UNIVERSIDAD DE SAN ANDRÉS

Seminario del Departamento de Economía

"Nonparametric estimation of nonadditive random functions"

Rosa L. Matzkin
(Northwestern University)

Martes 8 de junio de 1999
11 hs.
Aula Roberto J. Lebach

Sem.
Eco.
99/8

UdeSA

NONPARAMETRIC ESTIMATION OF NONADDITIVE RANDOM FUNCTIONS

Rosa L. Matzkin*

Department of Economics

Northwestern University

This version: April 1999

Abstract

We present an estimator for a nonparametric function that depends on an unobservable random variable in a nonadditive way. The unobservable random term is assumed to be distributed independently of the observable exogenous variables, with an unknown distribution. We show that, when the function satisfies some properties that are implied by economic theory, such as homogeneity of degree one, the nonparametric function and the distribution of the unobservable random term are identified subject to minor normalizations. For the cases in which the properties of economic theory are not satisfied, we provide a convenient normalization. The estimators are shown to be consistent and asymptotically normal.

*The support of NSF is gratefully acknowledged. I have benefitted from my interaction with Joe Altonji, Jim Heckman, Chuck Manski, and Whitney Newey, and the research assistance of Ulrich Doraszelski and Elie Tamer. This paper is to be presented at the Invited Lecture on New Developments in the Estimation of Preferences and Production Functions, Latin American Meeting of the Econometric Society, August 1999.

UNIVERSIDAD DE SAN ANDRES
BIBLIOTECA

1 Introduction

A common practice when estimating many economic models proceeds by first specifying the relationship between a vector of observable exogenous variables, X, and a dependent variable, Y, and then, adding a random unobservable term, ε, to the relationship. In the resulting model, ε is typically interpreted as the difference between the observed value of the dependent variable, Y, and the conditional expectation of Y given X. This procedure has been criticized on the grounds that instead of adding an unobservable random term to the relationship, as an after-thought, one should be able to generate this unobservable random term from within the model. When approaching the random relationship in the latter way, ε may represent an heterogeneity parameter in a utility function, some productivity shock in a production function, or some other relevant unobservable variable. When using this approach, the random term ε rarely appears in the model as a term added to the conditional expectation of Y given X. (See McElroy (1981, 1987) and Brown and Walker (1989, 1995).) In general, even when parametric functions are used to specify the underlying functions in the economic model, the resulting function by which the values of Y are determined from X and ε is nonlinear in ε.

Most nonparametric methods that are currently used to specify the relationship between the vector of observable exogenous variables, X, an unobservable term, and the observable dependent variable, Y, define the unobservable random term as being the difference between Y and the conditional expectation. The resulting model is then one where the unobservable random term is added to the relationship. Although one could interpret this added unobservable random term as being a function of the observable and unobservable variables, the existent methods do not provide a way of studying this function, which has information about the important interaction between the observable and unobservable variables.

In this paper, we present a nonparametric method of estimating the relationship between a dependent variable Y, an observable vector of exogenous variables, X, and an unobservable exogenous variable ε. The method does not require that the unobservable variable ε be additive. It also does not require that either the function or the distribution of the unobservable random term be parametric. More specifically, we consider the model $Y = m(X,\varepsilon)$

where ε is distributed independently of X, and both the function m and the distribution of ε are unknown. While, without any restrictions, the function m and the distribution of ε are not jointly identified, we present some restrictions that, in some economic models such as when m is a cost or profit function, may be derived from economic theory. When these restrictions are satisfied, only minor normalizations, such as fixing the value of the function m at one point, may suffice to jointly identify the function m and the distribution of ε. For the cases in which economic theory does not imply these restrictions, we present convenient normalizations.

The estimators are very easy to calculate. They are defined as nonlinear functionals of a nonparametric estimator of the joint cumulative distribution function of the observable variables. We present the asymptotic properties of the estimators for the case in which the nonparametric estimator of the cumulative distribution function is obtained by kernel methods. The estimators are shown to be consistent and asymptotically normal.

Other papers that consider nonparametric models where the random terms do not enter in an additive form are Roehrig (1988), Brown and Matzkin (1996), Altonji and Ichimura (1997), and Altonji and Matzkin (1997). Roehrig (1988) provides a general condition for the identification of nonparametric systems of equations. Brown and Matzkin (1996) extend Roehrig (1988)'s conditions and provide an extremum estimator for estimating nonparametric simultaneous equations of the from studied in Roehrig (1988). Altoniji and Ichimura (1997) consider models with one dependent variable, and estimate an average derivative. Altonji and Matzkin (1997) consider the estimation of models for panel data. In nonparametric models where the unobservable random term is additive, shape restrictions have been used in previous work to identify otherwise unidentified functions, and to estimate nonparametric models. (Matzkin (1994) reviews some of the existent literature for limited dependent variable models and nonparametric regression functions.) In a different vein, Manski (1997), uses restrictions of economic theory to determine bounds for the distribution of a dependent variable.

The outline of the paper is as follows. In the next section, we present the model and two sets of conditions under which the function m and the distribution of ε are identified. In Section 3, we present the estimators and their asymptotic properties. Section 4 presents additional conditions under which m and the distribution of ε are identified. The results of some simulations are described in Section 5.

2 The Model

The model that we will deal with is

(1) $Y = m(X, \varepsilon)$

where $m : A \times E \to R$ is strictly increasing in ε, $A \subset R^K$, $E \subset R$, y and X are observable, and ε is an unobservable random term which is distributed independently of X. If the economic model contains several unobservable random terms, then the function m is assumed to be weakly separable in all these terms, and ε denotes the value of the function that aggregates these terms. We will denote the distribution of ε by F_ε.

One possible example of this model, to which we will return below, is where m denotes the profit function of a typical firm, X is a vector of the observable output and input prices and ε is the unobservable price of a (possibly unobservable) input. As another example, let Y denote the cost of undertaking a particular project by a typical firm, and suppose that all the inputs prices are observed except for one, which is distributed independently of the others. If X denotes the vector of observable input prices and ε denotes the unobservable price, then m is the cost function, which, in general, is not additively separable in ε. Note that in these two examples, economic theory implies that the function m is homogenous of degree one in X and ε.

The first question that arises when specifying the model in (1) is whether one can identify the function m and the distribution of ε. Following the standard definition of identification, we say that (m, F_ε) is identified if we can uniquely recover it from the distribution of the observable variables. More specifically, let M denote a set to which the function m belongs, and let Γ denote a set to which F_ε. Let $F_{Y,X}(\cdot; m', F'_\varepsilon)$ denote the joint cdf of the observable variables when $m = m'$ and $F_\varepsilon = F'_\varepsilon$. Then,

Definition: *The pair* (m, F_ε) *is identified in the set* $(M \times \Gamma)$ *if* (i) $(m, F_\varepsilon) \in (M \times \Gamma)$ *and (ii) for all* (m', F'_ε) *in* $(M \times \Gamma)$
$[F_{Y,X}(\cdot; m, F_\varepsilon) = (F_{Y,X}(\cdot; m', F'_\varepsilon))] \Longrightarrow (m', F'_\varepsilon) = (m, F_\varepsilon)$

If for any two functions, m' and $m"$ in M, we can find distributions, F'_ε and $F_\varepsilon"$ in Γ such that the pairs (m', F'_ε) and $(m", F_\varepsilon")$ generate the same distribution of observable variables, m' and $m"$ are said to be observationally equivalent.

Definition: *Any two functions, m' and $m"$ in M are said to be observationally equivalent if there exist $F'_\varepsilon, F''_\varepsilon$ in Γ such that for all (y, x)*
$F_{y,x}(y, x; m', F'_\varepsilon) = F_{y,x}(y, x; m'', F''_\varepsilon)$.

Below, we present some sets of functions in which (m, F_ε) is identified. The following assumptions will be made:

Assumption I.1: ε is distributed independent of X. The support of ε includes the set E.

Assumption I.2: $\forall x$, $m(x, \cdot)$ is strictly increasing in ε.

Assumption I.3: The support of X includes the set A.

The independence between ε and X implies that the conditional probability density of Y given X is generated, for all values of X, by a common distribution of ε. In particular, if the distribution of ε can be recovered from the conditional cdf of Y given X at some value of X, then it can be used to identify m from the conditional cdf of Y given X at any other value of X. Assumption I.2 guarantees that, given m and X, the pdf of ε can be obtained from the conditional pdf of Y given X. Our results can be easily modified if m is strictly deceasing, instead of increasing, in ε. The support conditions on ε and X are made to guarantee that all the values of (x, ε) in the domain of the function m are in the support of the joint distribution of (x, ε). If m were a parametric function, these support conditions would not, in general, be necessary to identify m.

We describe below two possible sets of functions to which the function m may belong. In Section 4, we consider two other sets of functions. We will show that when the distribution of ε belongs to the set, Γ, of strictly increasing distributions whose support includes the set E, and m belongs to

QUAERERE VERUM Universidad de San Andrés

any one of these sets of functions, the pair (m, F_ε) is identified.

Let I denote the set of functions on (x, ε) that are strictly increasing in ε, i.e.

$$I = \{m' : A \times E \to R | \text{for all } X \in A,\ m'(x, \cdot) \text{ is strictly increasing}\}$$

The first is the set of functions that are strictly increasing in ε, attain a given value at one point, and are homogeneous of degree one along the ray that goes through that point. Let $\overline{x} \in R^K$, $\overline{\varepsilon} \in R$, and $\alpha \in R$ be given. Then, the set $M1$ is defined by:

$$M1{=}\{m' : A \times E \to R |\, m' \in I,\ m'(\overline{x}, \overline{\varepsilon}) = \alpha \ \ and \ \ \forall \lambda \in R,\ m'(\lambda \overline{x}, \lambda \overline{\varepsilon}) = \lambda \alpha\}$$

In some economic models, such as the ones described above where Y denotes either the profit or the cost of a typical firm, m is known to be homogenous of degree one in (x, ε) and strictly increasing (decreasing) in ε. In such cases, the only restriction in m guaranteeing that it belongs to $M1$ is that it attains the value α at $(\overline{x}, \overline{\varepsilon})$. It will follow from our analysis below that the latter is the only normalization needed in the set of homogenous of degree one functions that are strictly increasing in ε. Hence, as long as we specify the value of m at one point, we can identify m and the distribution of ε.

