmatrix} \hat{m}_\zeta(\mathfrak{z}_1)-m_\zeta(\mathfrak{z}_1) \\ \cdots\cdots\cdots\cdots \\ \hat{m}_\zeta(\mathfrak{z}_f)-m_\zeta(\mathfrak{z}_f) \end{pmatrix} \xrightarrow{d} N(0,\Gamma(\mathfrak{z}_1,..,\mathfrak{z}_f)),$$

and $\hat{\Gamma}(\mathfrak{z}_1,..,\mathfrak{z}_f) \xrightarrow{p} \Gamma(\mathfrak{z}_1,..,\mathfrak{z}_f)$. ■

As a corollary to theorems 2 and 3 we have,

COROLLARY 2: Assume that (2.1), (2.6) *hold,* (ζ_1,Z_1), ..., (ζ_n,Z_n) *are i.i.d. random vectors and* $\mathfrak{z}_j\in\mathcal{D}$ $(j=1,..,f)$.

a) If (2.2) *and* (2.4) *hold, then*

$$n^{1/2}\begin{pmatrix} \mathring{m}_\zeta(\mathfrak{z}_1)-m_\zeta(\mathfrak{z}_1) \\ \cdots\cdots\cdots\cdots \\ \mathring{m}_\zeta(\mathfrak{z}_f)-m_\zeta(\mathfrak{z}_f) \end{pmatrix} \xrightarrow{d} N(0,\Gamma(\mathfrak{z}_1,..,\mathfrak{z}_f)).$$

b) If (2.3) *and* (2.4) *hold, then*

$$n^{1/2}\begin{pmatrix} \tilde{m}_{\zeta}(\mathfrak{z}_1)-m_{\zeta}(\mathfrak{z}_1) \\ \cdots\cdots\cdots \\ \tilde{m}_{\zeta}(\mathfrak{z}_f)-m_{\zeta}(\mathfrak{z}_f) \end{pmatrix} \xrightarrow{\quad d \quad} N(0,\Gamma(\mathfrak{z}_1,\ldots,\mathfrak{z}_f)).$$

c) If $\breve{m}_{\zeta}(\mathfrak{z}_1)$ *is a* k-*NN estimate and* (2.5) *holds, then*

$$n^{1/2}\begin{pmatrix} \breve{m}_{\zeta}(\mathfrak{z}_1)-m_{\zeta}(\mathfrak{z}_1) \\ \cdots\cdots\cdots \\ \breve{m}_{\zeta}(\mathfrak{z}_f)-m_{\zeta}(\mathfrak{z}_f) \end{pmatrix} \xrightarrow{\quad d \quad} N(0,\Gamma(\mathfrak{z}_1,\ldots,\mathfrak{z}_f)). \qquad \blacksquare$$

Of course, $\Gamma(\mathfrak{z}_1,\ldots,\mathfrak{z}_f)$ in *a), b)* and *c)* can be consistently estimated as in theorem 3.

A similar result to corollary 2.b was established by Bierens (1987) under different conditions.

All theorems and corollaries stated in this section will be used in section 3 to prove asymptotic results in various semiparametric estimation problems.

3. ESTIMATING SEMIPARAMETRIC MODELS WITH DISCRETE REGRESSORS

Discrete regressors with possibly unbounded support are not a problem in some semiparametric models in which the focus of interest is to improve efficiency of the estimates. Stone's (1977) results with *k*-NN weights, allowing for very general regressors, were first applied by Robinson (1987) in semiparametric estimation in order to achieve asymptotic efficiency in regression models in the presence of heteroskedasticity of unknown form (the same result had been obtained by Carroll 1982, using kernels and under much more restrictive conditions on the regressors). These weights have been also applied to other semiparametric estimation problems by Newey (1990), Delgado (1992) and Delgado and Stengos (1993).

In many semiparametric inference problems, however, a bias term, which increases with the dimension of the regressors set, makes it difficult to achieve root-n-consistency results. Robinson (1988) introduced the use of higher order kernels as a bias reduction technique in semiparametric problems. This technique has been also applied to other semiparametric procedures, like the average derivative method (Powell et al. 1989, Härdle and Stoker 1989) and shape-invariant modelling (Pinkse and Robinson 1993), among others.

When regressors are discrete, the bias term exactly equals 0 and, hence, no bias reduction techniques are required. In this section we discuss how this fact can be exploited to obtain asymptotic properties in semiparametric models with discrete regressors. We analyse in detail the partially linear regression model and shape-invariant modelling, and make some remarks about how the same procedure may be used in other semiparametric estimation problems.

As expected, in the mixed continuous-discrete case stronger conditions have to be imposed on the continuous part. However, no new techniques are required and theorems can be proved by combining the arguments in section 2 with the well-developed asymptotic theory for continuous variables. We only analyse the mixed case in the partially linear regression model.

3.1. Partially linear regression model

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

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

where β is an $\mathbb{R}^p$-valued unknown parameter vector and θ is an unknown real function. Given a random sample $\{(Y_i,X_i,Z_i),\ i=1,...n\}$ 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,

$$\varepsilon_{Yi} = \beta'\varepsilon_{Xi} + U_i, \quad i=1,2,\ldots,n,$$

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

$$E[U_i^2|X_i,Z_i] = E[U_i^2] = \sigma^2 < \infty, \tag{3.2}$$

$$\Phi \equiv E[\varepsilon_{Xi}\varepsilon'_{Xi}] \text{ is positive definite (p.d.).} \tag{3.3}$$

Let us define $\bar{\Phi} = n^{-1}\sum_i \varepsilon_{Xi}\varepsilon'_{Xi}$ and the unfeasible estimate $\bar{\beta} = \bar{\Phi}^{-1}n^{-1}\sum_i \varepsilon_{Xi}\varepsilon_{Yi}$. Under (3.1), (3.2) and (3.3), $\bar{\beta}$ is asymptotically normal and

$$\text{AsyVar}(n^{1/2}(\bar{\beta}-\beta)) = \sigma^2\Phi^{-1}. \tag{3.4}$$

Chamberlain (1992) has shown that (3.4) is a semiparametric asymptotic bound for model (3.1) in the absence of heteroskedasticity. 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 is 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 root-n-consistency is achieved. Speckman (1988) and Robinson (1988, 1993) proposed feasible estimates of β by estimating the conditional expectations in ε_{Yi} and ε_{Xi}. We follow here this approach.

Given (ζ_1,Z_1), (ζ_1,Z_1), ..., (ζ_{i-1},Z_{i-1}), (ζ_{i+1},Z_{i+1}), .., (ζ_n,Z_n) i.i.d. random vectors, $m_{\zeta i}(Z_i) \equiv E[\zeta_i|Z_i]$ is estimated by,

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

where now, for $i\neq j$

$$W_{nj}(Z_i) = I(Z_j{=}Z_i)/(\textstyle\sum_{k\neq i} I(Z_k{=}Z_i)). \tag{3.5}$$

Note that this is a "leave-one-out" estimate because ζ_i is not used to estimate $E[\zeta_i|Z_i]$. We use this estimate instead of an ordinary one in order to apply straightforwardly the global consistency results obtained in section 2. Specifically, applying theorem 1 we have that if (2.1) holds, $E\|\zeta\|^r<\infty$ and (ζ_1,Z_1), ..., (ζ_n,Z_n) are i.i.d. random vectors, then

$$E\|\hat{m}_{\zeta i}-m_{\zeta i}\|^r = o(1), \tag{3.6}$$

where $\hat{m}_{\zeta i}\equiv \hat{m}_{\zeta i}(Z_i)$, $m_{\zeta i}\equiv m_{\zeta i}(Z_i)$. With these estimates we can obtain residuals $\hat{\varepsilon}_{\zeta i} = \zeta_i-\hat{m}_{\zeta i}$ for any random variable ζ. 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 (3.5), 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

$$I_i = I(\textstyle\sum_{k\neq i}I(Z_k{=}Z_i)>0).$$

We can now construct $\hat{\Phi} = n^{-1}\sum_i\hat{\varepsilon}_{Xi}\hat{\varepsilon}'_{Xi}I_i$, $\hat{\beta} = \hat{\Phi}^{-1}n^{-1}\sum_i\hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Yi}I_i$ and $\hat{\sigma}^2 = n^{-1}\sum_i(\hat{\varepsilon}_{Yi}-\hat{\beta}'\hat{\varepsilon}_{Xi})^2I_i$. The estimate $\hat{\beta}$ achieves the semiparametric bound (3.4) under certain regularity conditions as stated in the following theorem.

THEOREM 4: If (2.1), (3.1), (3.2), (3.3) *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*

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

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,

$$\textstyle\sum_{j\neq i} W_{nj}(Z_i)\theta(Z_j) = \theta(Z_i). \qquad (3.7)$$

And, therefore, we also have the following equalities,

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

$$\hat{\beta} = \beta + \hat{\Phi}^{-1} n^{-1} \textstyle\sum_i \hat{\varepsilon}_{Xi} \hat{\varepsilon}_{Ui} I_i, \qquad (3.8)$$

$$E[\hat{\beta}-\beta \,|\, (X_i,Z_i),\ i=1,\dots n] = 0. \qquad (3.9)$$

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 be also obtained using smoothers, as has been discussed in section 2. However, the non-smoothing approach avoids the choice of a smoothing value and, on the other hand, if smoothers are used, (3.7), and then (3.9), 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 on computing $\hat{\beta}$ and the actual sample size will decrease dramatically. In such cases it seems reasonable to smooth.

Let us define $\tilde{\Phi}$, $\tilde{\beta}$ and $\tilde{\sigma}^2$ in the same way as $\hat{\Phi}$, $\hat{\beta}$, and $\hat{\sigma}^2$ but using instead of (3.5) leave-one-out kernel weights defined, for $i\neq j$, as

$$\tilde{W}_{nj}(Z_i) = \psi((Z_i - Z_j)/h_n)/\textstyle\sum_{k\neq i}\psi((Z_i - Z_k)/h_n).$$

Then, it is straightforward to obtain,

COROLLARY 3: If (2.1), (2.3), (2.4), (3.1), (3.2), (3.3) *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,

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

The same result follows when a regressogram estimate of $E[\zeta|Z=z]$ is used (i.e., corollary 3 also holds when $\tilde{\Phi}$, $\tilde{\beta}$ and $\tilde{\sigma}^2$ are replaced by $\dot{\Phi}$, $\dot{\beta}$ and $\dot{\sigma}^2$ -constructed using leave-one-out estimates based on regressogram weights- and assumption (2.3) is replaced by assumption (2.2)).

When we use k-NN weights theorem 3 also holds, but this corollary is not as straightforward as with regressogram or kernel weights. Let $\check{m}_{\zeta i}$ be a uniform k-NN estimate of $m_{\zeta}(Z_i)$ obtained as defined in section 2.1 and using as observations the n-1 random variables Z_1, ..., Z_{i-1}, Z_{i+1}, ..., Z_n. Define now $\check{\varepsilon}_{\zeta i} = \zeta_i-\check{m}_{\zeta i}$ for any random variable ζ, and

$$\check{\Phi} = n^{-1}\textstyle\sum_i\check{\varepsilon}_{Xi}\check{\varepsilon}'_{Xi},$$

$$\check{\beta} = \check{\Phi}^{-1}n^{-1}\textstyle\sum_i\check{\varepsilon}_{Xi}\check{\varepsilon}_{Yi},$$

$$\check{\sigma}^2 = n^{-1}\textstyle\sum_i(\check{\varepsilon}_{Yi}-\check{\beta}'\check{\varepsilon}_{Xi})^2.$$

Then, we have,

COROLLARY 4: If (2.1), (2.5), (3.1), (3.2), (3.3), hold, $E[\theta(Z)^2]<\infty$, $E[U^4]<\infty$, $E\|X\|^4<\infty$ and (Y_1,X_1,Z_1), ..., (Y_n,X_n,Z_n) are i.i.d. random vectors, then,

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

Note that $\hat{\beta}$, $\tilde{\beta}$ and $\check{\beta}$ employ non-parametric leave-one-out estimates. In addition, in $\hat{\beta}$ and $\tilde{\beta}$ a previous trimming is made. Such a trimming is not necessary for $\check{\beta}$.

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). Let us assume that, instead of (3.2), we have

$$E[U^2|X,Z] = \sigma^2(X,Z) > 0 \quad a.s. \tag{3.10}$$

The following theorem summarises the results for the heteroskedastic model.

THEOREM 5: If (2.1), (3.1), (3.3), (3.10) *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,*

$$n^{1/2}\hat{\Psi}^{-1/2}(\hat{\beta}-\beta) \xrightarrow{d} N(0,I_k),$$

where the matrix $\hat{\Psi}$ *is defined by*

$$\hat{\Psi} = \hat{\Phi}^{-1}(n^{-1}\textstyle\sum_i(\hat{\varepsilon}_{Yi}-\hat{\beta}'\hat{\varepsilon}_{Xi})^2\hat{\varepsilon}_{Xi}\hat{\varepsilon}'_{Xi}I_i)\hat{\Phi}^{-1}. \quad ■$$

Up to now we have only analysed the case when all regressors are discrete. A similar methodology can be applied when there are both discrete and continuous regressors, though notation and proofs become more lengthy and less intuitive[2]. Suppose that (3.1) holds for a random vector Z such that

$$\left.\begin{array}{l} Z=(Z^{(1)},Z^{(2)}), \text{ where } Z^{(1)}\subset \mathbb{R}^s \text{ is discrete and} \\ Z^{(2)}\subset \mathbb{R}^q \text{ is absolutely continuous; } q+s=r,\ q\geq 1,\ s\geq 1. \end{array}\right\} \tag{3.11}$$

We estimate $m_{\zeta i} \equiv E[\zeta_i|Z_i]$ using Nadaraya-Watson kernel weights (Nadaraya 1964, Watson 1964) for the continuous regressors and the non-smoothing weights for the discrete regressors, i.e.

$$W_{nj}(Z_i)=K_{ij}(a_n)I(Z_i^{(1)}=Z_j^{(1)})/\textstyle\sum_k K_{ik}(a_n)I(Z_i^{(1)}=Z_k^{(1)}),$$

[2] All notation used earlier in this subsection will be redefined now in order to adapt it to the new assumptions.

where hereafter we denote

$$K_{ij}(a_n) \equiv K((Z_i^{(2)}-Z_j^{(2)})/a_n),$$

K is a function from $\mathbb{R}^q$ to $\mathbb{R}$ defined as $K(z) = k(z_1)k(z_2)\cdots k(z_q)$, k is a function from $\mathbb{R}$ to $\mathbb{R}$ ("kernel function") and a_n is a sequence of positive real numbers ("smoothing values"). We estimate $m_{\zeta i}$ by

$$\hat{m}_{\zeta i} = \textstyle\sum_j \zeta_j W_{nj}(Z_i),$$

for any random variable ζ. (Note that, unlike in previous sections, this is not a "leave-one-out" estimator). Using these estimates it is possible to construct estimated residuals $\hat{\varepsilon}_{\zeta i}$ and estimates of the parameters of interest $\hat{\Phi}$, $\hat{\beta}$ and $\hat{\sigma}^2$ as in the discrete case, but now

$$I_i = I(\textstyle\sum_k K_{ik}(a_n) I(Z_i^{(1)}=Z_k^{(1)})/na_n^q > b_n),$$

where b_n is a sequence of positive real numbers (trimming values).

Some additional assumptions are required to prove that a similar result to theorem 4 holds when there are both continuous and discrete regressors in the unknown part of the model. Given $d\in\mathcal{D}$, we denote $\theta_d(u) \equiv \theta(d,u)$, $\xi_d(u) \equiv E[X|Z^{(1)}=d,\ Z^{(2)}=u]$ and $f_d(u)$ is the probability density function of $Z^{(2)}|Z^{(1)}=d$. We will assume that

$$\exists\ t\in\mathbb{N} : \theta_d \in \mathcal{G}^4_{tq},\ \xi_d \in \mathcal{G}^2_{tq},\ f_d \in \mathcal{G}^\infty_{tq} \text{ uniformly in } \mathcal{D}, \tag{3.12}$$

$$\text{the kernel function } k \text{ is in the class } \mathcal{K}_{2tq} \text{ and} \tag{3.13}$$

$$b_n \longrightarrow 0,\quad nb_n^{-4}a_n^{4tq} \longrightarrow 0,\quad nb_n^4a_n^{2q} \longrightarrow \infty,\quad (\text{as } n \longrightarrow\infty). \tag{3.14}$$

Classes $\mathcal{G}^\alpha_\mu$ and $\mathcal{K}_w$ are defined in Robinson 1988, and "uniformly in $\mathcal{D}$" means that the constants which appear in the definition do not depend on the value d. The following theorem justifies asymptotic inferences on β.

THEOREM 6: If (3.1), (3.2), (3.3), (3.11), (3.12), (3.13), (3.14) *hold, U is independent of (X,Z),* $E\|X\|^4<\infty$ *and* (Y_1,X_1,Z_1), ..., (Y_n,X_n,Z_n) *are i.i.d. random vectors, then,*

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

Unlike when all regressors are discrete, in theorem 6 it is required independence between regressors and regression errors. Hence, this result does not follow straightforwardly in the heteroskedastic model.

As noted in previous sections, when the sample size is small and there are many different values of $Z^{(1)}$ in the sample, it may be necessary to smooth in the discrete part as well. In such a case theorem 6 does not apply directly, but similar results to corollaries 3 and 4 may be easily deduced from this theorem.

Assumption (3.12) is difficult to verify as the functions f_d, θ_d and ξ_d are not known. Assumption (3.14) restricts the choice of a_n and b_n: if we suppose that $a_n = Cn^{-c}$ and $b_n = Dn^{-d}$ for real numbers *C*, *D*, *c*, *d*, then (3.14) means that in a two-dimensional *c/d* graphic, the point *(c,d)* must lie within the triangle whose vertices are *(1/q(1+2t),(2t-1)/4(2t+1))*, *(1/4tq,0)*, and *(1/2q,0)*. In practice, if we try to maintain *c* as close as possible to $(q+4tq)^{-1}$ (the optimal smoothing value for the nonparametric estimate), possible values are *c=1/(4tq-1)* and *d=1/(16tq-3)*. In semiparametric models, the choice of a_n is not as critical as in nonparametric ones. In empirical applications these admissible values may be used as a reference.

3.2. Shape-invariant modelling

Let us assume that (ζ,Z), (ζ^*,Z^*) are both $\mathbb{R}\times\mathbb{R}^q$-valued observable random variables such that Z and Z^* are discrete that is,

$$\exists\ \mathcal{D}\subset\mathbb{R}^q,\ \mathcal{D} \text{ countable set, such that } P(Z\subset\mathcal{D})=1,\ P(Z^*\subset\mathcal{D})=1. \qquad (3.15)$$

We will denote $\mathcal{F}$ as the following subset of $\mathcal{D}$:

$$\mathcal{F} \equiv \{z\in\mathcal{D} : P(Z=z)>0 \text{ and } P(Z^*=z)>0\}.$$

Note that we do not require that the probability function of Z and Z^* is positive in exactly the same points, but an assumption on $\mathcal{F}$ will be necessary -see (3.20) below.

Let us suppose that there exists a linear relationship between the regression functions $m(z) \equiv E[\zeta|Z=z]$ and $m^*(z) \equiv E[\zeta^*|Z^*=z]$, that is,

$$\exists\ \theta_0=(\theta_{10},\theta_{20})\in\mathbb{R}^2\ (\theta_{20}\neq 0) \text{ such that } m^*(z)=\theta_{10}+\theta_{20}m(z)\ \forall z\in\mathcal{F}. \tag{3.16}$$

Given independent random samples $\{(\zeta_i,Z_i),i=1,..,n)\}$ and $\{(\zeta^*_j,Z^*_j),j=1,..,n\}$[3], the objective of this section is to propose root-n-consistent estimates of the unknown parameter θ_0. We also discuss how our results may be extended to non-linear semiparametric relationships when regressors are discrete.

The relationship specified in equation (3.16) appears when the two curves $m(z)$ and $m^*(z)$ are noisy versions of a similar function, but there is no reasonable parametric model for each regression function. Figures 1-2 show two sets of 1000 simulated observations for which $(\theta_{10},\theta_{20}) = (10,5)$ and

$$m(z) = z/5 \qquad \text{(Figure 1)}$$

$$m(z) = (z-8)^2/10 \qquad \text{(Figure 2)}$$

(Figures 1 and 2 about here)

In these simulations, Z and Z^* were taken to be Poisson variables with

[3] We assume that the size of both random samples is the same for the sake of simplicity. This assumption is, obviously, not necessary.

mean 8 and the errors were taken independent normal variables with zero mean and variance 1.

Lawton et al. (1972) and Gasser et al. (1984) (among others) provide with examples in which similar models to (3.16) may apply. In econometrics practice, these models are likely to appear when analysing certain microeconomic data. Consider, for instance, the case in which (ζ,Z) are, respectively, "percentage of expenditure on food" and "age of the reference person" for households in a low level of income and (ζ^*,Z^*) are the same variables but considered for households in a high level of income. After a nonparametric analysis of data, it may seem unreasonable to assume that $m(\mathfrak{z})$ and $m^*(\mathfrak{z})$ are the same function; but it may be possible that a relationship as (3.16) holds and then it will be of interest to estimate θ_0.

Some recent papers have analysed similar models to (3.16) in settings which are different from ours. Härdle and Marron (1990) consider a (possibly non-linear) parametric relationship between the two unknown regression functions when regressors are fixed and taken equally spaced on the unit interval. Pinkse and Robinson (1993) consider the same kind of relationship as Härdle and Marron (1990) when regressors are continuous random variables, and prove that a more efficient estimate is obtained by pooling the two data sets.

The true parameter θ_0 satisfies that $\theta_0 = argmin\ Q(\theta)$, where

$$Q(\theta) \equiv Q(\theta_1,\theta_2) = \sum_{\mathfrak{z}\in\mathcal{F}}(m^*(\mathfrak{z})-\theta_1-\theta_2 m(\mathfrak{z}))^2 v(\mathfrak{z}) \qquad (3.17)$$

($v(.)$ is any positive real "weight function", chosen in such a way that the summation is finite). We can obtain a feasible estimate replacing the unknown regression functions by the non-smoothing estimates $\hat{m}(\mathfrak{z})$ and $\hat{m}^*(\mathfrak{z})$ defined in section 2. Thus, let us define the *least squares estimate*

$$\hat{\theta} \equiv (\hat{\theta}_1,\hat{\theta}_2) = argmin \sum_{\mathfrak{z}\in\mathcal{F}}(\hat{m}^*(\mathfrak{z})-\theta_1-\theta_2\hat{m}(\mathfrak{z}))^2\hat{w}(\mathfrak{z}).$$

where the weight function we consider here is

$$\hat{w}_n(\mathfrak{z}) = I(n^{-1}\textstyle\sum_j I(Z_j=\mathfrak{z})\geq\varepsilon) \times I(n^{-1}\textstyle\sum_j I(Z^*_j=\mathfrak{z})\geq\varepsilon),$$

for a fixed real value $\varepsilon>0$. We assume that ε is taken in such a way that

$$\varepsilon \notin \{p \in (0,1) : \exists\ \mathfrak{z}\in\mathcal{D} \text{ such that } P(Z=\mathfrak{z})=p \text{ or } P(Z^*=\mathfrak{z})=p\} \qquad (3.18)$$

This is a mere technical condition which does not restrict, in practice, the choice of ε. This condition is introduced in order to ensure that $\forall\ \mathfrak{z}\in\mathcal{F}$ $\hat{w}_n(\mathfrak{z})$ converges to $w(\mathfrak{z})$, where we denote

$$w(\mathfrak{z}) = I(P(Z=\mathfrak{z})>\varepsilon) \times I(P(Z^*=\mathfrak{z})>\varepsilon).$$

The value ε must also satisfy condition (3.20) below. We choose this weight function in order to consider only those points in $\mathcal{F}$ for which there are enough observations in our random samples to construct accurate estimates of the conditional expectations $m(\mathfrak{z})$ and $m^*(\mathfrak{z})$.