The second set of functions provides a convenient normalization for the set of functions that are strictly increasing in ε. It is the set of functions that are strictly increasing in ε and whose values when X equals some vector, $\overline{x}$, are known to equal ε. Let $\overline{x} \in R^K$. Then, the second set of functions is

$$M2{=}\{m' : A \times E \to R \mid m' \in I \text{ and } m'(\overline{x}, \varepsilon) = \varepsilon\}$$

For any function $\widetilde{m}(x, \varepsilon)$ that is strictly increasing in ε, there exists a function $m'(x, \varepsilon)$ that is observationally equivalent to $\widetilde{m}(x, \varepsilon)$ and satisfies $m'(\overline{x}, \varepsilon)$ $= \varepsilon$. The function m' is defined by $m'(x, \varepsilon) = \widetilde{m}(x, \widetilde{m}^{-1}(\overline{x}, \varepsilon))$, where by $\widetilde{m}^{-1}$ we mean the inverse with respect to the last coordinate (Altonji and Matzkin (1997). Hence, the requirement that $m(\overline{x}, \varepsilon) = \varepsilon$ does not restrict the set of identified functions.

A common property of the two sets of functions described above is that the functions in each of these sets attain common known values on a set that is mapped into the real line. In $M1$, this is the ray from the origin that passes through the point $(\overline{x}, \overline{\varepsilon})$. In $M2$, this is the set of all vectors for which $X = \overline{x}$. Let

$\Gamma = \{F : R \rightarrow R \mid F \text{ is strictly increasing}\}$.

Our first identification result is the following:

Theorem 1 *Suppose that Assumptions I.1-I.3 are satisfied. Then,*
(i) if m belongs to $M1$, (m, F_ε) is identified in $(M1 \times \Gamma)$
(ii) if m belongs to $M2$, (m, F_ε) is identified in $(M2 \times \Gamma)$

Proof of Theorem 1: We first note that for all $e \in E$ and $X \in A$,

(1) $F_\varepsilon(e) = F_{Y|X=x}(m(x,e))$.

This is because $F_\varepsilon(e) = \Pr(\varepsilon \leq e) = P(\varepsilon \leq e|X = x) = \Pr(m(x,\varepsilon) \leq m(x,e) \mid X = x) = \Pr(Y \leq m(x,e) \mid X = x) = F_{Y|X=x}(m(x,e))$. The first equality follows by the definition of F_ε, the second follows by the independence between ε and X, the third follows by the monotonicity of $m(\overline{x}, \cdot)$, the fourth follows by the definition of Y, and the fifth equality follows by the definition of $F_{Y|X}$.

Suppose first that $m \in M2$. Then, letting $X = \overline{x}$ in (1) and noticing that $m(\overline{x}, \varepsilon) = \varepsilon$, we get that

(2) $F_\varepsilon(e) = F_{Y|X=\overline{x}}(e)$.

Hence, F_ε is identified from the conditional cdf of Y given $X = \overline{x}$. Next, from (1) and (2) it follows that for all $e \in E$ and $X \in A$, $F_{Y|X=\overline{x}}(e) = F_{Y|X=x}(m(x,e))$. Since $Y = m(X,\varepsilon)$ and $m(x,\cdot)$ is strictly increasing, it follows that the conditional cdf of Y given $X = x$ is strictly increasing on the set $m(x,E) = \{y|y = m(x,\varepsilon),\ \varepsilon \in E\}$; hence $F_{Y|X}$ has an inverse on $m(x,E)$. It follows then that

(3) $m(x,e) = F^{-1}_{Y|X=x}\left(F_{Y|X=\overline{x}}(e)\right).$

Hence, the function m is identified.

Suppose, next, that $m \in M1$. Then, given any $\lambda \in R$ and letting $x = \lambda\overline{x}$ and $e = \lambda\overline{\varepsilon}$, we have that $F_\varepsilon(\lambda\overline{\varepsilon}) = F_{Y|X=\lambda\overline{x}}\left(m(\lambda\overline{x},\lambda\overline{\varepsilon})\right) = F_{Y|X=\lambda\overline{x}}(\lambda\alpha)$, where the second equality follows because $m(\lambda\overline{x},\lambda\overline{\varepsilon}) = \lambda m(\overline{x},\overline{\varepsilon}) = \lambda\alpha$. In particular, for any $e \in E$,

(4) $F_\varepsilon(e) = F_{Y|X=(e/\overline{\varepsilon})\overline{x}}\left((e/\overline{\varepsilon})\alpha\right),$

by letting $\lambda = (e/\overline{\varepsilon})$. Hence, $F_\varepsilon(e)$ is identified from the conditional cdf of Y given X, when $y = (e/\overline{\varepsilon})\alpha$ and $x = (e/\overline{\varepsilon})\overline{x}$. From (1) and (4), we have that for any $e \in E$ and $x \in A$, $F_{Y|X=(e/\overline{\varepsilon})\overline{x}}\left((e/\overline{\varepsilon})\alpha\right) = F_{Y|X=x}\left(m(x,e)\right)$. Using the fact that $F_{Y|X=x}(\cdot)$ has an inverse, we get that

(5) $m(x,e) = F^{-1}_{Y|X=x}\left(F_{Y|X=(e/\overline{\varepsilon})\overline{x}}\left((e/\overline{\varepsilon})\alpha\right)\right).$

Hence, the function m is identified. ■

If the functions in the sets were required to be differentiable, and the joint cdf of (Y,X) were assumed to be absolutely continuous and with support R^{K+1}, then the above theorem could have been proved by showing that the functions in $M1$ and $M2$ satisfy the rank conditions given in Roehrig (1988). Instead, we used the above argument, because besides not requiring the additional conditions, suggests a method of obtaining estimators for the function m and the distribution of ε.

3 Estimation

Since, in equations (2)-(5), the functions m and F_ε are expressed in terms of the conditional cdf's of Y given X, we can obtain estimators for m and F_ε by substituting the conditional cdf of Y given X by a nonparametric estimator of it. While one could consider using any type of nonparametric estimators for the conditional cdf's of Y given X, we present here the details and asymptotic properties for the case in which the conditional cdf's are estimated using the method of kernels.

Let the data be denoted by $\{X^i, Y^i\}_{i=1}^N$. Let $f(y,x)$ and $F(y,x)$ denote, respectively, the joint pdf and cdf of (Y,X), let $\hat{f}(y,x)$ and $\hat{F}(y,x)$ denote, respectively, their kernel estimators, and let $\hat{f}_{Y|X=x}(y)$ and $\hat{F}_{Y|X=x}(y)$ denote the kernel estimators of, respectively, the conditional pdf and conditional cdf of Y given $X = x$. Then,

$$\hat{f}(y,x) = \frac{1}{N\sigma_N^{K+1}} \sum_{i=1}^N K\left(\frac{y-Y^i}{\sigma}, \frac{x-X^i}{\sigma}\right) \quad \text{for all } (y,x) \in R^d,$$

$$\hat{F}(y,x) = \int_{-\infty}^y \int_{-\infty}^x \hat{f}_N(s,z)\, ds\, dz,$$

$$\hat{f}_{Y|X=x}(y) = \frac{\hat{f}_N(y,x)}{\int_{-\infty}^\infty \hat{f}_N(s,x)\, ds}, \text{ and}$$

$$\hat{F}_{Y|X=x}(y) = \frac{\int_{-\infty}^y \hat{f}_N(s,x)\, ds}{\int_{-\infty}^\infty \hat{f}_N(s,x)\, ds}$$

where $K : R \times R^K \to R$ is a kernel function and σ_N is the bandwidth. Note that the estimator for the conditional cdf of Y given X is different from the Nadaraya-Watson estimator for $F_{Y|X=x}(y)$. The latter is the kernel estimator for the conditional expectation of $Z \equiv 1[Y \leq y]$ given $X = x$. For any t and x, $\hat{F}^{-1}_{Y|X=x}(t)$ will denote the set of values of y for which $\hat{F}_{Y|X=x}(y)$ $= t$. When m belongs to $M1$, the estimators for F_ε and m are defined by

$$\widehat{F}_\varepsilon(e) = \widehat{F}_{Y|X=(e/\overline{\varepsilon})\overline{x}}\left((e/\overline{\varepsilon})\alpha\right) \text{ and}$$

$$\widehat{m}(x,e) = \widehat{F}^{-1}_{Y|X=x}\left(\widehat{F}_{Y|X=(e/\overline{\varepsilon})\overline{x}}\left((e/\overline{\varepsilon})\alpha; x\right)\right).$$

When m belongs to $M2$, they are defined by

$$\widehat{F}_\varepsilon(e) = \widehat{F}_{Y|X=\overline{x}}(e) \text{ and}$$

$$\widehat{m}(x,e) = \widehat{F}^{-1}_{Y|X=x}\left(\widehat{F}_{Y|X=\overline{x}}(e; x)\right).$$

To derive the asymptotic properties of the estimators, we make the following assumptions:

Assumption S.1: The sequence $\{Y^i, X^i\}$ is a strictly stationary $\beta-$mixing sequence satisfying $k^\nu \beta_k \to 0$ as $k \to \infty$, for some $\nu > 1$.

Assumption S.2: $f(y,x)$ is continuously differentiable up to the order s, where s is the first even integer larger than or equal to K+1. The support of f is compact. The derivatives of $f(y,x)$ up to order s are bounded and in $L^2(R^{K+1})$.

Assumption S.3: The kernel function $K(\cdot,\cdot)$ is an even function, integrates to 1, is of order s, is continuously differentiable up to the order $s+K+1$, and its derivatives of order up to s are in $L^2(R^{K+1})$. The value of K is zero outside a compact set.

Assumption S.4: As $N \to \infty$ $\sqrt{N}\sigma_N^{s+\frac{K}{2}} + \frac{1}{\sqrt{N}\sigma_N^{2m}} \to 0$ where $r > 0$ and $0 \leq m < s/2 + K/4$.