We assume in our model that

$$E[\zeta^2] < \infty,\ E[\zeta^{*2}] < \infty, \qquad (3.19)$$

$$\left.\begin{array}{l} \exists\ \mathfrak{z}_1,\mathfrak{z}_2 \in\mathcal{F} \text{ such that:} \\ \qquad \text{a) } m(\mathfrak{z}_1)\neq m(\mathfrak{z}_2) \\ \qquad \text{b) } P(Z=\mathfrak{z}_1)>\varepsilon,\ P(Z^*=\mathfrak{z}_1)>\varepsilon \text{ for } i=1,2, \end{array}\right\} \qquad (3.20)$$

$$\text{If } \mathfrak{z}\in\mathcal{F},\ Var(\zeta\,|\,Z=\mathfrak{z}) > 0,\ Var(\zeta^*\,|\,Z^*=\mathfrak{z}) > 0. \qquad (3.21)$$

Assumption (3.19) ensures that we can apply the asymptotic results proved in section 2.3. Assumption (3.20) is an identifiability condition: it ensures that θ_0 is the only solution to (3.17) when $w(\mathfrak{z})$ is used as weight function. Assumption (3.21) avoids degenerate cases which could be treated in a simpler way. Let us define

$$\lambda(\mathfrak{z}) = \Gamma^{*}(\mathfrak{z}) + \theta_{20}^{2}\Gamma(\mathfrak{z}),$$

$$\hat{\lambda}(\mathfrak{z}) = \hat{\Gamma}^{*}(\mathfrak{z}) + \hat{\theta}_{2}^{2}\hat{\Gamma}(\mathfrak{z}),$$

where $\Gamma(\mathfrak{z})$, $\Gamma^{*}(\mathfrak{z})$, $\hat{\Gamma}^{*}(\mathfrak{z})$ and $\hat{\Gamma}(\mathfrak{z})$ are as defined in section 2.3. Then we have the following result,

THEOREM 7.- If (3.15), (3.16), (3.18), (3.19), (3.20), (3.21) *hold and* (ζ_1, Z_1), (ζ_1^{*}, Z_1^{*}), ..., (ζ_n, Z_n), (ζ_n^{*}, Z_n^{*}) *are i.i.d. random vectors then,*

$$n^{1/2}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,\ A^{-1}VA^{-1}),$$

where the matrices A and V are defined by

$$A \equiv \sum_{\mathfrak{z}\in\mathfrak{Z}} \begin{pmatrix} 1 & m(\mathfrak{z}) \\ m(\mathfrak{z}) & m(\mathfrak{z})^2 \end{pmatrix} w(\mathfrak{z}),$$

$$V \equiv \sum_{\mathfrak{z}\in\mathfrak{Z}} \begin{pmatrix} 1 & m(\mathfrak{z}) \\ m(\mathfrak{z}) & m(\mathfrak{z})^2 \end{pmatrix} \lambda(\mathfrak{z})w(\mathfrak{z}).$$

Furthermore, the asymptotic variance-covariance matrix may be consistently estimated by $\hat{A}^{-1}\hat{V}\hat{A}^{-1}$, where $\hat{A}$ and $\hat{V}$ are defined as A, V, replacing $m(\mathfrak{z})$, $w(\mathfrak{z})$ and $\lambda(\mathfrak{z})$ by $\hat{m}(\mathfrak{z})$, $\hat{w}(\mathfrak{z})$ and $\hat{\lambda}(\mathfrak{z})$. ■

According to the definition of $w(\mathfrak{z})$, the summation in A and V extend only over a finite number of terms. Moreover, A is positive definite as a consequence of (3.20) and Cauchy inequality.

As in section 3.1, the non-smoothing estimates used in this theorem may be replaced by regressogram, kernel or k-NN estimates (the same proof applies, changing references to theorem 3 by references to corrresponding corollaries).

On implementing this estimate, the practitioner only has to choose the fixed value ε. If the asymptotic variance-covariance matrix were known, obviously ε should be such that the most efficient estimate were obtained. In practice, the choice of this value must depend on the sample size and variance of Z and Z^*, the objective of this choice being to consider only those points for which we have accurate estimates.

The asymptotic variance-covariance matrix of $\hat{\theta}$ reminds us of the "heteroskedastic" nature of the model. Observe that

$$\mathrm{AsyVar}(n^{1/2}(\hat{m}^*(\mathfrak{z})-\theta_{10}-\theta_{20}\hat{m}(\mathfrak{z})) = \lambda(\mathfrak{z}).$$

As usual, we can obtain a more efficient estimate in a second stage if we use weighted least squares. Specifically, let us define the *generalised least squares estimate* as

$$\tilde{\theta} \equiv (\tilde{\theta}_1,\tilde{\theta}_2) = argmin \sum\nolimits_{\mathfrak{z}\in\mathcal{F}}(\hat{m}^*(\mathfrak{z})-\theta_1-\theta_2\hat{m}(\mathfrak{z}))^2\hat{\lambda}(\mathfrak{z})^{-1}\hat{u}_n(\mathfrak{z}),$$

where the trimming function we consider now is

$$\hat{u}_n(\mathfrak{z}) = I(n^{-1}\sum\nolimits_j I(Z_j=\mathfrak{z})\ge\rho/n^{\alpha}) \times I(n^{-1}\sum\nolimits_j I(Z^*_j=\mathfrak{z})\ge\rho/n^{\alpha}),$$

for fixed positive real values ρ and α. Observe that, unlike $\hat{w}_n(\mathfrak{z})$, the trimming function $\hat{u}_n(\mathfrak{z})$ satisfies that

$$\forall\ \mathfrak{z}\in\mathcal{F},\ \hat{u}_n(\mathfrak{z}) \xrightarrow{\ p\ } 1.$$

Hence, asymptotically all values in $\mathcal{F}$ are taking into account on computing $\tilde{\theta}$ whatever the values ρ and α we choose. We assume that

$$\exists\ \delta > 0 \text{ such that } \forall\ \mathfrak{z}\in\mathcal{F}\ Var(\zeta|Z=\mathfrak{z})>\delta \text{ and } Var(\zeta^*|Z^*=\mathfrak{z})>\delta, \qquad (3.22)$$

THEOREM 8.- If (3.15), (3.16), (3.18), (3.19), (3.20), (3.22) *hold and* (ζ_1,Z_1), (ζ^*_1,Z^*_1), ..., (ζ_n,Z_n), (ζ^*_n,Z^*_n) *are i.i.d. random vectors then,*

$$n^{1/2}(\tilde{\theta}-\theta_0) \xrightarrow{d} N(0,\, \Omega^{-1}),$$

where the matrix Ω *is defined by*

$$\Omega = \sum_{\mathfrak{z}\in\mathcal{Z}} \begin{pmatrix} 1 & m(\mathfrak{z}) \\ m(\mathfrak{z}) & m(\mathfrak{z})^2 \end{pmatrix} \lambda(\mathfrak{z})^{-1}.$$

Furthermore, Ω *may be consistently estimated by* $\hat{\Omega}$, *defined in the same way as* Ω *replacing* $m(\mathfrak{z})$ *and* $\lambda(\mathfrak{z})$ *by* $\hat{m}(\mathfrak{z})$ *and* $\hat{\lambda}(\mathfrak{z})$. ■

If we compare theorems 7 and 8 we observe that there are at least two reasons why $\tilde{\theta}$ is preferable to $\hat{\theta}$: on the one hand, $\tilde{\theta}$ is more efficient than $\hat{\theta}$ (it is easy to prove that $A^{-1}VA^{-1}-\Omega^{-1}$ is positive definite); on the other hand, the asymptotic distribution of $n^{1/2}(\tilde{\theta}-\theta)$ does not depend on the choice of any real number, whereas the asymptotic distribution of $n^{1/2}(\hat{\theta}-\theta)$ may be severely affected by a bad choice of ε. In section 3.4 we analyse the finite-sample behaviour of both estimates in various statistical models.

The linear relationship considered in (3.16) may be too simple to capture the true nature of the observations. More general parametric relationships may be considered. Specifically,

$$m^*(\mathfrak{z}) = S(\theta, m(\mathfrak{z})),$$

(where $S(.,.)$ is a known real function and θ is an unknown vector of parameters) may be a more realistic assumption than (3.16). But, essentially, the same ideas which underlie our proposed estimate may be also used in this case -we prefer the simpler model (3.16) for the sake of clarity. Even more general models can be considered, such as

$$m^*(\mathfrak{z}) = S(\theta_1, m(T(\theta_2, \mathfrak{z}))$$

for known real functions $S(.,.)$, $T(.,.)$ and unknown vector parameters

θ_1, θ_2. But here the function T and the parameter space must be such that $T(\theta_2,\mathfrak{z}) \in \mathcal{D}$. Hence, strong conditions should be imposed on $T(.,.)$, the parameter space and the estimates of θ_2.

3.3. Other semiparametric models

In some semiparametric problems it is not straightforward to achieve root-n-consistency owing to the bias introduced by the nonparametric estimate, as in the models studied in previous sections or in the "average derivative estimation (ADE) method" (see e.g. Powell et al. 1989, Härdle and Stoker 1989 or Robinson 1989). In the ADE model Chamberlain (1986) proved that if all regressors are discrete then the parameter of interest may not be identifiable (even up to a scale coefficient). In the mixed continuous-discrete case, it would be possible to achieve root-n-consistency, but the involved resulting model will probably not capture the true relationship between the variables concerned (see Stoker 1991, section 5.2.a).

In other semiparametric problems, the goal is to improve efficiency rather than achieve root-n-consistency. In most of these models, implementation of discrete regressors using our methods is straightforward. For instance, in the asymptotic efficient estimation in the presence of heteroskedasticity of unknown form, Robinson (1987) proved (using k-NN regression estimates) that the semiparametric estimate is asymptotically efficient even when regressors have discrete or mixed distribution. As a consequence of our results in section 2, when all regressors are discrete the same asymptotic distribution is obtained using non-smoothing, regressogram or kernel weights. Nonparametric k-NN weights have been also used in other semiparametric inference problems in which weights presented in this chapter are also straightforwardly applicable (see e.g. Newey 1990 and Delgado 1992).

3.4. Simulations

We have generated observations from the regression models discussed in sections 3.1 and 3.2 and computed the various semiparametric estimates discussed there. The results are contained in Tables 1, 2, 3, 4 and 5.

First we have generated observations from eight partially linear regression models. In models 1-6 we have taken X and Z to be scalar random variables. In these six models Z was taken from a Poisson distribution with mean λ (specified below) and X was taken as $X = Z + V$, where V was generated from a normal population independent of Z with zero mean and variance 1. In all models the error term U is independent from V and was generated from a normal population with zero mean and variance $\sigma^2(Z)$. The complete description of models 1-6 is as follows:

Model	λ	$\sigma_U^2(Z)$	Underlying model for Y
1	0.3	1	$Y = 1 + X + Z + U$
2	3.0	1	$Y = 1 + X + Z + U$
3	0.3	1	$Y = 1 + X - 3(Z-1)^2 + U$
4	3.0	1	$Y = 1 + X - 3(Z-1)^2 + U$
5	0.3	$(1+Z/3)^2$	$Y = 1 + X - 3(Z-1)^2 + U$
6	3.0	$(1+Z/3)^2$	$Y = 1 + X - 3(Z-1)^2 + U$

Note that models 1 and 2 are linear and homoskedastic, models 3 and 4 are nonlinear and homoskedastic and models 5 and 6 are nonlinear and heteroskedastic. In models with uneven label, the variance of Z is small and in every sample the majority of values will be 0 or 1; however, in models with even label samples will contain many different values of Z.

In models 7 and 8, Z was taken to be a bivariate Poisson distribution, $Z = (Z_1,Z_2)$ (both Z_1 and Z_2 with mean λ), V and U were as in models 1-6 and $X = Z_1 + Z_2 + V$. The complete description of these models is:

Model	λ	$\sigma^2_U(Z)$	Underlying model for Y
7	*0.3*	*1*	$Y = 1 + X - 3(Z-1)^2 - 3(Z-1)^2 + U$
8	*3.0*	*1*	$Y = 1 + X - 3(Z-1)^2 - 3(Z-1)^2 + U$

In all models the semiparametric estimates $\hat{\beta}$ $\tilde{\beta}$ and $\check{\beta}$ (non-smoothing, kernel and uniform k-NN estimates, respectively) were computed. In models 1-6 the kernel we used was the ***Epanechnikov kernel*** (the most efficient one in nonparametric estimation), defined as

$$k(u) = 0.75(1-u^2)I(|u|\leq 1).$$

In models 7-8 the kernel used was the product of two univariate Epanechnikov kernels. On computing both the kernel and the k-NN estimates smoothing values (h_n and k_n respectively) have to be selected. We have simply selected three possible h_n and k_n trying to cover meaningful intervals for them. Observe that, according to our selection of the support $\mathcal{D}$ and the kernel function k, if $h_n<1$ then the kernel estimate is the same as the non-smoothing one.

From the results in section 3.1, the asymptotic distribution of the non-smoothing estimate $\hat{\beta}$ is

Models 1-4, 7-8: $n^{1/2}(\hat{\beta}-1) \xrightarrow{d} N(0,1),$

Model 5: $n^{1/2}(\hat{\beta}-1) \xrightarrow{d} N(0,1.1^2).$

Model 6: $n^{1/2}(\hat{\beta}-1) \xrightarrow{d} N(0,2^2).$

The same asymptotic distributions hold for the kernel and k-NN estimates.

We report the sample mean (M) and mean square error (E) of each estimate. Table 1 contains results corresponding to a sample size of $n=40$ observations; the reported values are based on $r=5000$ replications.

Tables 2 and 3 contain corresponding results for $n=200$, $r=2000$ and $n=1000$, $r=500$, respectively.

In nonparametric estimation, the typical trade-off between bias and variance is closely related with the degree of smoothing. Specifically, bias increases/decreases as the amount of smoothing increases/decreases and variance increases/decreases as the amount of smoothing decreases/increases. This behaviour is observed using any smoother. However, in semiparametric estimation problems this relationship is not so evident. In fact, we find in the simulations reported here that, for fixed sample size, the non-smoothing estimate can perform better than the others in terms of bias and variance, and this fact is stressed when the nonparametric part of the model exhibits high volatility.

In models 1 and 2 (both linear) all estimates have similar behaviour; the k-NN estimates perform slightly better than the others in model 1 and the kernel estimates seem to be the better ones in model 2 (though, as expected, in both cases the non-smoothing estimate is the one with lowest bias). In models 3, 5 and 7 (nonlinear in Z and with low variance for Z) the non-smoothing estimate is the most adequate one, but the other nonparametric estimates also behave properly. In models 4, 6 and 8 (nonlinear in Z and with high variance for Z) the non-smoothing estimate is, again, the better one but, unlike in previous models, kernel and k-NN estimates perform rather poorly. In the heteroskedastic models the variance varies in the expected direction. In the two-dimensional models there is an increase in variance as a result of the poorer performance of the nonparametric estimate.

These results are not a surprise and can be explained in terms of the closeness between $m_{\zeta}(\mathfrak{z}_1)$ and $m_{\zeta}(\mathfrak{z}_2)$ when $\mathfrak{z}_1$ and $\mathfrak{z}_2$ are close values within $\mathcal{D}$. Since the set $\mathcal{D}$ is discrete, the traditional concept of continuous function is useless to assess this relationship of closeness. But observe that,

a) In models 1 and 2, we have $m_Y(0)=1$, $m_Y(1)=3$, $m_Y(2)=5$, $m_Y(3)=7$, $m_Y(4)=9$ and so on. In these models, close values in $\mathcal{D}$ have fairly close conditional expectations and, as a result, if the sample size is small (as in Table 1), smoothing may improve the behaviour of the estimates. For a fixed sample size, the higher the variance of Z the better it will be to smooth: in this case, it will be likely to have points $\mathcal{z}$ for which $Z=\mathcal{z}$ in only a few observations and, then, smoothing will improve the accuracy of the nonparametric estimate. This is what we see when comparing models 1 and 2 in table 1: in the latter, we achieve by smoothing a comparatively more important improvement when we smooth.

b) In models 3, 4, 5 and 6, we have $m_Y(0)=-2$, $m_Y(1)=2$, $m_Y(2)=0$, $m_Y(3)=-8$, $m_Y(4)=-22$ and so on. Thus, close values in $\mathcal{D}$ do not have close conditional expectations. As a result, in no case is smoothing advisable. Even more, the higher the variance of Z, the worse it will be to smooth: if Z has small variance we will have plenty of information for each observed data point and the smoothing will not worsen dramatically the performance of the nonparametric estimate; however, if Z has large variance, then "noisy" information which comes from smoothing will seriously affect the performance of the nonparametric estimate. In tables 1, 2 and 3 we observe that in models 3-8 the non-smoothing estimate is the best one and the other nonparametric estimates only seem adequate in those models in which $Var(Z) = 0.3$.

To sum up, if the unknown part of the partially linear regression model does not exhibit high volatility, then the k-NN and the kernel estimates may perform slightly better than the non-smoothing one if the smoothing values are properly chosen. Otherwise, smoothing techniques are not adequate and may produce extremely misleading results, as in models 4, 6 and 8 -and observe that this may happen even though there exist continuous functions from $\mathbb{R}^q$ to $\mathbb{R}$ $m_Y(.)$ and $m_X(.)$ such that $\forall\ \mathcal{z}\in\mathcal{D}$ $E[Y|Z=\mathcal{z}]=m_Y(\mathcal{z})$ and $E[X|Z=\mathcal{z}]=m_X(\mathcal{z})$.

We have also generated observations from five pairs of regression curves with similar shape and computed the semiparametric estimates described in section 3.2. In all cases Z and Z^* were taken as independent random variables from a Poisson distribution with mean λ

(specified below), V and V^* were taken as independent random variables (also independent from Z and Z^*) from a normal distribution with zero mean and variance 1 and, finally, $\zeta = m(Z) + V$ and $\zeta^* = m^*(Z^*) + V^*$, where $m(Z)$ is specified below and $m^*(.)$ and $m(.)$ satisfy (3.16) for $\theta_0 = (10,2)$. The complete description of all models is as follows:

Model	9	10	11	12	13
λ	1.0	3.0	2.0	0.5	5.0
$m(Z)$	$2+Z$	$2+Z$	$(2-Z)^2$	$3(Z-1)^2$	$log(Z+2)$

A trimming value ε had to be chosen in order to compute $\hat{\theta}$ and, additionally, positive real values ρ and α had to be selected to compute $\tilde{\theta}$. According to theorem 7, the performance of $\hat{\theta}$ depends crucially on the choice of ε; according to theorem 8, the performance of $\tilde{\theta}$ does not depend on the choice of ε, δ and α. In order to analyse how to choose ε, we have first computed in models 9-13 what values in $\mathcal{F}$ should satisfy $w(\mathfrak{z}) = 1$ to achieve as good an estimate of θ as possible[4]. We obtained that these values are: $\{0,1,2,3\}$ in model 9, $\{1,2,3,4,5,6\}$ in model 10, $\{0,1,2,3,4,5,6\}$ in model 11, $\{0,1\}$ in model 12 and $\{1,2,3,4,5,6,7,8,9\}$ in model 13. Thus, we observe that the higher the variance of Z, the greater the number of values in $\mathcal{F}$ which must satisfy $w(\mathfrak{z}) = 1$ -and, hence, the smallest the positive real number ε should be. Therefore, in our simulations we have selected two values of ε which are inversely proportional to the standard deviation of Z. Specifically, we chose $\varepsilon_1 = 0.05 \times Var(Z)^{1/2}$ and $\varepsilon_2 = 0.1 \times Var(Z)^{1/2}$. With this choice, according to theorems 7 and 8, the asymptotic distribution of $n^{1/2}(\hat{\theta}-\theta)$ is $N(0,\Sigma_1)$ for ε_1 and $N(0,\Sigma_2)$ for ε_2, and the asymptotic distribution of $n^{1/2}(\tilde{\theta}-\theta)$ is $N(0,\Sigma_3)$, where the symmetric matrices Σ_1, Σ_2 and Σ_3 are specified below for each model:

[4] We say that the estimate $\hat{\theta}$ obtained with a value ε is the best one if for every other positive real number η, the determinant of the asymptotic variance-covariance matrix $A^{-1}VA^{-1}$ (see th. 7) obtained for η is greater than or equal to the determinant of the matrix $A^{-1}VA^{-1}$ obtained for ε.

M.	9	10	11	12	13
Σ_1	$\begin{pmatrix} 80.3 & -26.0 \\ & 8.97 \end{pmatrix}$	$\begin{pmatrix} 75.0 & -13.7 \\ & 2.78 \end{pmatrix}$	$\begin{pmatrix} 11.5 & -3.25 \\ & 1.72 \end{pmatrix}$	$\begin{pmatrix} 16.5 & -5.50 \\ & 3.89 \end{pmatrix}$	$\begin{pmatrix} 260 & -133.4 \\ & 70.6 \end{pmatrix}$
Σ_2	$\begin{pmatrix} 84.2 & -28.3 \\ & 10.2 \end{pmatrix}$	$\begin{pmatrix} 94.5 & -18.4 \\ & 3.84 \end{pmatrix}$	$\begin{pmatrix} 10.5 & -3.54 \\ & 2.50 \end{pmatrix}$	$\begin{pmatrix} 16.5 & -5.50 \\ & 2.75 \end{pmatrix}$	$\begin{pmatrix} 320 & -166.5 \\ & 88.5 \end{pmatrix}$
Σ_3	$\begin{pmatrix} 50 & -15 \\ & 5 \end{pmatrix}$	$\begin{pmatrix} 46.6 & -8.33 \\ & 1.67 \end{pmatrix}$	$\begin{pmatrix} 7 & -1 \\ & 0.5 \end{pmatrix}$	$\begin{pmatrix} 10.6 & -2.50 \\ & 1.11 \end{pmatrix}$	$\begin{pmatrix} 161 & -82.6 \\ & 43.7 \end{pmatrix}$

We report in table 4 the mean (M) and variance (V) of $\hat{\theta}$ and $\tilde{\theta}$ computed using non-smoothing weights for the nonparametric estimates and $\alpha=0.01$, $\rho=0.1$. In table 5 we report corresponding results when the nonparametric estimates are computed using kernel weights (with Epanechnikov kernel) and $h=1.2$. All reported values are based on $n=40$ observations and $r=10000$ replications.

We observe that in models 9, 10, 11 and 12 the non-smoothing estimate performs better than the kernel one, whereas in model 13 the kernel estimate seems to be the most adequate one. Again, these results are not a surprise: in model 13 the regression function has low variability (i.e. close points in $\mathcal{D}$ have close conditional expectations) and as $Var(Z)=5$ in every sample there are many different values -therefore, smoothing improves the accuracy of estimates.

If we compare $\hat{\theta}$ and $\tilde{\theta}$, we observe that, surprisingly, in some cases the former performs better than the latter (models 10 and 11 when ε_1 is used). This also happens with some other well-known two-stage estimators, and the reason for this fact is because the weights $\lambda(\mathfrak{z})$ are so poorly estimated in the first stage that no improvement is achieved in the second stage. However, in this specific model, the generalised least squares estimate still has an advantage over the ordinary least squares estimate: results do not depend on the choice of ε when using $\tilde{\theta}$, unlike what happens with $\hat{\theta}$ (see, for instance, models 9 and 12). This is the main reason why $\tilde{\theta}$ seems preferable to $\hat{\theta}$ when estimating the parameters relating two regression curves with similar shape.

We observe that in models 9, 10, 11 and 12 the non-smoothing estimate performs better than the kernel one, whereas in model 13 the kernel estimate seems to be the most adequate one. Again, these results are not a surprise: in model 13 the regression function has low variability (i.e. close points in $\mathfrak{D}$ have close conditional expectations) and as $Var(Z)=5$ in every sample there are many different values -therefore, smoothing improves the accuracy of estimates.

If we compare $\hat{\theta}$ and $\tilde{\theta}$, we observe that, surprisingly, in some cases the former performs better than the latter (models 10 and 11 when ε_1 is used). This also happens with some other well-known two-stage estimators, and the reason for this fact is because the weights $\lambda(\mathfrak{z})$ are so poorly estimated in the first stage that no improvement is achieved in the second stage. However, in this specific model, the generalised least squares estimate still has an advantage over the ordinary least squares estimate: results do not depend on the choice of ε when using $\tilde{\theta}$, unlike what happens with $\hat{\theta}$ (see, for instance, models 9 and 12). This is the main reason why $\tilde{\theta}$ seems preferable to $\hat{\theta}$ when estimating the parameters relating two regression curves with similar shape.

APPENDIX.- Proofs.