Assumption S.1 allows for dependence across observations. Assumption S.2 requires that the pdf of (Y, X) be sufficiently smooth. Note that this requires ε to have a smooth enough density. The support of f is required to be compact in order to guarantee that f can be approximated by functions that vanish outside a compact set. Assumption S.3 restricts the kernel function that may be used. Assumption S.4 restricts the rate at which the bandwidth, σ_N goes to zero. In the next two results, we make use of these assumptions to establish the consistency and asymptotic normality of our estimators. The proofs of these results are presented in the Appendix. The first result concerns the estimator for the cdf of ε.

Theorem 2 : *Suppose that Assumptions I.1-I.3 and S.1-S.4 are satisfied, and that the function* m *belongs to either* $M1$ *or* $M2$. *Then,*

(i) $\sup_{e \in R} \left| \widehat{F}_\varepsilon(e) - F_\varepsilon(e) \right| \to 0$ *in probability and*

(ii) *for all* e, $\sqrt{N}\sigma^{(K/2)} \left(\widehat{F}_\varepsilon(e) - F_\varepsilon(e) \right) \to N(0, V_F)$,

where, if $m \in M1$, $V_F = \left\{ \int \left(\int K(s,z)\, ds \right)^2 dz \right\} [F_\varepsilon(e)\ (1 - F_\varepsilon(e)] \, [1/f((e/\overline{\varepsilon})\overline{x})]$,

and if $m \in M2$, $V_F = \left\{ \int \left(\int K(s,z)\, ds \right)^2 dz \right\} [F_\varepsilon(e)\ (1 - F_\varepsilon(e)] \, [1/f(\overline{x})]$.

The above asymptotic distribution is the same as that of the estimator for the conditional distribution of Y given X, which, itself, is the same as the asymptotic distribution of the Nadaraya-Watson estimator for the conditional expectation of $Z = [Y \leq y]$ given X. The next result concerns the estimator for the function m. Let $int(C)$ denote the interior of the set C.

Theorem 3 *Suppose that Assumptions I.1-I.3 and S.1-S.4 are satisfied. Let* $\delta > 0$. *Let* $C = \{s | f(s,x) \geq \delta\}$. *Suppose that* $m(x,e) \in int(C)$ *and* $f(x) > 0$. *If* $m \in M1$, *assume that* $f((e/\overline{\varepsilon})\overline{x}) > 0$. *If* $m \in M2$, *assume that* $f(\overline{x}) > 0$. *Then,*

(i) $\widehat{m}(x,e)$ *converges in probability to* $m(x,e)$, *and*

(ii) $\sqrt{N}\sigma_N^{K/2} (\hat{m}(x,e) - m(x,e)) \to N(0, V_m)$ *in distribution,*

where if $m \in M1$, $V_m = \left\{ \int \left(\int K(s,z)\, ds \right)^2 dz \right\} \left[\frac{F_\varepsilon(e)\ (1-F_\varepsilon(e))}{f_{Y|X=x}(m(x,e))^2} \right] \left[\frac{1}{f(x)} + \frac{1}{f((e/\overline{\varepsilon})\overline{x}))} \right]$

and if $m \in M2$, $V_m = \left\{ \int \left(\int K(s,z)\, ds \right)^2 dz \right\} \left[\frac{F_\varepsilon(e)\ (1-F_\varepsilon(e))}{f_{Y|X=x}(m(x,e))^2} \right] \left[\frac{1}{f(x)} + \frac{1}{f(\overline{x})} \right]$

The rate of convergence in (ii) is the same as that of the Nadaraya-Watson estimator for the conditional expectation of Y given X. The kernel function influences the asymptotic variance of our estimator in the same way as it does for the Nadaraya-Watson estimator. However, while the asymptotic variance of the Nadaraya-Watson estimator depends on the pdf of X only through the value that this attains at x. The asymptotic variance of our estimator depends also on the value of the pdf of X at the value of X that is used to estimate the value of the cdf of ε at e. This is because $\widehat{m}(x,e)$ is derived from $\widehat{F}_{Y|X=((e/\overline{\varepsilon})\overline{x})}((e/\overline{\varepsilon})\alpha)$ (or $\widehat{F}_{Y|X=\overline{x}}(e)$, if $m \in M2$) and $\widehat{F}_{Y|X=x}(m(x,e))$. Hence, the asymptotic variance of $\widehat{m}(x,e)$ depends on the asymptotic variances of $\widehat{F}_{Y|X=\overline{x}}(e)$ and $\widehat{F}_{Y|X=x}(m(x,e))$. The asymptotic variance of our estimator depends also on the value of the conditional pdf of Y given $X = x$ at $m(x,e)$. If $m(x,\cdot)$ is differentiable, $f_{Y|X=x}(m(x,e)) = f_{\varepsilon}(e) \;/\; |\partial m(x,e)/\partial\varepsilon|$. Hence, the asymptotic variance decreases the larger is the value of the pdf of ε at e and the "flatter" $m(x,\varepsilon)$ is with respect to ε when $\varepsilon = e$.

4 Additive Separable Functions

In some cases, the economic model might imply that the function m is the addition of two functions, of which only one of them depends on epsilon. Consider, for example, the case where Y is the cost of undertaking a particular project. Y may be the sum of a fixed and a variable cost. If ε denotes the unobservable price of a variable input, we may specify the model as $Y = v_1(x_1) + v_2(x_2,\varepsilon)$, where x_1 are variables that affect the fixed cost, and (x_2,ε_1) represents the vector of prices of the variable inputs.

When the function m is additively separable, we may identify m and F_ε under weaker restrictions than the ones presented in Section 2. We present below two sets of additive separable functions such that when m belongs to any of these functions, m and F_ε are identified.

Let $\overline{x} \in R^K$, $\overline{\varepsilon} \in R$, and $\alpha \in R$ be given. Let $K_1, K_2 > 0$ be such that $K_1 + K_2 = K$. Let $\gamma \in R$ be also given, and denote $x = (x_1,x_2) \in A_1 \times A_2 \subset R^{K_1} \times R^{K_2}$. We consider sets of functions of the form $m'(x_1,x_2) = v_1'(x_1) + v_2'(x_2,\varepsilon)$ where v_1' belongs to the set

$$V_1 = \{v_1' : A_1 \to R|\; v_1'(\overline{x}_1) = \gamma\},$$

and v_2' belongs to either $M1$ or $M2$. More specifically, let

$$I_2 = \{v_2' : A_2 \times E \to R|\forall x_2 \in A_2 \;\; v_2'(x_2,\cdot) \text{ is strictly increasing}\},$$

$$V_{2a} = \{v_2' : A_2 \times E \to R | v_2' \in I_2,\ v_2'(\overline{x}_2, \overline{\varepsilon}) = \alpha, \text{ and } \forall \lambda \in R,\ v_2'(\lambda \overline{x}_2, \lambda \overline{\varepsilon}) = \lambda \alpha\}.$$

and

$$V_{2b} = \{v_2' : A_2 \times E \to R | v_2' \in I_2 \text{ and } v_2'(\overline{x}_2, \varepsilon) = \varepsilon\}.$$

Then, we define the sets

$$M3a = \{m' | \exists v_1' \in V_1, v_2' \in V_{2a} \text{ such that } \forall x_1, x_2, \varepsilon,\ m'(x_1, x_2, \varepsilon) = v_1'(x_1) + v_2'(x_2, \varepsilon)\}$$

and

$$M3b = \{m' | \exists v_1 \in V_1, v_2 \in V_{2b} \text{ such that } \forall x_1, x_2, \varepsilon,\ m'(x_1, x_2, \varepsilon) = v_1'(x_1) + v_2'(x_2, \varepsilon)\}.$$

So, when v_1 denotes a fixed cost and v_2 denotes a variable cost, the function m belongs to $M3b$ if the values of v_1 and v_2 are specified at one value.

Note that, as it was the case with $M1$ and $M2$, all the functions in $M3a$ and $M3b$ attain common known values on a set that is mapped into the real line. In $M3a$, all the functions attain the same known values on the set $\{(\overline{x}_1, \lambda \overline{x}_2) | \lambda \in R\}$. In $M3b$, all the functions attain the same known values on the set $(x_1, x_2) = (\overline{x}_1, \overline{x}_2)$. This property allows us to identify the distribution of epsilon from the conditional cdf of Y given X. The cdf of epsilon can then be used to identify the function m from the conditional cdf of Y given X when $X = x$. The next result formally states the identification result for these types of functions.

Theorem 4 *Suppose that Assumptions I.1-I.3 are satisfied. Then,*
(i) if m belongs to $M3a$, $(m, v_1, v_2, F_\varepsilon)$ is identified in $(M3a \times V_1 \times V_{2a} \times \Gamma)$,
(ii) if m belongs to $M3b$, $(m, v_1, v_2, F_\varepsilon)$ is identified in $(M3b \times V_1 \times V_{2b} \times \Gamma)$.

The proof, which is presented in the Appendix, provides expressions for F_ε, m, v_1, and v_2 in terms of the conditional cdf of Y given X. When this conditional cdf is substituted by its kernel estimator, we obtain the estimators for F_ε, m, v_1, and v_2. When $m \in M3a$,

$$\widehat{F}_{\varepsilon}(e) = \widehat{F}_{Y|X=(\overline{x}_1,(e/\overline{\varepsilon})\overline{x}_2)}\left(\gamma + ((e/\overline{\varepsilon})\alpha)\right),$$

$$\widehat{m}(x,e) = \widehat{F}^{-1}_{Y|X=(x_1,x_2)}\left(\widehat{F}_{Y|X=(\overline{x}_1,(e/\overline{\varepsilon})\overline{x}_2)}\left(\gamma + (e/\overline{\varepsilon})\alpha\right)\right),$$

$$\widehat{v}_2(x_2,\varepsilon) = \widehat{m}(\overline{x}_1, x_2, e) - \gamma \text{ and}$$

$$\widehat{v}_1(x_1) = \widehat{m}(x_1, \overline{x}_2, \overline{\varepsilon}) - \alpha.$$

And, when m belongs to $M3b$,

$$\widehat{F}_{\varepsilon}(e) = \widehat{F}_{Y|X=(\overline{x}_1,\overline{x}_2)}\left(\gamma + e\right),$$

$$\widehat{m}(x,e) = \widehat{F}^{-1}_{Y|X=(x_1,x_2)}\widehat{F}_{Y|X=(\overline{x}_1,\overline{x}_2)}\left(\gamma + e\right),$$

$$\widehat{v}_2(x_2,\varepsilon) = \widehat{m}(\overline{x}_1, x_2, e) - \gamma \quad \text{and for any } e \in E$$

$$\widehat{v}_1(x_1) = \widehat{m}(x_1, \overline{x}_2, e) - e.$$

The asymptotic properties of the above estimators for F_{ε} are the same as those stated in Theorem 2, except that, if $m \in M3a$,

$$V_F = \left\{\int\left(\int K(s,z)\, ds\right)^2 dz\right\} \cdot \left[F_{\varepsilon}(e)\ (1 - F_{\varepsilon}(e)\right]\left[1/f(\overline{x}_1,(e/\overline{\varepsilon})\overline{x}_2)\right],$$

and if $m \in M3b$,

$$V_F = \left\{\int\left(\int K(s,z)\, ds\right)^2 dz\right\}\left[F_{\varepsilon}(e)\ (1 - F_{\varepsilon}(e)\right]\left[1/f(\overline{x}_1,\overline{x}_2)\right].$$

Properties (i) and (ii) of Theorem 3 are satisfied by the estimators for $m(x,e)$ in the additive separable case. If $m \in M3a$,

$$V_m = \left\{\int\left(\int K(s,z)\, ds\right)^2 dz\right\}\left[\frac{F_{\varepsilon}(e)\ (1-F_{\varepsilon}(e))}{f_{Y|X=x}(m(x,e))^2}\right]\left[\frac{1}{f(x)} + \frac{1}{f(\overline{x}_1,(e/\overline{\varepsilon})\overline{x}_2)}\right]$$

and if $m \in M3b$,

$$V_m = \left\{\int\left(\int K(s,z)\, ds\right)^2 dz\right\}\left[\frac{F_{\varepsilon}(e)\ (1-F_{\varepsilon}(e))}{f_{Y|X=x}(m(x,e))^2}\right]\left[\frac{1}{f(x)} + \frac{1}{f(\overline{x}_1,\overline{x}_2)}\right].$$

Universidad de SanAndrés

When $m \in M3a$, the asymptotic properties of $\widehat{v}_1(x_1)$ and $\widehat{v}_2(x_2, \varepsilon)$ are, respectively, those of $\widehat{m}(x_1, \overline{x}_2, \overline{e})$ and $\widehat{m}(\overline{x}_1, x_2, e)$. The asymptotic properties of $\widehat{v}_1(x_1)$ and $\widehat{v}_2(x_2, \varepsilon)$ when $m \in M3b$ are those of $\widehat{m}(x_1, \overline{x}_2, e)$ and $\widehat{m}(\overline{x}_1, x_2, e)$.

We could exploit the additivity of m to develop estimators for v_1 and v_2 that possess better asymptotic properties than the estimators described above. This is, however, beyond the scope of this paper.

5 Simulations

To evaluate the performance of the new estimators in small samples, we performed some simulations. We used the following designs:

- *Design I:* $Y = X + \epsilon$,
  where $X \sim N(0,1)$ and $\epsilon \sim N(0, \frac{1}{4})$.
- *Design II:* $Y = X + \epsilon$,
  where $X \sim N(0,1)$ and $\epsilon \sim N(0,1)$.
- *Design III:* $Y = \frac{4}{3^{3/4}} X^{3/4} \epsilon^{1/4}$,
  where $X \sim N(0,6)$ and $\epsilon \sim N(0,6)$.
- *Design IV:* $Y = \frac{3^3}{4^4} X^4 \, (-\varepsilon)^{-3}$
  where $X \sim N(6,1)$ and $\epsilon \sim N(-6,1)$.

The first design was chosen to evaluate how badly the estimator may perform, relative to the best estimator that one can use when the function is additively separable in ε and its parametric form is known. Also, since the function belongs to both $M1$ and $M2$, it allows one to evaluate the effect of the two normalizations. This design was estimated using the first normalization with $\overline{x} = \overline{\varepsilon} = 1$, $\alpha = 2$, and using the second normalization

QUAERERE VERUM
Universidad de San Andrés

with $\overline{x} = 0$. The second design is identical to the first, except that the variance of ε is four times as large as that in the first design. The data simulated from this model was estimated using the second normalization with $\overline{x} = 1$. Design III is the cost function of a Cobb-Douglas production function of the form $n(x_1, x_2) = x_1^{3/4} x_2^{1/4}$, when X is the price of x_1 and ε is the price of x_2. This design was used with the fist normalization with $\overline{x} = \overline{\varepsilon} = 6$ and $\alpha = 24/(3)^{3/4}$. Note that in this design, the function m is close to being linear in X, but is not additive separable in ε. Design IV is the profit function generated from a production function of the form $p(z) = z^a$ where $a = .75$, X is the price of the output, and $-\varepsilon$ is the price of the input z. We write this function in terms of $-\varepsilon$ to transform it so that it is strictly increasing in ε. Alternatively, we could have calculated the estimators under the restriction that m is strictly decreasing in ε. This would have only modified the estimator for $\widehat{F}_\varepsilon(e)$. Instead of deriving $\widehat{F}_\varepsilon(e)$ from the value of $\widehat{F}_{Y|X=x}(y)$ at a particular y and x, we would have derived $\widehat{F}_\varepsilon(e)$ from $1 - \widehat{F}_{Y|X=x}(y)$ at the same particular y and x. The expression for $\widehat{m}$ would have been the same as for the strictly increasing case. We used this design with $\overline{x} = \overline{\varepsilon} = 6$ and $\alpha = 6 \cdot 3^3/4^4$. The normal distributions, which were chosen for X and ε in these designs, violate Assumption S.2, but, since we are dealing with a finite set of data, we could obtain the same results if we specified the distributions of X and ε so that they are equal to the chosen distributions only on a large enough compact set.

For each design, we run 500 simulations of 500 observations each. The estimators of the joint pdf and cdf of (Y, X) were obtained using a multiplicative Gaussian kernel. The bandwidths were chosen to roughly minimize the integrated squared error of $\widehat{f}_{Y,X} : \int (\widehat{f}_{Y,X}(y,x) - f_{Y,X}(y,x))^2 dydx$. The following table specifies the bandwidth sizes that were used for each design:

	σ_Y	σ_X
Design I	0.2156	0.1850
Design II	0.4031	0.2928
Design III	0.0596	0.2619
Design IV	0.2017	0.1302

The results obtained for Design I, Normalization I (where m is specified to be in $M1$) and for Design I, Normalization II (where m is specified to be in $M2$) are presented below, together with the results obtained when the function m is estimated by Least Squares. The latter estimator is denoted by $\widetilde{m}_{LS}$. The estimators are evaluated at points where x equals the 5th, 25th, 50th, 75th, and 95th quantile of the distribution of X. The values of ε were chosen similarly. For each point, the tables show the bias, variance, mean square error, and asymptotic variance of the estimator.

Graphs corresponding to Design I, Normalization I are presented in Pages DI/NIa and DI/NIb. The first graph in page DI/NIa plots $\widehat{m}$ for one simulation. The second graph plots the true distribution F_ε (in a solid line) and $\widehat{F}_\varepsilon$ for one simulation. The last two graphs plot $\widehat{m}(\cdot,0)$ and $\widehat{m}(1,\cdot)$ (in the - - lines) versus $m(\cdot,0)$ and $m(1,\cdot)$ (in the solid lines). In page DI/NIb, the first graphs plots the average of $\widehat{m}$ over the 500 simulations. The other graphs plot the mean (in a $-\cdot-$ line), and the 5th and 95th percentiles (in the $\cdots\cdot$ lines) of $\widehat{F}_\varepsilon$, $\widehat{m}(\cdot,0)$ and $\widehat{m}(1,\cdot)$, together with the true values of F_ε, $m(\cdot,0)$ and $m(1,\cdot)$ (in the solid lines).

Design I / Normalization I

(x,e)	m	$Bias(\widehat{m})$	$Var(\widehat{m})$	$MSE(\widehat{m})$	$AVar(\widehat{m})$
(-1.6449,-0.8224)	-2.467280	0.035774	0.028337	0.029617	0.044973
(-1.6449, 0.8224)	-0.822427	0.078586	0.024113	0.030289	0.044973
(1.6449,-0.8224)	0.822427	-0.070478	0.024894	0.029861	0.044973
(1.6449, 0.8224)	2.467280	-0.020636	0.025500	0.025926	0.044973
(-0.6745,-0.3372)	-1.011735	0.006765	0.004384	0.004429	0.008209
(-0.6745, 0.3372)	-0.337245	0.034004	0.007438	0.008595	0.008209
(0.6745,-0.3372)	0.337245	-0.036290	0.006890	0.008207	0.008209
(0.6745, 0.3372)	1.011735	-0.010081	0.004487	0.004589	0.008209
(0.0000, 0.0000)	0.000000	0.000000	0.000000	0.000000	0.006003

e	F	$Bias(\widehat{F})$	$Var(\widehat{F})$	$MSE(\widehat{F})$	$AVar(\widehat{F})$
-0.8224	0.05	0.020432	0.000400	0.000818	0.000509
-0.3372	0.25	0.023079	0.001056	0.001589	0.001517
0.0000	0.50	-0.002542	0.001332	0.001338	0.001911
0.3372	0.75	-0.022267	0.001041	0.001537	0.001517
0.8224	0.95	-0.019337	0.000370	0.000744	0.000509