Proof of theorem 1: We must prove that the sequence of non-smoothing weights 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 $f:\mathbb{R}^q \longrightarrow \mathbb{R}$,

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

PROOF: $E[\sum_j W_{nj}(Z)f(Z_j)] \leq \quad 2E[\sum_j f(Z_j)I(Z_j=Z)/(1+\sum_k I(Z_k=Z))]$

$$= 2nE[f(Z_1)I(Z_1=Z)/(1+\sum_k I(Z_k=Z))]$$

$$= 2E\{f(Z_1)I(Z=Z_1)E[n/(2+\sum_{k=2}^{n} I(Z_k=Z))|Z,Z_1]\}$$

Given $\mathfrak{z} \in \mathcal{D}$, if we define $B_n^* \equiv \sum_{k=2}^{n} I(Z_k=\mathfrak{z})$, and $p_{\mathfrak{z}} \equiv P(Z=\mathfrak{z})$, then

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

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

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

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

$$E[\sum_j W_{nj}(Z)f(Z_j)] \le 2E\{f(Z_1)E[I(Z=Z_1)P(Z)^{-1}|Z_1]\} = 2E[f(Z_1)] \qquad ■$$

Lemma 1.- Let Z be a discrete random variable with support $\mathcal{D}$ and probability function $P(Z=\mathfrak{z}) = p_{\mathfrak{z}}\ \forall\ \mathfrak{z}\in\mathcal{D}$; let Z, Z_1, ..., Z_n be i.i.d. random variables and $m\in\mathbb{Z}$, $m\ge0$ (m fixed). Then

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

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

$$\begin{aligned} nP(\sum_k I(Z_k=Z)=m) &= n\sum_{\mathfrak{z}\in\mathcal{D}} p_{\mathfrak{z}} \binom{n}{m} p_{\mathfrak{z}}^{m}(1-p_{\mathfrak{z}})^{n-m} \\ &= n\sum_{\mathfrak{z}\in\mathcal{D}} p_{\mathfrak{z}}^{m+1} \binom{n}{m} (\sum_{s=0}^{n-m}(-1)^{s}\binom{n-m}{s} p_{\mathfrak{z}}^{s}) \end{aligned}$$

$$= n\sum_{s=0}^{n-m}\binom{n}{m}\binom{n-m}{s}(-1)^s\Big(\sum_{\mathfrak{z}\in\mathcal{D}} p_{\mathfrak{z}}^{s+m+1}\Big).$$

Define $p_0 = \sup_{\mathfrak{z}\in\mathcal{D}} p_{\mathfrak{z}} < 1$ and $q \in (0,1-p_0)$. (If $p_0 = 1$, Z is degenerate and Lemma 1 is straightforward). Then,

$$\forall\ k\geq 1 \text{ and } \forall\ \mathfrak{z}\in\mathcal{D},\ (p_{\mathfrak{z}}/(1-q))^k \leq p_{\mathfrak{z}}/(1-q) < p_{\mathfrak{z}}/p_0,$$

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

$$\Rightarrow \quad \sum_{\mathfrak{z}\in\mathcal{D}} p_{\mathfrak{z}}^k < (1-q)^k/p_0 < (1-q)^{k-1}/p_0.$$

Therefore, the previous equality implies that

$$\begin{aligned} nP(\textstyle\sum_k I(Z_k=Z)=m) &\leq np_0^{-1}\binom{n}{m}\sum_{s=0}^{n-m}\binom{n-m}{s}(-1)^s(1-q)^{s+m} \\ &= p_0^{-1}\binom{n}{m}n(1-q)^m q^{n-m} = o(1). \end{aligned}$$ ■

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

PROOF: $\sum_k W_{nk}(Z) = I(\sum_k I(Z_k=Z)\neq 0)$. So, 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.- $\max_j W_{nj}(Z) \xrightarrow{P} 0.$

PROOF: $\forall\varepsilon>0$, $P(|\max_j W_{nj}(Z)|>\varepsilon) = P(\sum_k I(Z_k=Z)\neq 0,\ 1/\sum_k I(Z_k=Z)>\varepsilon) = P(0<\sum_k I(Z_k=Z)<1/\varepsilon)$. Define $\mathcal{J}(\varepsilon) = \mathbb{N} \cap (0,1/\varepsilon)$, which is a finite subset of $\mathbb{N}$. Then, $P(0<\sum_k I(Z_k=Z)<1/\varepsilon) = \sum_{m\in\mathcal{J}(\varepsilon)} P(\sum_k I(Z_k=Z)=m) = o(1)$, since the sum contains a finite number of terms, all converging to 0 (Lemma 1). ■

Proof of Corollary 1: We only prove the second statement here (the first one follows in a similar way). By (2.4) we know that $\exists\ M : \|x\|\geq M \Rightarrow \psi(x) = 0$, and there exists n_0 such that

$$n\geq n_0 \Rightarrow \mu/h_n \geq M \text{ and } \|\mathfrak{z}-Z_j\|/h_n \geq MI(\mathfrak{z}\neq Z_j) \text{ if } \mathfrak{z}\in\mathcal{D}.$$

Hence if $n\geq n_0$ and $\mathfrak{z}\in\mathcal{D}$, then $\tilde{W}_{nj}(\mathfrak{z}) = W_{nj}(\mathfrak{z})$ and $\tilde{m}_\zeta(\mathfrak{z}) = \hat{m}_\zeta(\mathfrak{z})$; so this corollary follows from theorem 1. ■

Proof of Theorem 2: Observe that if $\sum_j I(Z_j=Z)\geq k$, then

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

Therefore, $$P(\hat{m}_\zeta(Z)\neq \check{m}_\zeta(Z)) \leq P(\textstyle\sum_j I(Z_j=Z)<k)$$

$$=\sum_{m=0}^{k-1} P(\textstyle\sum_j I(Z_j=Z)=m)$$

$$\leq p_0^{-1}\sum_{m=0}^{k-1}\binom{n}{m}(1-q)^m q^{n-m},$$

where $k\equiv k_n$, p_0^{-1} is as in Lemma 1, and the last equality holds for $q\in(0,1)$ (as in Lemma 1). By (2.5) there exists n_0 such that $n\geq n_0 \Rightarrow k\leq n/2$. So, if $n\geq n_0$,

$$P(\hat{m}(Z)\neq\check{m}(Z)) \leq p_0^{-1}\sum_{m=0}^{k-1}\binom{n}{m}(1-q)^m q^{n-m} \leq p_0^{-1}k\binom{n}{k}q^{n-k},$$

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,

$$q_0^{-n}P(m_\zeta^{(1)}(Z)\neq m_\zeta^{(4)}(Z)) \sim q_0^{n-2k} \times (2\pi)^{-1/2}(q_0 n/(n-k))^{n-k} \times q_0^{n-k}(n/k)^k k^{1/2}$$

and all terms in this product converge to 0 by (2.5) (the third term is equal to $exp\{n \times [(n-k)log(q_0)/n + (k/n)log(k/n) + (1/2n)log(k)]\})$. ■

Proof of Theorem 3: Given $\mathfrak{z}\in\mathcal{D}$, let us define $U_j(\mathfrak{z}) = (\zeta_j-m_\zeta(\mathfrak{z}))I(Z_j=\mathfrak{z})$. Then,

$$n^{1/2}\begin{pmatrix} \hat{m}_\zeta(\mathfrak{z}_1)-m_\zeta(\mathfrak{z}_1) \\ \cdots\cdots\cdots\cdots\cdots \\ \hat{m}_\zeta(\mathfrak{z}_f)-m_\zeta(\mathfrak{z}_f)\end{pmatrix} = (P_n \otimes I_s)^{-1}n^{-1/2}\sum_j\begin{pmatrix} U_j(\mathfrak{z}_1) \\ \cdots\cdots \\ U_j(\mathfrak{z}_f)\end{pmatrix},$$

where $s=dim(\zeta)$, I_s is the identity matrix of order s and P_n is the $f\times f$ diagonal matrix $P_n \equiv diag[n^{-1}\sum_j I(Z_j=\mathfrak{z}_1), \ldots, n^{-1}\sum_j I(Z_j=\mathfrak{z}_f)]$. Now, by Khinchine's Law of Large Numbers,

$$(P_n \otimes I_s)^{-1} \xrightarrow{p} diag[p(\mathfrak{z}_1), \ldots, p(\mathfrak{z}_f)] \otimes I_s;$$

and by Lindenberg-Levy's Central Limit Theorem

$$n^{-1/2}\sum_j \begin{pmatrix} U_j(\mathfrak{z}_1) \\ \cdots\cdots \\ U_j(\mathfrak{z}_f) \end{pmatrix} \xrightarrow{d} \mathrm{N}(0, \begin{bmatrix} p(\mathfrak{z}_1)\Sigma(\mathfrak{z}_1) & \ldots & 0 \\ 0 & \ldots & p(\mathfrak{z}_f)\Sigma(\mathfrak{z}_f) \end{bmatrix}.$$

Combining both results we obtain theorem 3. ∎

Proof of Corollary 2: Follows from theorem 3 in the same way as corollary 1 and theorem 2 were proved. ∎

Proof of Theorem 4: From equation (3.8), it suffices to prove that

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

$$\hat{\Phi} \xrightarrow{p} \Phi,\ \hat{\sigma}^2 \xrightarrow{p} \sigma^2. \qquad \text{(A.2)}$$

Propositions 4.1-4.4 below prove (A.1); (A.2) is easily proved with similar arguments.

<u>Proposition 4.1.</u>- $E\|n^{-1/2}\sum_i(m_{Xi}-\hat{m}_{Xi})\hat{m}_{Ui}I_i\|^2 = o(1)$.

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

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

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

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

For the first term, applying Cauchy-Schwartz inequality,

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

$E\|m_{X1}-\hat{m}_{X1}\|^4$ converges to 0 (applying (3.6)); $\hat{m}_{U1}$ 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 (3.6).

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

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

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

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

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

All terms in this last expression are 0 because if we denote

$$W^*_{jl}(Z_1,Z_2,\dots,Z_n) \equiv W_{nj}(Z_1)W_{nl}(Z_2)W_{n3}(Z_1)W_{n3}(Z_2),$$

Then, we have

$$W^*_{jl}(Z_1,Z_2,\dots,Z_n) = I(Z_1=Z_2=Z_3=Z_j=Z_l)/(\textstyle\sum_{k=2}^{n}I(Z_k=Z_1))^4 \Rightarrow$$

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

$$E[I_1W^*_{jl}(Z_1,Z_2,\dots,Z_n)I_2(m_{X1}-m_{Xj})'(m_{X2}-m_{Xl})] = 0.$$

The last equality holds because the variable whose expectation is taken is 0; note that, if $W^*_{jl}(Z_1,Z_2,\dots,Z_n) \neq 0$, then $Z_1=Z_j$ and $Z_2=Z_l$, hence $(m_{X1}-m_{Xj})'(m_{X2}-m_{Xl}) \equiv (E[X|Z_1]-E[X|Z_j])'(E[X|Z_2]-E[X|Z_l]) = 0)$. ∎

Proposition 4.2 $\quad E\|n^{-1/2}\sum_i(X_i-m_{Xi})\hat{m}_{Ui}I_i\|^2 = o(1).$

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

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

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

The first term converges to 0 as in proposition 4.1. As for the second one,

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

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

Proposition 4.3.- $E\|n^{-1/2}\sum_i(m_{Xi}-\hat{m}_{Xi})U_iI_i\|^2 = o(1).$

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 Cauchy-Schwartz inequality as in previous propositions) and the second one is 0 (because U_1 and U_2, conditional on $\mathfrak{Z}$, are independent random variables whose expectation is exactly equal to 0). ∎

Proposition 4.4.- $n^{-1/2}\sum_i(X_i-m_{Xi})U_iI_i \xrightarrow{d} N(0,\sigma^2\Phi)$.

PROOF: By Central Limit Theorem it follows that,

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

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

On the other hand,

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

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

$$= E[\|(X_i-m_{Xi})U_i\|^2(1-I_i)] = \sigma^2E[\|X_i-m_{Xi}\|^2(1-I_i)].$$

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 Corollary 3: Follows from theorem 2 in the same way as corollary 2 was deduced from theorem 1. ∎

Proof of Corollary 4: When k-NN weights are used, equation (3.8) no longer holds. Instead,

$$n^{1/2}(\check{\beta}-\beta) = \check{\Phi}^{-1}(n^{-1/2}\sum_i\check{\varepsilon}_{Xi}\check{\varepsilon}_{Ui} + n^{-1/2}\sum_i\check{\varepsilon}_{Xi}\check{\varepsilon}_{\theta i}).$$

The second term converges to 0 as proposition 4.5 below proves. As for the first term, the proof of theorem 4 applies except for proposition 4.1 (in all propositions, references to theorem 1 must be replaced by references to Stone's corollary 3 (Stone 1977), where it is proved that uniform k-NN weights are universally consistent). Proposition 4.1 must be replaced by proposition 4.6 below.

Proposition 4.5 .- $E\|n^{-1/2}\sum_i\check{\varepsilon}_{Xi}\check{\varepsilon}_{\theta i}\| = o(1)$.