Design I / Normalization II

(x,e)	m	$Bias(\widehat{m})$	$Var(\widehat{m})$	$MSE(\widehat{m})$	$AVar(\widehat{m})$
(-1.6449,-0.8224)	-2.467280	0.060026	0.027209	0.030812	0.041541
(-1.6449, 0.8224)	-0.822427	0.044748	0.021093	0.023095	0.041541
(1.6449,-0.8224)	0.822427	-0.045390	0.020585	0.022645	0.041541
(1.6449, 0.8224)	2.467280	-0.053573	0.024551	0.027421	0.041541
(-0.6745,-0.3372)	-1.011735	0.015474	0.006338	0.006578	0.008003
(-0.6745, 0.3372)	-0.337245	0.018240	0.007030	0.007363	0.008003
(0.6745,-0.3372)	0.337245	-0.028173	0.006022	0.006815	0.008003
(0.6745, 0.3372)	1.011735	-0.027071	0.006619	0.007352	0.008003
(0.0000, 0.0000)	0.000000	0.000000	0.000000	0.000000	0.006003

Universidad de SanAndrés

e	F	$Bias(\widehat{F})$	$Var(\widehat{F})$	$MSE(\widehat{F})$	$AVar(\widehat{F})$
-0.8224	0.05	0.026224	0.000304	0.000991	0.000363
-0.3372	0.25	0.027458	0.001032	0.001785	0.001433
0.0000	0.50	-0.002542	0.001332	0.001338	0.001911
0.3372	0.75	-0.031512	0.001093	0.002086	0.001433
0.8224	0.95	-0.026939	0.000315	0.001041	0.000363

Design I / Least Squares

(x,e)	m	$Bias(\widetilde{m}_{LS})$	$Var(\widetilde{m}_{LS})$	$MSE(\widetilde{m}_{LS})$
(-1.6449,-0.8224)	-2.467280	-0.004959	0.001823	0.001848
(-1.6449, 0.8224)	-0.822427	-0.004959	0.001823	0.001848
(1.6449,-0.8224)	0.822427	0.004884	0.002029	0.002053
(1.6449, 0.8224)	2.467280	0.004884	0.002029	0.002053
(-0.6745,-0.3372)	-1.011735	-0.002056	0.000716	0.000720
(-0.6745, 0.3372)	-0.337245	-0.002056	0.000716	0.000720
(0.6745,-0.3372)	0.337245	-0.002056	0.000716	0.000720
(0.6745, 0.3372)	1.011735	0.001980	0.000801	0.000805
(0.0000, 0.0000)	0.000000	-0.000038	0.000522	0.000522

In general, we do not see a big difference between the results from the two normalizations. This is probably because, except for the point $(0,0)$ the chosen points are sufficiently away from the points at which the value of the function m is known. Comparison with the Least Squares (LS) estimator shows that the MSE of the new estimator may be up to 16 times larger than that of the LS estimator, specially at points where the pdf of both X and ε is very small. For points where these pdf's are larger, the MSE of the new estimator is around 6 times larger than that of the LS estimator.

To evaluate the effect that the variance of ε has on the new estimators, we can compare the table for Design I / Normalization II with the results of the table below, for which the simulated data was generated with a variance of ε four times larger than that in Design I. We can see that this has the effect of increasing the variance of the estimator by, roughly, a factor of 2.5. Hence, the new estimator seems to be much less sensitive to the variance of ε than the Nadaraya-Watson or the LS estimator, whose asymptotic variances increase linearly in the variance of ε. The graphs corresponding to this design are presented in pages DII/NIIa and DII/NIIb.

Design II / Normalization II

(x,e)	m	$Bias(\widehat{m})$	$Var(\widehat{m})$	$MSE(\widehat{m})$	$AVar(\widehat{m})$
(-1.6449,-1.6449)	-3.289707	0.134679	0.066571	0.084710	0.105033
(-1.6449, 1.6449)	0.000000	0.111552	0.051961	0.064405	0.105033
(1.6449,-1.6449)	0.000000	0.119248	0.051053	0.065273	0.105033
(1.6449, 1.6449)	-3.289707	-0.128687	0.063948	0.080509	0.105033
(-0.6745,-0.6745)	-1.348980	0.043760	0.013495	0.015410	0.020226
(-0.6745, 0.6745)	0.000000	0.045323	0.014521	0.016575	0.020226
(0.6745,-0.6745)	0.000000	-0.057992	0.012395	0.015758	0.020226
(0.6745, 0.6745)	1.348980	-0.056681	0.013435	0.016647	0.020226
(0.0000, 0.0000)	0.000000	0.000000	0.000000	0.000000	0.015174

e	F	$Bias(\widehat{F})$	$Var(\widehat{F})$	$MSE(\widehat{F})$	$AVar(\widehat{F})$
-1.6449	0.05	0.020293	0.000195	0.000607	0.000229
-0.6745	0.25	0.022256	0.000676	0.001172	0.000906
0.0000	0.50	-0.001386	0.000909	0.000911	0.001207
0.6745	0.75	-0.024429	0.000717	0.001313	0.000906
1.6449	0.95	-0.020745	0.000194	0.000625	0.000229

Universidad de San Andrés

The superiority of the LS estimator gets reversed when the function m is nonlinear and nonadditive in ε. We present below the results for the cost function example (Design III) and the profit function example (Design IV), together with the corresponding results for the LS estimators. For Design III, the LS estimator, $\widetilde{m}_{LS}$, at any point (x,e) is $\widetilde{m}_{LS}(x,e) = \widehat{\beta}_0 + \widehat{\beta}_1 x + e - E\varepsilon$, where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the LS estimators for the coefficients of a linear function in X. For Design IV, we present two LS estimators for $m(x,e)$, $\widetilde{m}^{+}_{LS}(x,e) = \widehat{\beta}_0 + \widehat{\beta}_1 x + e - E\varepsilon$, and $\widetilde{m}^{-}_{LS}(x,e) = \widehat{\beta}_0 + \widehat{\beta}_1 x - e + E\varepsilon$. While the former is the obvious LS estimator for $m(x,e)$, the latter yields a smaller bias, probably because the true function is strictly decreasing in e, so we present the results for both. The points at which the estimators were evaluated were the 2.5th, 27.5th, 50th, 77.5th, and 92.5th quantiles of the distribution of X and of the distribution of ε. We did not use the same quantiles as in the previous designs to avoid points along the 45 degree line, where the value of

m is known. The graphs for these designs are presented in pages DIII/NIIa and DIII/NIIb for Design III, and DIV/NIIa and DIV/NIIb for design IV.

Design III / Normalization II

x	e	m	$Bias(\widehat{m})$	$Var(\widehat{m})$	$MSE(\widehat{m})$
4.0400	4.5605	7.307357	-0.013097	0.043489	0.043660
4.0400	7.4395.	8.258371	0.145575	0.022846	0.044038
7.9600	4.5605	12.152195	-0.014446	0.117609	0.117817
7.9600	7.4395	13.733738	-0.031865	0.062351	0.063367
5.4022	5.2446	9.409731	0.002017	0.003799	0.003804
5.4022	6.7554	10.024494	0.045864	0.005687	0.007791
6.5978	5.2446	10.931864	-0.014342	0.010321	0.010526
6.5978	6.7554	11.646071	0.003830	0.003186	0.003201
6.0000	6.0000	10.528592	0.000000	0.000000	0.000000

e	F	$Bias(\widehat{F})$	$Var(\widehat{F})$	$MSE(\widehat{F})$
4.5605	0.025	0.029452	0.001740	0.002608
5.2446	0.275	0.034205	0.001712	0.002882
6.0000	0.500	0.003859	0.001769	0.001784
6.7554	0.725	0.017013	0.001404	0.001693
7.4395	0.975	-0.021691	0.001676	0.002146

Design III / Least Squares

x	e	m	$Bias(\widetilde{m}_{LS})$	$Var(\widetilde{m}_{LS})$	$MSE(\widetilde{m}_{LS})$
4.0400	4.5605	7.307357	-0.863639	0.001667	0.747538
4.0400	7.4395.	8.258371	1.064411	0.001667	1.134637
7.9600	4.5605	12.152195	-0.529689	0.002733	0.283304
7.9600	7.4395	13.733738	0.767830	0.002733	0.592296
5.4022	5.2446	9.409731	-0.482230	0.000435	0.232981
5.4022	6.7554	10.024494	0.413838	0.000435	0.171697
6.5978	5.2446	10.931864	-0.424908	0.000761	0.181308
6.5978	6.7554	11.646071	0.371714	0.000761	0.138932
6.0000	6.0000	10.528592	-0.055948	0.000434	0.003564

Universidad de San Andrés

Design IV / Normalization II

x	e	m	$Bias(\widehat{m})$	$Var(\widehat{m})$	$MSE(\widehat{m})$
4.0400	-7.4395	0.068238	-0.021932	0.000302	0.000783
4.0400	-4.5605	0.296235	0.030986	0.002574	0.003534
7.9600	-7.4395	1.028328	0.009417	0.017004	0.017093
7.9600	-4.5605	4.464161	-0.607415	1.013710	1.382663
5.4022	-6.7554	0.291382	0.013688	0.000528	0.000716
5.4022	-5.2446	0.622710	-0.005127	0.000388	0.000414
6.5978	-6.7554	0.648267	0.001158	0.000286	0.000287
6.5978	-5.2446	1.385404	0.096138	0.012280	0.021523
6.0000	-6.0000	0.632813	0.000000	0.000000	0.000000

e	F	$Bias(\widehat{F})$	$Var(\widehat{F})$	$MSE(\widehat{F})$
-7.4395	0.025	0.031442	0.000982	0.001970
-6.7554	0.275	0.032348	0.001119	0.002165
-6.0000	0.500	-0.006567	0.001260	0.001303
-5.2446	0.725	-0.039862	0.001179	0.002768
-4.5605	0.925	-0.037608	0.000990	0.002405