PROOF: $E\|n^{-1/2}\sum_i\check{\varepsilon}_{Xi}\check{\varepsilon}_{\theta i}\| \le n^{1/2}E[\|\check{\varepsilon}_{X1}\check{\varepsilon}_{\theta 1}\|] \le$

$$E[\|X_1-m_{X1}\|^2]^{1/2}E[n\|\check{\varepsilon}_{\theta 1}\|^2]^{1/2} + E[\|m_{X1}-\check{m}_{X1}\|^2]^{1/2}E[n\|\check{\varepsilon}_{\theta 1}\|^2]^{1/2}$$

Thus, it suffices to prove that $E[n\|\check{\varepsilon}_{\theta 1}\|^2] = o(1)$. Now, $\check{\varepsilon}_{\theta 1} = \theta(Z_1)-\check{m}_{\theta(Z1)}$. Let A be the event $\{\check{m}_{\theta(Z1)} = \hat{m}_{\theta(Z1)},\ I_1 = 1\}$. If A is true, then, according to (3.7), $\check{\varepsilon}_{\theta 1} = 0$. Therefore, if $P(A) = 1$, this prop. is already proved. Otherwise, $P(A^c) = \delta > 0$, and

$$E[n\|\check{\varepsilon}_{\theta 1}\|^2] = nP(A^c)E[\|\theta(Z_1)-\check{m}_{\theta(Z1)}\|^2|A^c];$$

now, $nP(A^c)$ converges to 0 (theorem 1 and lemma 1) and the second term is bounded because $E[\theta(Z_1)^2] < \infty$. ■

Proposition 4.6.- $E\|n^{-1/2}\sum_i(m_{Xi}-\check{m}_{Xi})\check{m}_{Ui}\|^2 = o(1)$.

PROOF: As in proposition 4.1, it suffices to prove that

$$(n-1)E[\check{m}_{U1}(m_{X1}-\check{m}_{X1})'(m_{X2}-\check{m}_{X2})\check{m}_{U2}] \longrightarrow 0.$$

Let A_i be the event $\{\check{m}_{\theta(Zi)} = \hat{m}_{\theta(Zi)}\}$, for $i=1,2$. If $A_1\cap A_2$ is true, then

$$E[\check{m}_{U1}(m_{X1}-\check{m}_{X1})'(m_{X2}-\check{m}_{X2})\check{m}_{U2}] = 0$$

as in proposition 4.1. Hence if $P(A_1\cap A_2) = 1$, the prop. is already proved. Otherwise,

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

$$P((A_1\cap A_2)^c)E[\check{m}_{U1}(m_{X1}-\check{m}_{X1})'(m_{X2}-\check{m}_{X2})\check{m}_{U2}|(A_1\cap A_2)^c];$$

now, $(n-1)P((A_1\cap A_2)^c) \longrightarrow 0$ and the second factor is bounded. ■

Proof of Theorem 5: As in theorem 4, it suffices to prove that

$$n^{-1/2}\sum_i\hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Ui}I_i \xrightarrow{d} N(0,\Psi), \qquad (A.3)$$

$$\hat{\Psi} \xrightarrow{p} \Psi. \qquad (A.4)$$

(A.4) follows in a similar way to (A.2); (A.3) follows in the same way as (A.1) (e.g.: in proposition 4.1, we have

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

$$(n-2)\sum_{j,\, j\neq 1}\sum_{l,\, l\neq 2}E[I_1\sigma^2(X_3,Z_3)(m_{X1}-X_j)'(m_{X2}-X_l)W^*_{jl}(Z_1,Z_2,\ldots,Z_n)I_2],$$

and all terms in this double summation are *0* because

$$E\{I_1W^*_{jl}(Z_1,Z_2,\ldots,Z_n)I_2E[\sigma^2(X_3,Z_3)(m_{X1}-X_j)'(m_{X2}-X_l)\,|\,Z_1,\ldots,Z_n]\} =$$

$$E[I_1W^*_{jl}(Z_1,Z_2,\ldots,Z_n)I_2\sigma^2(m_{X3},Z_3)(m_{X1}-m_{Xj})'(m_{X2}-m_{Xl})] = 0).$$

Oddly enough, the moment condition required in both the homoskedastic model and the heteroskedastic one is the same. In the homoskedastic model, second order moments are required to prove (A.2) and fourth order moments are required to prove (A.1); in the heteroskedastic model fourth order moments are required to prove both (A.3) and (A.4). ■

Proof of Theorem 6: The following lemmas will be used in the proof. They are versions of Robinson's (1988) lemmas adapted to the mixed case. Throughout this proof, Robinson will mean Robinson (1988).

In the following lemmas, Z is a random variable which satisfies (3.11), $Z^{(2)}(d)$ denotes the conditional random variable $Z^{(2)}|Z^{(1)}=d$, f_d is the conditional probability density function of $Z^{(2)}(d)$, k is a function from $\mathbb{R}$ to $\mathbb{R}$ such that $\int|uk(u)|du < \infty$, K is a function from $\mathbb{R}^q$ to $\mathbb{R}$ defined by $K(u_1,\ldots,u_q) = k(u_1)\cdots k(u_q)$ and a_n is a sequence of positive real numbers. All notation here refers to that introduced in section 3.1 after (3.11).

Lemma 2.- If there exist real numbers M, M' such that $f_d(u) < M$ ($\forall u$, $\forall d\in\mathcal{D}$) and $|k(u)| < M'$ ($\forall$ $u\in\mathbb{R}$) then,

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

PROOF: $$h(d,u) = P(Z^{(1)}=d) \times E[\,|K((Z^{(2)}-u)/a_n)|\,|Z^{(1)}=d]$$

$$= P(Z^{(1)}=d) \times \int|K((v-c)/a_N)|f_d(v)dv$$

$$\leq P(Z^{(1)}=d) \times M\left(\int|k(u)|du\right)^q a_N^q \leq Ca_N^q, \text{ where } C= M\left(\int|k(u)|du\right)^q<\infty. \blacksquare$$

Lemma 3.- If there exist real numbers M, M' such that $f_d(u)< M$ ($\forall u$, $\forall d\in\mathcal{D}$) and $|k(u)| < M'$ ($\forall$ $u\in\mathbb{R}$) and $g(d,u)$ is a function from $\mathbb{R}^r$ to $\mathbb{R}$ such that $E[|g(Z)|] < \infty$, then,

$$E[|g(Z_1)K_{12}(a_n)|I(Z_2^{(1)}=Z_1^{(1)})] = O(a_N^q).$$

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

$$E[|g(Z_1)K_{12}(a_n)|I(Z_2^{(1)}=Z_1^{(1)})] =$$

$$E\{|g(Z_1)|E[|K_{12}(a_n)|I(Z_2^{(1)}=Z_1^{(1)})|Z_1]\} =$$

$$E[|g(Z_1)|h(Z)] \le Ca_n^q E[|g(Z_1)|] = C'a_n^q$$

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

Lemma 4.- If $f_d\in\mathcal{G}_\lambda^\infty$ and $k\in\mathcal{K}_l$ ($l-1 < \lambda \le l$) then,

$$E\{(a_n^{-q}E[K_{12}(a_n)|Z_1^{(2)}]-f_d(Z_1^{(2)}))^2|Z_1^{(1)}=d\} = O(a^{2\lambda})$$

PROOF: Similar to Robinson's lemma 4. ■

Lemma 5.- Let $g(d,u)$ be as in lemma 3 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 $l-1<\lambda\le l$, $m-1<\mu\le m$ and $\eta=\min(\mu,\lambda+1)$), then

$$E|E[(g(Z_1)-g(Z_2)K_{12}(a_n)I(Z_1^{(1)}=Z_2^{(1)})|Z_1|^\alpha=O(a_n^{\alpha(q+\eta)})$$

PROOF: Similar to lemma 3's proof and applying Robinson's lemma 5 to $E\{|E[(g_d(Z_1^{(2)}(d))-g_d(Z_2^{(2)}(d)))K((Z_1^{(2)}(d)-Z_2^{(2)}(d))/a_n)]|^\alpha\}$. ■

We can now prove theorem 6. It will suffice to prove that

$$n^{-1/2}\sum_i(X_i-\hat{m}_{Xi})(U_i-\hat{m}_{Ui})I_i \xrightarrow{d} N(0,\sigma^2\Phi), \tag{A.5}$$

$$n^{-1}\sum_i(X_i-\hat{m}_{Xi})(X_i-\hat{m}_{Xi})'I_i \xrightarrow{P} \Phi, \tag{A.6}$$

$$n^{-1/2}\sum_i(X_i-\hat{m}_{Xi})(\theta_i-\hat{m}_{\theta i})I_i \xrightarrow{P} 0, \tag{A.7}$$

$$\hat{\sigma}^2 \xrightarrow{P} \sigma^2. \tag{A.8}$$

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 Cauchy-Schwartz inequality may be used instead. Lemmas in Robinson's appendix B do not apply any more; instead, the lemmas specified above must be used. ■

Proof of Theorem 7: By solving the optimization problem, we obtain

$$\begin{pmatrix}\hat\theta_1\\ \hat\theta_2\end{pmatrix} = \left\{\sum_{\mathfrak{z}\in\mathcal{F}}\hat{w}_n(\mathfrak{z})\begin{pmatrix}1 & \hat{m}(\mathfrak{z})\\ \hat{m}(\mathfrak{z}) & \hat{m}(\mathfrak{z})^2\end{pmatrix}\right\}^{-1}\sum_{\mathfrak{z}\in\mathcal{F}}\hat{w}_n(\mathfrak{z})\begin{pmatrix}\hat{m}^*(\mathfrak{z})\\ \hat{m}(\mathfrak{z})\hat{m}^*(\mathfrak{z})\end{pmatrix}$$

Let us now consider the unfeasible estimate

$$\begin{pmatrix}\bar\theta_1\\ \bar\theta_2\end{pmatrix} = \left\{\sum_{\mathfrak{z}\in\mathcal{F}}w(\mathfrak{z})\begin{pmatrix}1 & \hat{m}(\mathfrak{z})\\ \hat{m}(\mathfrak{z}) & \hat{m}(\mathfrak{z})^2\end{pmatrix}\right\}^{-1}\sum_{\mathfrak{z}\in\mathcal{F}}w(\mathfrak{z})\begin{pmatrix}\hat{m}^*(\mathfrak{z})\\ \hat{m}(\mathfrak{z})\hat{m}^*(\mathfrak{z})\end{pmatrix}$$

(It is unfeasible because $w(\mathfrak{z})$ is unknown). First we prove

Lemma 6: $P(\hat\theta \neq \bar\theta) = o(1)$.

PROOF: Let us define the following subsets of $\mathcal{F}$:

$$\mathcal{F}_1 = \{\mathfrak{z}\in\mathcal{F} : P(Z=\mathfrak{z})<\varepsilon\},$$

$$\mathcal{F}_2 = \{\mathfrak{z}\in\mathcal{F} : P(Z=\mathfrak{z})>\varepsilon \text{ and } P(Z^*=\mathfrak{z})<\varepsilon\},$$

$$\mathcal{F}_3 = \{\mathfrak{z}\in\mathcal{F} : P(Z=\mathfrak{z})>\varepsilon \text{ and } P(Z^*=\mathfrak{z})>\varepsilon\};$$

so, $\mathcal{F} = \mathcal{F}_1\cup\mathcal{F}_2\cup\mathcal{F}_3$ and, hence,

$$P(\hat\theta\neq\bar\theta) \leq \sum_{\mathfrak{z}\in\mathcal{F}}P(w(\mathfrak{z})\neq\hat{w}_n(\mathfrak{z})) = S_1 + S_2 + S_3,$$

where $S_i \equiv \sum_{\mathfrak{z}\in\mathcal{F}_i}P(w(\mathfrak{z})\neq\hat{w}_n(\mathfrak{z}))$. We prove that S_3 converges to 0 (the proof for S_1 and S_2 is similar). Let us define

$$\Xi \equiv \{p\in\mathbb{R} : \exists\ \mathfrak{z}\in\mathbb{R} \text{ such that } P(Z=\mathfrak{z})=p \text{ or } P(Z^*=\mathfrak{z})=p\}.$$

This set is closed in $\mathbb{R}$ (note that $0\in\Xi$); as $\varepsilon\notin\Xi \Rightarrow \exists\ \delta>0$ such that $(\varepsilon-\delta,\varepsilon+\delta)\cap\Xi = \emptyset$. Then, if $\mathfrak{z}\in\mathcal{F}_3$, applying Chebychev inequality we have,

$$\begin{aligned}P(w(\mathfrak{z})\neq\hat{w}_n(\mathfrak{z})) &\leq P(n^{-1}\textstyle\sum_j I(Z_j=\mathfrak{z})<\varepsilon) + P(n^{-1}\sum_j I(Z^*_j=\mathfrak{z})<\varepsilon)\\ &\leq P(|n^{-1}\textstyle\sum_j I(Z_j=\mathfrak{z})-P(Z=\mathfrak{z})|>\delta) + P(|n^{-1}\sum_j I(Z_j=\mathfrak{z})-P(Z^*=\mathfrak{z})|>\delta)\end{aligned}$$

$$\le n^{-1}\delta^{-1}P(Z=z)(1-P(Z=z)) + n^{-1}\delta^{-1}P(Z^*=z)(1-P(Z^*=z)).$$