Universidad de SanAndrés

Design IV / Least Squares

x	e	m	$Bias(\widetilde{m}^{+}_{LS})$	$Var(\widetilde{m}^{+}_{LS})$	$MSE(\widetilde{m}^{+}_{LS})$
4.0400	-7.4395	0.068238	1.172196	0.007153	1.381196
4.0400	-4.5605	0.296235	-1.934864	0.007153	3.750849
7.9600	-7.4395	1.028328	2.412921	0.019719	5.841907
7.9600	-4.5605	4.464161	-3.901975	0.019719	15.245128
5.4022	-6.7554	0.291382	1.029735	0.000748	1.061102
5.4022	-5.2446	0.622710	-0.812423	0.000748	0.660779
6.5978	-6.7554	0.648267	1.344066	0.004581	1.811095
6.5978	-5.2446	1.385404	-0.903901	0.004581	0.821618
6.0000	-6.0000	0.632813	0.268498	0.001560	0.073651

x	e	m	$Bias(\widetilde{m}_{LS}^{-})$	$Var(\widetilde{m}_{LS}^{-})$	$MSE(\widetilde{m}_{LS}^{-})$
4.0400	-7.4395	0.068238	-1.706867	0.007153	2.920547
4.0400	-4.5605	0.296235	0.944199	0.007153	0.898665
7.9600	-7.4395	1.028328	-0.466142	0.019719	0.237007
7.9600	-4.5605	4.464161	-1.022912	0.019719	1.066068
5.4022	-6.7554	0.291382	-0.481095	0.000748	0.232201
5.4022	-5.2446	0.622710	0.698407	0.000748	0.488521
6.5978	-6.7554	0.648267	-0.166764	0.004581	0.032391
6.5978	-5.2446	1.385404	0.606929	0.004581	0.372943
6.0000	-6.0000	0.632813	0.268498	0.001560	0.073651

In these designs, the outperformance of the new estimator over that of the LS estimator is quite large, even at points where the values of the pdf's of X and ε are small. For example, in Design III, at the point $(x, \varepsilon) = (4.0400,$ $7.4395)$ the MSE of the LS estimator is 25 times larger than the MSE of the new estimator. At the point $(x, \varepsilon) = (6.5978,\ 5.2446)$ the ratio of the MSE's is around 17. The difference is more striking in Design IV, where at the point $(x, \varepsilon) = (4.0400,\ 7.4395)$ the ratio of the MSE's is around 1700, and at (x, ε) $= (5.4022,\ -6.7554)$, the ratio is around 300.

Universidad de SanAndrés

6 Summary

We have presented an estimator for a model in which the value of the dependent variable is determined by a nonparametric function that is nonadditive in an unobservable exogenous variables. The unobservable variable is assumed to be distributed, with an unknown distribution, independently of the observable exogenous variables. Under some normalizations, estimators for both, the distribution of the unobservable random variable and the nonparametric function are derived and are shown to be consistent and asymptotically normal. Both estimators converge at the same rate as the Nadaraya-Watson estimator of the conditional expectation of the dependent variable given the observed exogenous variables. When the nonparametric function is known to be homogenous of degree one, or additively separable into an homogenous of degree one function, the only normalization needed is the specification of the value of the function, or functions in the additively separable case, at one point. The estimators were defined as nonlinear functionals of a kernel estimator for the distribution of the observable variables. To derive their asymptotic distributions, we first linearized the functionals, by calculating their Hadamard-derivatives, and then applied the Delta method developed by Ait-Sahalia (1994).

The results of some simulations indicate that the method may outperform estimators that require specifying a parametric structure for the function to be estimated, when the specified structure is incorrect. The extent of the outperformance seems to depend on the degree of the misspecification. Since one can rarely find a parametric specification that would perfectly fit the true function, there seems to be a benefit to using the new estimators. Our simulation results indicate that for large data sets and values of the exogenous variables at which their pdf's are not too small, the increase in variance that is due to the weak structure may be reasonable.

QUAERERE VERUM
Universidad de
SanAndrés

7 Appendix

Proof of Theorem 2: Let F denote the joint cdf of (Y, X). For any x in A and y in R, define the functional $\Delta(F)$ by $\Delta(F) = F_{Y|X=x}(y)$. Let $\|H\|$ denote the sum of the L^2 norms of all the derivatives of H up to the order $2K+1$. Then, for any H such that H is constant outside a compact set and $\|H\|$ is sufficiently small, we have that, $|h(x)| \leq \|H\|$, $\left|\int_{-\infty}^{y} h(s,x)ds\right| \leq \|H\|$, and $|f(x) + h(x)| \geq a\,|f(x)|$ for some $a\,(0 < a < \infty)$ where $h(s,z) = \partial^{K+1}H(s,z)/\partial s\partial z_1 \cdots \partial z_K$ and $h(x) = \int_{-\infty}^{\infty} h(s,x)ds$. Then,

$$(1)\ \Delta(F+H) - \Delta(F) = (F+H)_{Y|X=x}(y) - F_{Y|X=x}(y)$$

$$= D\Delta(F,H) + R\Delta(F,H), \text{ where}$$

$$D\Delta(F,H) = \frac{\int_{-\infty}^{y} h(s,x)ds - h(x)\ F_{Y|X=x}(y)}{f(x)},$$

$$R\Delta(F,H) = \left[\frac{\int_{-\infty}^{y} h(s,x)ds - h(x)\ F_{Y|X=x}(y)}{f(x)}\right]\left[\frac{h(x)}{f(x)+h(x)}\right],$$

and for some $0 < c < \infty$,

$$(2)\ |D\Delta(F,H)| \leq \tfrac{c}{f(x)}\|H\|, \text{ and } |R\Delta(F,H)| \leq \tfrac{c}{f(x)^2}\|H\|^2.$$

Let $H = \left\|\widehat{F} - F\right\|$. It follows from (1) and (2) that $\left|\widehat{F}_{Y|X=x}(y) - F_{Y|X=x}(y)\right| \leq \frac{c}{f(x)}\left\|\widehat{F} - F\right\| + \frac{c}{f(x)^2}\left\|\widehat{F} - F\right\|^2$. Under Assumptions S.1-S.4, $\left\|\widehat{F} - F\right\| \to 0$ in probability (see Ait-Sahalia (1994). Hence, for any $x \in X$ and $y \in R$, $\sup_{y\in R}\left|\widehat{F}_{Y|X=x}(y) - F_{Y|X=x}(y)\right| \to 0$ in probability. By the definition of $\widehat{F}_{\varepsilon}$, this implies that $\sup_{e\in R}\left|\widehat{F}_{\varepsilon}(e) - F_{\varepsilon}(e)\right| \to 0$ in probability.

Next, by (1) and (2), Δ is $L(2, 2K+1)-$ Hadamard differentiable at F. $D\Delta(F,H) = \frac{\int_{-\infty}^{y} h(s,x)ds\ -\ h(x)\ F_{Y|X=x}(y)}{f(x)} = \int\int\left[\frac{1[s\leq y] - F_{Y|X=x}(y)}{f(x)}\right] 1_{(s,x)}(s,z)\ h(s,z)\ ds\ dz$, where $1[\cdot] = 1$ if $[\cdot]$ is correct and it equals 0 otherwise; $1_{(t,x)}(s,z) = 1$ if $(t,z) = (s,x)$ and it equals 0 otherwise. It then follows by the Delta method of Ait-Sahalia (1994) that for any $x \in A$ and $y \in R$,

$\sqrt{N}\sigma^{(K/2)}\left(\widehat{F}_{Y|X=x}(y) - F_{Y|X=x}(y)\right) \to N(0, V_F)$ in distribution, where

$$V_F = \left\{\int\left(\int K(s,z)ds\right)^2 dz\right\}\left\{(1/f(x)^2)\int\left(1\left[s \le y\right] - F_{Y|X=x}(y)\right)^2 f(s,x)ds\right\}$$
$$= \left\{\int\left(\int K(s,x)\, ds\right)^2 ds\right\}\left((1/f(x)^2)\right)\left[F_{Y|X=x}(y)\left(1 - F_{Y|X=x}(y)\right)\right].$$

Substituting the appropriate values of x and y, for which $\widehat{F}_\varepsilon(e) = \widehat{F}_{Y|X=x}(y)$, it follows that

$$\sqrt{N}\sigma^{(K/2)}\left(\widehat{F}_\varepsilon(e) - F_\varepsilon(e)\right) \to N\left(0, V_F\right),$$

where $V_F = \left\{\int\left(\int K(s,z)\, ds\right)^2 dz\right\}\ \left[F_\varepsilon(e)\ (1 - F_\varepsilon(e))\right]\ L$ and where $L = 1/f((e/\overline{\varepsilon})\overline{x})$ if $m \in M1$ and $L = 1/f(\overline{x})$ if $m \in M2$.

Proof of Theorem 3: Suppose first that $m \in M2$. Define the functional Φ by $\Phi(F) = F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)$. Then, $\widehat{m}(x,e) = \Phi(\widehat{F})$. If more than one value satisfies the definition of $\widehat{m}(x,e)$, then, let $\widehat{m}(x,e)$ be an arbitrary element within the set of all those values.

Let H be such that H vanishes outside a compact set and $\|H\| \le f(x)/2$. By $\|H\|$ we mean the sum of the L^2 norms of all derivatives of H up to the order $2K+2$. As in the proof of the previous theorem, we can show that for all x there exists $c < \infty$ such that

(0) $\sup_{y\in R}\left|(F+H)_{Y|X=x}(y) - F_{Y|X=x}(y)\right| \le \frac{c\|H\|}{f(x)}$.

Hence, since for all $y \in C$, $f(y,x) \ge \delta$, there exists $\rho > 0$ such that if $\|H\| \le \min\{\rho, f(x)/2)\}$, then $\left|(F+H)^{-1}_{Y|X=x}(F_{Y|X=\overline{x}}(e)) - F^{-1}_{Y|X=x}(F_{Y|X=\overline{x}}(e))\right| < \eta$. It follows that for all r that is between $(F+H)^{-1}_{Y|X=x}\ (F_{Y|X=\overline{x}}(e))$ and $F^{-1}_{Y|X=x}\ (F_{Y|X=\overline{x}}(e))$, r must belong to C. Moreover, if $\|H\| \le \min\{\rho,\ f(x)/2,\ \delta/2\}$, $|f(x) + h(x)| \ge \delta/2$. Hence, if $\|H\| \le \min\{\rho,\ f(x)/2,\ \delta/2\}$, then for all r that is between $(F+H)^{-1}_{Y|X=x}(F_{Y|X=\overline{x}}(e))$ and $F^{-1}_{Y|X=x}(F_{Y|X=\overline{x}}(e))$,

(1) $(f+h)_{Y|X=x}(r) > 0$.