Hence, $S_3 \le 2/n\delta = o(1)$. ■

Now we prove theorem 7: let us denote $\hat{v}(z) \blacksquare \hat{m}^*(z)-\theta_{10}-\theta_{20}\hat{m}(z)$. Then

$$n^{1/2}\begin{pmatrix} \bar\theta_1-\theta_{10} \\ \bar\theta_2-\theta_{20} \end{pmatrix} = \left\{\sum_{z\in\mathcal{Z}} w(z)\begin{bmatrix} 1 & \hat{m}(z) \\ \hat{m}(z) & \hat{m}(z)^2 \end{bmatrix}\right\}^{-1} \times n^{1/2}\sum_{z\in\mathcal{Z}} w(z)\begin{bmatrix} 1 \\ \hat{m}(z) \end{bmatrix}\hat{v}(z),$$

where now both summations run only through a finite number of terms which does not depend on n. Now, by theorem 3

$$\sum_{z\in\mathcal{Z}} w(z)\begin{bmatrix} 1 & \hat{m}(z) \\ \hat{m}(z) & \hat{m}(z)^2 \end{bmatrix} - A = o_P(1).$$

On the other hand, if $f = \#\,\{z\in\mathcal{Z} : w(z) = 1\}$ and we denote $z_1, \ldots, z_f$ the points in $\mathcal{Z}$ satisfying that $w(z) = 1$, then

$$n^{1/2}\sum_{z\in\mathcal{Z}} w(z)\begin{bmatrix} 1 \\ \hat{m}(z) \end{bmatrix}\hat{v}(z) = \begin{bmatrix} 1 & \ldots & 1 \\ \hat{m}(z_1) & \ldots & \hat{m}(z_f) \end{bmatrix} \times n^{1/2}\begin{pmatrix} \hat{v}(z_1) \\ \vdots \\ \hat{v}(z_f) \end{pmatrix}.$$

Now, as $\hat{v}(z) = (\hat{m}^*(z)-m^*(z)) + \theta_{20}(\hat{m}(z)-m(z))$, and the random samples in which each nonparametric estimate is based are independent, by theorem 3

$$n^{1/2}\begin{pmatrix} \hat{v}(z_1) \\ \vdots \\ \hat{v}(z_f) \end{pmatrix} \xrightarrow{d} \mathrm{N}\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, diag[\lambda(z_1), \ldots, \lambda(z_f)]\right],$$

and hence theorem 7 follows for the unfeasible estimate $\bar\theta$ and, applying lemma 6, also for the feasible estimate $\hat\theta$. ■

Proof of Theorem 8.- By solving the optimization problem we obtain

$$n^{1/2}\begin{pmatrix} \tilde\theta_1-\theta_{10} \\ \tilde\theta_2-\theta_{20} \end{pmatrix} = \left\{\sum_{z\in\mathcal{Z}}\begin{bmatrix} 1 & \hat{m}(z) \\ \hat{m}(z) & \hat{m}(z)^2 \end{bmatrix}\hat\lambda(z)^{-1}\hat{u}_n(z)\right\}^{-1} \times$$

$$\times\ n^{1/2}\sum_{\mathfrak{z}\in\mathfrak{Z}}\begin{pmatrix}1\\ \hat{m}(\mathfrak{z})\end{pmatrix}\hat{v}(\mathfrak{z})\hat{\lambda}(\mathfrak{z})^{-1}\hat{u}_n(\mathfrak{z})$$

where $\hat{v}(\mathfrak{z})$ is as in Theorem 7. Thus it suffices to prove that

$$\sum_{\mathfrak{z}\in\mathfrak{Z}}(1,\ \hat{m}(\mathfrak{z}))'(1,\ \hat{m}(\mathfrak{z}))\hat{\lambda}(\mathfrak{z})^{-1}\hat{u}_n(\mathfrak{z}) \xrightarrow{p} \Omega, \qquad \text{(A.9)}$$

$$n^{1/2}\sum_{\mathfrak{z}\in\mathfrak{Z}}(1,\ \hat{m}(\mathfrak{z}))'\hat{v}(\mathfrak{z})\hat{\lambda}(\mathfrak{z})^{-1}\hat{u}_n(\mathfrak{z}) \xrightarrow{d} N(0,\Omega). \qquad \text{(A.10)}$$

We prove (A.10); (A.9) follows in a similar way.

$n^{1/2}\sum_{\mathfrak{z}\in\mathfrak{Z}}(1,\ \hat{m}(\mathfrak{z}))'\hat{v}(\mathfrak{z})\hat{\lambda}(\mathfrak{z})^{-1}\hat{u}_n(\mathfrak{z}) = T_1 + T_2 + T_3 + T_4 + T_5$, where

$$T_1 = n^{1/2}\sum_{\mathfrak{z}\in\mathfrak{Z}}(1,\ \hat{m}_\zeta\mathfrak{z}))'\hat{v}(\mathfrak{z})(\hat{\lambda}(\mathfrak{z})^{-1}-\lambda(\mathfrak{z})^{-1})\hat{u}_n(\mathfrak{z})$$

$$T_2 = n^{1/2}\sum_{\mathfrak{z}\in\mathfrak{Z}}(0,\ \hat{m}(\mathfrak{z})-m_\zeta(\mathfrak{z}))'\hat{v}(\mathfrak{z})\lambda(\mathfrak{z})^{-1}\hat{u}_n(\mathfrak{z})$$

$$T_3 = n^{1/2}\sum_{\mathfrak{z}\in\mathfrak{Z}}(1,\ m_\zeta(\mathfrak{z}))'\hat{v}(\mathfrak{z})\lambda(\mathfrak{z})^{-1}(\hat{u}_n(\mathfrak{z})-u_n(\mathfrak{z}))$$

$$T_4 = n^{1/2}\sum_{\mathfrak{z}\in\mathfrak{Z}}(1,\ m_\zeta(\mathfrak{z}))'(1,\ \theta_{20})(\hat{\Pi}(\mathfrak{z})^{-1}-\Pi(\mathfrak{z})^{-1})\psi(\mathfrak{z})\lambda(\mathfrak{z})^{-1}u_n(\mathfrak{z})$$

$$T_5 = n^{1/2}\sum_{\mathfrak{z}\in\mathfrak{Z}}(1,\ m_\zeta(\mathfrak{z}))'(1,\ \theta_{20})\Pi(\mathfrak{z})^{-1}\psi(\mathfrak{z})\lambda(\mathfrak{z})^{-1}u_n(\mathfrak{z})$$

where we denote $u_n(\mathfrak{z}) = I(P(Z^*=\mathfrak{z})\geq\rho/n^\alpha) \times I(P(Z=\mathfrak{z})\geq\rho/n^\alpha)$,

$$\hat{\Pi}(\mathfrak{z}) = \begin{pmatrix}\sum_j I(Z^*_j=\mathfrak{z})/n & 0\\ 0 & \sum_j I(Z_j=\mathfrak{z})/n\end{pmatrix},\quad \Pi(\mathfrak{z}) = \begin{pmatrix}P(Z^*=\mathfrak{z}) & 0\\ 0 & P(Z=\mathfrak{z})\end{pmatrix},$$

$$\psi(\mathfrak{z}) = \begin{pmatrix}\sum_j(\zeta^*_j-m^*(\mathfrak{z}))I(Z^*_j=\mathfrak{z})/n\\ \sum_j(\zeta_j-m(\mathfrak{z}))I(Z_j=\mathfrak{z})/n\end{pmatrix}.$$

Now, $T_5 \xrightarrow{d} N(0,\Omega)$ because if we define

$$X_{nj} = n^{1/2}\sum_{\mathfrak{z}\in\mathfrak{Z}}(1,m_\zeta(\mathfrak{z}))'(1,\theta_{20})\Pi(\mathfrak{z})^{-1}\begin{pmatrix}(\zeta^*_j-m_{\zeta^*}(\mathfrak{z}))I(Z^*_j=\mathfrak{z})/n\\ (\zeta_j-m_\zeta(\mathfrak{z}))I(Z_j=\mathfrak{z})/n\end{pmatrix}\lambda(\mathfrak{z})^{-1}u_n(\mathfrak{z}),$$

then $T_5 = \sum_j X_{nj}$ and X_{nj} is a triangular array with independent random variables within rows which satisfies the Lindeberg condition (see e.g. Serfling 1980, section 1.9.3). Using similar arguments, it is easily proved that $T_i \xrightarrow{p} 0$, for $1\leq i\leq 4$. ∎

REFERENCES

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. (1986), "Asymptotic efficiency in semiparametric models with censoring", *Journal of Econometrics*, 32, 189-218.

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.

Delgado, M.A. (1992), "Semiparametric generalised least squares in the multivariate nonlinear regression model", *Econometric Theory,* 8, 203-222.

Delgado, M.A. and Stengos, T. (1993), "Semiparametric specification testing of non-nested econometric models", forthcoming in *Review of Economic Studies.*

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

Devroye, L. (1978), "The uniform convergence of nearest neighbor regression function estimators and their application in optimization", *IEEE Transactions on Information Theory*, IT-24, 142-151.

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 regressions", *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.

Gasser, T., Müller, H.G., Köhler, W., Molinari, L. and Prader, A. (1984), "Nonparametric regression analysis of growth curves", *Annals of Statistics,* 12, 210-229.

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.

Härdle, W. and Marron, J.S. (1990), "Semiparametric comparison of regression curves", *Annals of Statistics*, 18, 63-89.

Härdle, W. and Stoker, T.M. (1989), "Investigating smooth multiple regression by the method of average derivatives", *Journal of the American Statistical Association*, 84, 986-995.

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

Lawton, W.H., Sylvestre, E.A. and Maggio, M.S. (1972), "Self modeling nonlinear regression", *Technometrics*, 14, 513-532.

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.

Pinkse C.A.P. and Robinson P.M. (1993), "Pooling nonparametric estimates of regression functions with a similar shape", forthcoming in *Statistical Methods of Econometric and Quantitative Econometrics,* vol. in honour of C.R. Rao, (G.S. Maddala et al., eds.).

Powell, J.L., Stock, J.L. and Stoker, T.M. (1989), "Semiparametric estimation of index coefficients", *Econometrica*, 57, 1403-1430.

Rice, J.A. (1986), "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.

Robinson, P.M. (1989), "Hypothesis testing in nonparametric and semiparametric models for economic time series", *Review of Economic Studies*, 56, 511-534.

Robinson, P.M. (1993), "Nearest-neighbour estimation of semiparametric regression models", Manuscript, London School of Economics.

Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

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

Stoker, T.M. (1991), *Lectures on Semiparametric Econometrics*, CORE Lecture Series, Universite Catholique de Louvain.

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.

TABLE 1

Sample size = *40*, Number of replications = *5000*

Non-Smoothing Estimate

	M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8
M	*1.0016*	*0.9993*	*1.0013*	*0.9994*	*1.0039*	*1.0008*	*1.0024*	*0.9993*
E	*0.0294*	*0.0379*	*0.0297*	*0.0384*	*0.0389*	*0.1666*	*0.0331*	*0.0926*

Kernel Estimates (h_1=*1.25*, h_2=*1.75*, h_3=*2.25*)

		M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8
h_1	M	*1.0728*	*1.0540*	*1.1846*	*0.4974*	*1.1831*	*0.4944*	*1.3411*	*0.3244*
	E	*0.0309*	*0.0320*	*0.0724*	*0.4048*	*0.0776*	*0.4972*	*0.1651*	*0.8043*
h_2	M	*1.1150*	*1.0859*	*1.2989*	*0.1647*	*1.2972*	*0.1651*	*1.5198*	*-0.254*
	E	*0.0380*	*0.0346*	*0.1436*	*1.0116*	*0.1469*	*1.1036*	*0.3460*	*2.3823*
h_3	M	*1.1487*	*1.1717*	*1.3307*	*-0.797*	*1.3320*	*-0.821*	*1.5735*	*-1.243*
	E	*0.0481*	*0.0526*	*0.1630*	*3.8732*	*0.1699*	*4.0873*	*0.4120*	*6.5136*

k-NN Estimates (k_1=*3*, k_2=*6*, k_3=*8*)

		M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8
k_1	M	*1.0227*	*1.0699*	*0.9078*	*-1.043*	*0.9035*	*-1.105*	*0.8547*	*-4.892*
	E	*0.0279*	*0.0406*	*0.0804*	*12.848*	*0.1040*	*14.045*	*0.1289*	*48.802*
k_2	M	*1.0321*	*1.1413*	*0.0872*	*-2.667*	*0.8805*	*-2.557*	*0.8615*	*-7.117*
	E	*0.0293*	*0.0532*	*0.0987*	*26.759*	*0.1044*	*25.420*	*0.1592*	*81.239*
k_3	M	*1.0346*	*1.1878*	*0.8786*	*-3.467*	*0.8824*	*-3.523*	*0.8932*	*-8.023*
	E	*0.0282*	*0.0676*	*0.1001*	*33.825*	*0.1150*	*34.891*	*0.1592*	*95.943*

TABLE 2

Sample size = *200*, Number of replications = *2000*

Non-Smoothing Estimate

	M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8
M	0.9998	0.9979	1.0006	0.9993	1.0008	0.9980	0.9991	1.0002
E	0.0052	0.0052	0.0050	0.0052	0.0069	0.0250	0.0058	0.0081

Kernel Estimates (h_1=1.15, h_2=1.45, h_3=1.75)

		M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8
h_1	M	1.0511	1.0284	1.0952	0.6958	1.0907	0.6973	1.1999	0.5076
	E	0.0075	0.0056	0.0152	0.1063	0.0155	0.1209	0.0466	0.2686
h_2	M	1.1022	1.0612	1.2130	0.3545	1.2077	0.3579	1.4037	-0.033
	E	0.0151	0.0083	0.0540	0.4461	0.0537	0.4590	0.1750	1.1400
h_3	M	1.1189	1.0775	1.2599	0.2147	1.2570	0.2129	1.4846	-0.249
	E	0.0188	0.0106	0.0780	0.6560	0.0774	0.6717	0.2486	1.6598

k-NN Estimates (k_1=5, k_2=12, k_3=16)

		M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8
k_1	M	1.0046	1.0210	0.9603	0.2001	0.9636	0.1968	0.9012	-1.971
	E	0.0051	0.0059	0.0140	1.6441	0.0139	1.7434	0.0299	11.352
k_2	M	1.0146	1.0468	0.9236	-0.705	0.9196	-0.645	0.8384	-4.100
	E	0.0053	0.0080	0.0193	4.8134	0.0206	4.5358	0.0472	29.590
k_3	M	1.0187	1.0656	0.9027	-1.133	0.9045	-1.112	0.8207	-5.067
	E	0.0055	0.0105	0.0242	6.9217	0.0249	6.6985	0.0550	40.999

TABLE 3

Sample size = *1000*, Number of replications = *500*

Non-Smoothing Estimate

	M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8
M	*0.9999*	*1.0007*	*0.9992*	*0.9981*	*1.0040*	*1.0012*	*0.9992*	*1.0009*
E	*0.0010*	*0.0011*	*0.0010*	*0.0010*	*0.0013*	*0.0042*	*0.0011*	*0.0012*

Kernel Estimates (h_1=*1.05*, h_2=*1.15*, h_3=*1.25*)

		M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8
h_1	M	*1.0158*	*1.0067*	*1.0133*	*0.9223*	*1.0166*	*0.9176*	*1.0433*	*0.8496*
	E	*0.0012*	*0.0011*	*0.0013*	*0.0071*	*0.0015*	*0.0116*	*0.0029*	*0.0244*
h_2	M	*1.0513*	*1.0317*	*1.0859*	*0.7077*	*1.0847*	*0.7113*	*1.1794*	*0.4674*
	E	*0.0035*	*0.0020*	*0.0086*	*0.0875*	*0.0086*	*0.0886*	*0.0335*	*0.2879*
h_3	M	*1.0777*	*1.0434*	*1.1302*	*0.5482*	*1.1342*	*0.5466*	*1.2657*	*0.1977*
	E	*0.0069*	*0.0027*	*0.0183*	*0.2069*	*0.0197*	*0.2111*	*0.0725*	*0.6505*

k-NN Estimates (k_1=*11*, k_2=*22*, k_3=*28*)

		M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8
k_1	M	*1.0016*	*1.0080*	*0.9768*	*0.6732*	*0.9787*	*0.6515*	*0.9359*	*-0.608*
	E	*0.0010*	*0.0009*	*0.0020*	*0.1909*	*0.0022*	*0.2356*	*0.0063*	*2.9525*
k_2	M	*1.0027*	*1.0153*	*0.9610*	*0.3599*	*0.9661*	*0.3448*	*0.9153*	*-1.553*
	E	*0.0009*	*0.0011*	*0.0036*	*0.5639*	*0.0038*	*0.6295*	*0.0107*	*7.0295*
k_3	M	*1.0029*	*1.0183*	*0.9639*	*0.2252*	*0.9630*	*0.2153*	*0.9073*	*-1.977*
	E	*0.0009*	*0.0013*	*0.0032*	*0.7873*	*0.0037*	*0.8683*	*0.0119*	*9.4193*

TABLE 4

Non-smoothing estimates (n=40, r=10000)

		$\hat{\theta}_1$		$\hat{\theta}_2$		$\tilde{\theta}_1$		$\tilde{\theta}_2$	
		ε_1	ε_2	ε_1	ε_2	ε_1	ε_2	ε_1	ε_2
M 8	M	9.969	9.658	2.025	2.141	10.059	10.062	1.990	1.989
	V	3.942	60.167	0.533	8.221	2.314	2.315	0.272	0.272
M 9	M	10.267	10.235	1.948	1.953	10.266	10.270	1.948	1.947
	V	2.191	3.063	0.087	0.127	2.529	2.520	0.101	0.101
M10	M	10.090	10.067	1.968	1.976	10.073	10.073	1.969	1.968
	V	0.281	0.338	0.076	0.192	0.297	0.298	0.085	0.085
M11	M	9.985	9.801	2.010	1.999	10.032	10.032	1.993	1.992
	V	0.548	1.934	0.090	0.141	0.490	0.490	0.082	0.082
M12	M	12.491	12.530	0.690	0.660	12.554	12.560	0.650	0.647
	V	1.443	1.955	0.377	0.522	2.425	2.465	0.647	0.656

TABLE 5

Kernel estimates (n=40, r=10000)

		$\hat{\theta}_1$		$\hat{\theta}_2$		$\tilde{\theta}_1$		$\tilde{\theta}_2$	
		ε_1	ε_2	ε_1	ε_2	ε_1	ε_2	ε_1	ε_2
M 8	M	9.746	9.583	2.093	2.155	9.804	9.807	2.071	2.070
	V	3.923	8.200	0.475	1.105	3.134	3.130	0.354	0.354
M 9	M	10.020	9.971	1.998	2.008	10.013	10.015	1.999	1.999
	V	2.271	3.070	0.090	0.127	2.543	2.538	0.102	0.102
M10	M	10.089	10.053	1.974	1.989	10.067	10.068	1.982	1.981
	V	0.509	0.600	0.234	0.381	0.501	0.501	0.245	0.244
M11	M	10.008	9.719	1.974	2.024	10.059	10.062	1.955	1.953
	V	2.228	3.276	0.251	0.259	2.487	2.504	0.286	0.290
M12	M	11.670	11.708	1.125	1.100	11.732	11.733	1.087	1.087
	V	2.405	3.567	0.643	0.980	1.087	1.087	1.152	1.198

FIGURE 1

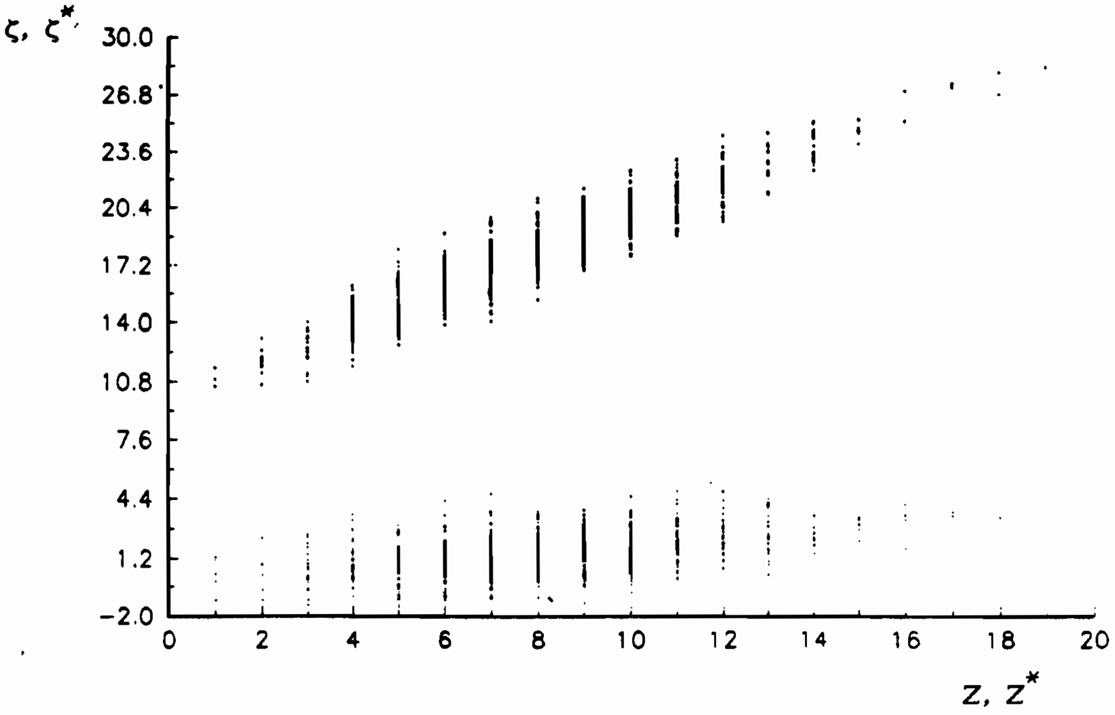
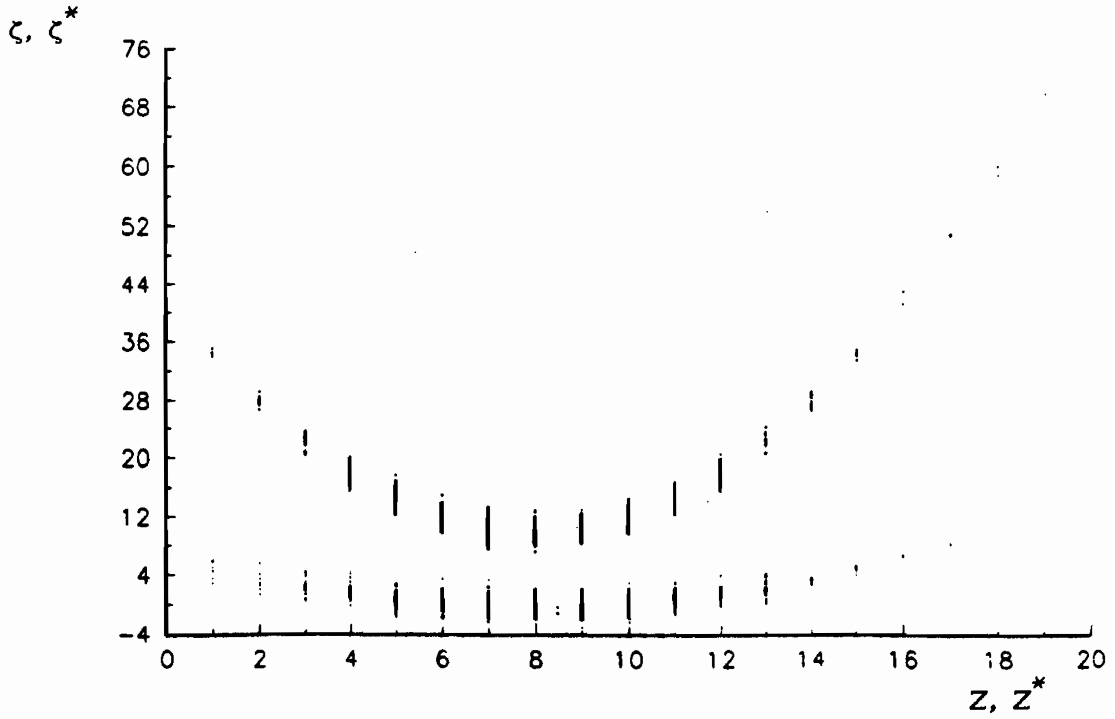


FIGURE 2


ζ, ζ*
76
68
60
52
44
36
28
20
12
4
-4
0
2
4
6
8
10
12
14
16
18
20
z, z*