Let then H be such that H vanishes outside a compact interval and $\|H\| \leq \min\{\rho,\ f(x)/2,\ \delta/2\}$. Using arguments similar to those used in Matzkin and Newey (1993), we will obtain a first order Taylor expansion for $\Phi(F+H)$. For this we note that

$$
\begin{aligned}
(2)\quad & \Phi(F+H) - \Phi(F) \\
&= (F+H)^{-1}_{y|x}\left((F+H)_{y|\overline{x}}(e)\right) - F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right) \\
&= \left\{(F+H)^{-1}_{y|x}\left((F+H)_{y|\overline{x}}(e)\right) - (F+H)^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right\} \\
&\quad + \left\{(F+H)^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right) - F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right\}
\end{aligned}
$$

To obtain an expression for the difference in the first bracket of (2), we note that by (1) and Taylor's Theorem,

$$
\begin{aligned}
& (F+H)^{-1}_{y|x}\left((F+H)_{y|\overline{x}}(e)\right) - (F+H)^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right) \\
&= \tfrac{\partial (F+H)^{-1}_{y|x}}{\partial y}\left(F_{y|\overline{x}}(e)\right)\left[(F+H)_{y|\overline{x}}(e) - F_{y|\overline{x}}(e)\right] + \mathrm{Re}\,m
\end{aligned}
$$

where, for some d $(0 < d < \infty)$, $\mathrm{Re}\,m \leq d\left|(F+H)_{y|\overline{x}}(e) - F_{y|\overline{x}}(e)\right|^2$. Hence,

$$
\begin{aligned}
(3)\quad & (F+H)^{-1}_{y|x}\left((F+H)_{y|\overline{x}}(e)\right) - (F+H)^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right) \\
&= \tfrac{\partial F^{-1}_{y|x}}{\partial y}\left(F_{y|\overline{x}}(e)\right)\left[(F+H)_{y|\overline{x}}(e) - F_{y|\overline{x}}(e)\right] + \mathrm{Re}\,m, \\
&+ \left[\tfrac{\partial (F+H)^{-1}_{y|x}}{\partial y}\left(F_{y|\overline{x}}(e)\right) - \tfrac{\partial F^{-1}_{y|x}}{\partial y}\left(F_{y|\overline{x}}(e)\right)\right]\left[(F+H)_{y|\overline{x}}(e) - F_{y|\overline{x}}(e)\right]. \\
&= \left[\tfrac{\partial F_{y|x}}{\partial y}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}\left[(F+H)_{y|\overline{x}}(e) - F_{y|\overline{x}}(e)\right] + \mathrm{Re}\,m \\
&+ \left\{\left[\tfrac{\partial (F+H)_{y|x}}{\partial y}\left((F+H)^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1} - \left[\tfrac{\partial F_{y|x}}{\partial y}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}\right\} \\
&\quad \cdot \left[(F+H)_{y|\overline{x}}(e) - F_{y|\overline{x}}(e)\right]
\end{aligned}
$$

Universidad de San Andrés
QUAERERE VERUM
UNIVERSIDAD DE SAN ANDRES
BIBLIOTECA

To obtain an expression for the difference in the second brackets of (2), we note that by (1) and the Mean Value Theorem, for any $t \in R$,

$$(F+H)_{y|x}\left((F+H)^{-1}_{y|x}(t)\right) - (F+H)_{y|x}\left(F^{-1}_{y|x}(t)\right)$$

$$= \frac{\partial (F+H)_{y|x}}{\partial y}(r^*)\ \left[(F+H)^{-1}_{y|x}(t) - F^{-1}_{y|x}(t)\right]$$

where r^* is between $(F+H)^{-1}_{y|x}(t)$ and $F^{-1}_{y|x}(t)$. Hence, since

$(F+H)_{y|x}\left((F+H)^{-1}_{y|x}(t)\right) = t = F_{y|x}\left(F^{-1}_{y|x}(t)\right)$, it follows that for any $t \in R$,

$$(4)\ (F+H)^{-1}_{y|x}(t) - F^{-1}_{y|x}(t) = \left[\frac{1}{(f+h)_{y|x}(r^*)}\right] \cdot \left[F_{y|x}\left(F^{-1}_{y|x}(t)\right) - (F+H)_{y|x}\left(F^{-1}_{y|x}(t)\right)\right],$$

where r^* is between $(F+H)^{-1}_{y|x}(t)$ and $F^{-1}_{y|x}(t)$. From (4) it follows that for any $t \in R$,

$$(5)\ (F+H)^{-1}_{y|x}(t) - F^{-1}_{y|x}(t) = \left[\frac{1}{f_{y|x}(F^{-1}_{y|x}(t))}\right] \cdot \left[F_{y|x}\left(F^{-1}_{y|x}(t)\right) - (F+H)_{y|x}\left(F^{-1}_{y|x}(t)\right)\right]$$
$$+ \left[\frac{1}{(f+h)_{y|x}(r^*)} - \frac{1}{f_{y|x}(F^{-1}_{y|x}(t))}\right] \cdot \left[F_{y|x}\left(F^{-1}_{y|x}(t)\right) - (F+H)_{y|x}\left(F^{-1}_{y|x}(t)\right)\right],$$

where r^* is between $(F+H)^{-1}_{y|x}(t)$ and $F^{-1}_{y|x}(t)$. Hence, by (2) and (3) and by (5) with $t = F_{y|\overline{x}}(e)$, it follows that

$$\Phi(F+H) - \Phi(F)$$

$$= \left[\frac{\partial F_{y|x}}{\partial y}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}\left[(F+H)_{y|\overline{x}}(e) - F_{y|\overline{x}}(e)\right] + \mathrm{Re}m$$

$$+ \left\{\left[\frac{\partial (F+H)_{y|x}}{\partial y}\left((F+H)^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1} - \left[\frac{\partial F_{y|x}}{\partial y}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}\right\}$$

$$\cdot \left[(F+H)_{y|\overline{x}}(e) - F_{y|\overline{x}}(e)\right].$$

$$+ \left[\frac{1}{f_{y|x}(F^{-1}_{y|x}(F_{y|\overline{x}}(e)))}\right] \cdot \left[F_{y|x}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right) - (F+H)_{y|x}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right)\right]$$

Universidad de SanAndrés

$$+\left[\frac{1}{(f+h)_{y|x}(r^*)}-\frac{1}{f_{y|x}(F_{y|x}^{-1}(F_{y|\overline{x}}(e)))}\right]\cdot\left[F_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)-(F+H)_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]$$

where r^* is between $(F+H)_{y|x}^{-1}(F_{y|\overline{x}}(e))$ and $F_{y|x}^{-1}(F_{y|\overline{x}}(e))$.

Let

(6) $A(F,H)=\left[\frac{\partial F_{y|x}}{\partial y}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}\cdot$

$$\left\{\left[(F+H)_{y|\overline{x}}(e)-F_{y|\overline{x}}(e)\right]+\left[F_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)-(F+H)_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]\right\},$$

and

(7) $D(F,H)=$

$$=\left\{\left[\frac{\partial (F+H)_{y|x}}{\partial y}\left((F+H)^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}-\left[\frac{\partial F_{y|x}}{\partial y}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}\right\}$$

$$\cdot\left[(F+H)_{y|\overline{x}}(e)-F_{y|\overline{x}}(e)\right]$$

$$+\ \mathrm{Re}\,m$$

$$+\left[\frac{1}{(f+h)_{y|x}(r^*)}-\frac{1}{f_{y|x}(F_{y|x}^{-1}(F_{y|\overline{x}}(e)))}\right]\cdot\left[F_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)-(F+H)_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right].$$

where r^* is between $(F+H)_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)$ and $F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)$. Then, $\Phi(F+H)-\Phi(F)=A(F,H)+D(F,H)$. Moreover, $A(F,H)=B(F,H)+C(F,H)$ where

(8) $B(F,H)=\left[\frac{\partial F_{y|x}}{\partial y}(r)\right]^{-1}\cdot$

$$\left\{\frac{f(\overline{x})\int_{-\infty}^{e}h(s,\overline{x})ds-h(\overline{x})\int_{-\infty}^{e}f(s,\overline{x})ds}{f(\overline{x})^2}-\frac{f(x)\int_{-\infty}^{r}h(s,x)ds-h(x)\int_{-\infty}^{r}f(s,x)ds}{f(x)^2}\right\},$$

and where, for $r=F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)$,

(9) $C(F,H)$

Universidad de San Andrés
QUAERERE VERUM

$$= \left[\frac{\partial F_{y|x}}{\partial y}(r)\right]^{-1} \cdot \left[f(\overline{x}) \int_{-\infty}^{e} h(s,\overline{x})ds \; - \; h(\overline{x}) \int_{-\infty}^{e} f(s,\overline{x})ds\right] \left[\frac{1}{f(\overline{x})^2 + h(\overline{x})f(\overline{x})} - \frac{1}{f(\overline{x})^2}\right]$$

$$- \left[\frac{\partial F_{y|x}}{\partial y}(r)\right]^{-1} \cdot \left[f(x) \int_{-\infty}^{r} h(s,x)ds \; - \; h(x) \int_{-\infty}^{r} f(s,x)ds\right] \left[\frac{1}{f(x)^2 + h(x)f(x)} - \frac{1}{f(x)^2}\right],$$

Let $D\Phi(F,H) = B(F,H)$ and $R\Phi(F,H) = C(F,H) + D(F,H)$. Then,

(10) $\Phi(F+H) - \Phi(F) = D\Phi(F,H) + R\Phi(F,H)$.

Moreover, since $m(x,e) \in C$, $f(x) > 0$, and $f(\overline{x}) > 0$, it follows from (8) that for some V_1 $(0 < V_1 < \infty)$,

(11) $|D\Phi(F,H)| \leq V \; \|H\|$

and it follows from (9) that for some V_2 $(0 < V_2 < \infty)$,

(13) $|C(F,H)| \leq V_2 \; \|H\|^2$.

We next show that for some V_3 $(0 < V_3 < \infty)$,

(14) $|D(F,H)| \leq V_3 \; \|H\|^2$

For this we note that, letting $t = F_{y|\overline{x}}(e)$, in (4),

(15) $(F+H)^{-1}_{y|x}(F_{y|\overline{x}}(e)) - \; F^{-1}_{y|x}(F_{y|\overline{x}}(e))$

$$= \left[\frac{1}{(f+h)_{y|x}(r^*)}\right] \cdot \left[F_{y|x}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right) - (F+H)_{y|x}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right)\right]$$

where r^* is between $(F+H)^{-1}_{y|x}(F_{y|\overline{x}}(e))$ and $F^{-1}_{y|x}(F_{y|\overline{x}}(e))$. Also, for some n such that $0 < n < \infty$, all $t_1, t_2 \in N(m(x,e),\eta)$ and all t between t_1 and t_2,

(16) $$\left|\frac{1}{(f+h)_{y|x}(t_1)} - \frac{1}{f_{y|x}(t_2)}\right| \leq \left(\frac{f(t_1,x)+f(x)}{f(t_1,x)^2}\right) n \, \|H\| + \left|\frac{(f(x)+h(x))\left(\frac{\partial f(\overline{t},x)}{\partial y} + \frac{\partial f(\overline{t},x)}{\partial y}\right)}{\left[f(\overline{t},x)+h(\overline{t},x)\right]^2}\right| |t_1 - t_2|.$$

This latter expression follows because

$$\left|\frac{1}{(f+h)_{y|x}(t_1)} - \frac{1}{f_{y|x}(t_2)}\right|$$
$$\leq \left|\frac{1}{(f+h)_{y|x}(t_1)} - \frac{1}{f_{y|x}(t_1)}\right| + \left|\frac{1}{f_{y|x}(t_1)} - \frac{1}{f_{y|x}(t_2)}\right|$$
$$\leq \left|\frac{f(x)+h(x)}{f(t_1,x)+h(t_1,x)} - \frac{f(x)}{f(t_1,x)}\right| + \left|\frac{\partial\left(\frac{1}{(f+h)_{y|x}(\bar{t})}\right)}{\partial y}\right| |t_1 - t_2|$$

where the last inequality follows by the Mean Value Theorem and $\bar{t}$ is between t_1 and t_2.

By (0), there exists T $(0 < T < \infty)$ such that

(17) $\left|(F+H)_{y|\overline{x}}(e) - F_{y|\overline{x}}(e)\right| \leq T \ \|H\|,$

(18) $\left|F_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right) - (F+H)_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right| \leq T \ \|H\|$ and

(19) $\operatorname{Re} m \leq T \ \|H\|^2$

Letting $t_1 = (F+H)_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)$ and $t_2 = F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)$ in (16), we get

$$\left|\left[\frac{\partial(F+H)_{y|x}}{\partial y}\left((F+H)^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1} - \left[\frac{\partial F_{y|x}}{\partial y}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}\right|$$
$$\leq \left|\frac{f((F+H)_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right),x)+f(x)}{f((F+H)_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right),x)^2}\right| n\|H\|$$
$$+\left|\frac{(f(x)+h(x))\left(\frac{\partial f(\bar{t},x)}{\partial y}+\frac{\partial f(\bar{t},x)}{\partial y}\right)}{\left[f(\bar{t},x)+h(\bar{t},x)\right]^2}\right| \left|(F+H)_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right) - F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right|$$
$$\leq \left|\frac{f((F+H)_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right),x)+f(x)}{f((F+H)_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right),x)^2}\right| n\|H\|$$
$$+\left|\frac{(f(x)+h(x))\left(\frac{\partial f(\bar{t},x)}{\partial y}+\frac{\partial f(\bar{t},x)}{\partial y}\right)}{\left[f(\bar{t},x)+h(\bar{t},x)\right]^2}\right| \cdot$$
$$\left|\frac{1}{(f+h)_{y|x}(r^*)}\right| \cdot \left|\left[F_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right) - (F+H)_{y|x}\left(F_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]\right|$$
$$\leq \left|\frac{f((F+H)_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right),x)+f(x)}{f((F+H)_{y|x}^{-1}\left(F_{y|\overline{x}}(e)\right),x)^2}\right| n\|H\|$$

QUAERERE VERUM
Universidad de San Andrés

$$+\left|\frac{(f(x)+h(x))\left(\frac{\partial f(\bar{t},x)}{\partial y}+\frac{\partial f(\bar{t},x)}{\partial y}\right)}{\left[f(\bar{t},x)+h(\bar{t},x)\right]^2}\right|\cdot\left|\frac{1}{(f+h)_{y|x}(r^*)}\right|\cdot T\left\|H\right\|,$$

where the second inequality follows by (15), the third inequality follows by (18), and where $\bar{t}$ and r^* are between $(F+H)^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)$ and $F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)$. Since

$\left|\frac{f((F+H)^{-1}_{y|x}(F_{y|\overline{x}}(e)),x)+f(x)}{f((F+H)^{-1}_{y|x}(F_{y|\overline{x}}(e)),x)^2}\right|$, $\left|\frac{(f(x)+h(x))\left(\frac{\partial f(\bar{t},x)}{\partial y}+\frac{\partial f(\bar{t},x)}{\partial y}\right)}{\left[f(\bar{t},x)+h(\bar{t},x)\right]^2}\right|$, and $\left|\frac{1}{(f+h)_{y|x}(r^*)}\right|$ are all bounded, there exists E $(0<E<\infty)$ such that

$$(20)\left|\left[\frac{\partial(F+H)_{y|x}}{\partial y}\left((F+H)^{-1}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}-\left[\frac{\partial F_{y|x}}{\partial y}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right)\right]^{-1}\right|\leq E\left\|H\right\|.$$

Letting $t_1=r^*$ and $t_2=F^{-1}_{y|x}(F_{y|\overline{x}}(e))$ in (16), it follows that

$$\left|\frac{1}{(f+h)_{y|x}(r^*)}-\frac{1}{f_{y|x}\left(F^{-1}_{y|x}(F_{y|\overline{x}}(e))\right)}\right|$$

$$\leq\left(\frac{f(r^*,x)+f(x)}{f(r^*,x)^2}\right)n\left\|H\right\|+\left|\frac{(f(x)+h(x))\left(\frac{\partial f(\tilde{t},x)}{\partial y}+\frac{\partial f(\tilde{t},x)}{\partial y}\right)}{\left[f(\tilde{t},x)+h(\tilde{t},x)\right]^2}\right|\left|r^*-F^{-1}_{y|x}(F_{y|\overline{x}}(e))\right|$$

$$\leq\left(\frac{f(r^*,x)+f(x)}{f(r^*,x)^2}\right)n\left\|H\right\|+\left|\frac{(f(x)+h(x))\left(\frac{\partial f(\tilde{t},x)}{\partial y}+\frac{\partial f(\tilde{t},x)}{\partial y}\right)}{\left[f(\tilde{t},x)+h(\tilde{t},x)\right]^2}\right|\left|(F+H)^{-1}_{y|x}(F_{y|\overline{x}}(e))-F^{-1}_{y|x}(F_{y|\overline{x}}(e))\right|$$

$$\leq\left(\frac{f(r^*,x)+f(x)}{f(r^*,x)^2}\right)n\left\|H\right\|+\left|\frac{(f(x)+h(x))\left(\frac{\partial f(\tilde{t},x)}{\partial y}+\frac{\partial f(\tilde{t},x)}{\partial y}\right)}{\left[f(\tilde{t},x)+h(\tilde{t},x)\right]^2}\right|\cdot$$

$$\left|\frac{1}{(f+h)_{y|x}(r^{**})}\right|\cdot\left|F_{y|x}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right)-(F+H)_{y|x}\left(F^{-1}_{y|x}\left(F_{y|\overline{x}}(e)\right)\right)\right|$$

$$\leq\left(\frac{f(r^*,x)+f(x)}{f(r^*,x)^2}\right)n\left\|H\right\|+\left|\frac{(f(x)+h(x))\left(\frac{\partial f(\tilde{t},x)}{\partial y}+\frac{\partial f(\tilde{t},x)}{\partial y}\right)}{\left[f(\tilde{t},x)+h(\tilde{t},x)\right]^2}\right|\left|\frac{1}{(f+h)_{y|x}(r^{**})}\right|T\left\|H\right\|,$$

where the second inequality follows because r^* is between $(F+H)^{-1}_{y|x}(F_{y|\overline{x}}(e))$ and $F^{-1}_{y|x}(F_{y|\overline{x}}(e))$, the third inequality follows by (15), the fourth follows by (18), and where $\tilde{t}$ is between r^* and $F^{-1}_{y|x}(F_{y|\overline{x}}(e))$, and r^{**} is between $(F+H)^{-1}_{y|x}(F_{y|\overline{x}}(e))$ and $F^{-1}_{y|x}(F_{y|\overline{x}}(e))$. Since

$\left(\frac{f(r^*,x)+f(x)}{f(r^*,x)^2}\right)$, $\left|\frac{(f(x)+h(x))\left(\frac{\partial f(\widetilde{t},x)}{\partial y}+\frac{\partial f(\widetilde{t},x)}{\partial y}\right)}{\left[f(\widetilde{t},x)+h(\widetilde{t},x)\right]^2}\right|$, and $\left|\frac{1}{(f+h)_{y|x}(r^{**})}\right|$ are all bounded, it follows that for some Q such that $0 < Q < \infty$,

(21) $$\left|\frac{1}{(f+h)_{y|x}(r^*)} - \frac{1}{f_{y|x}\left(F^{-1}_{y|x}(F_{y|\overline{x}}(e))\right)}\right| \leq Q \left\| H \right\|.$$

Hence, (14) follows from (7), (20), (17), (19), and (21).

Let $H = \widehat{F} - F$. We have shown that

(22) $$\widehat{m}(x,e) - m(x,e) = \Phi(\widehat{F}) - \Phi(F)$$

$$= D\Phi(F, \widehat{F} - F) + R\Phi(F, \widehat{F} - F),$$

and for some C $(0 < C < \infty)$.

(23) $$\left|D\Phi(F, \widehat{F} - F)\right| \leq C \left\|\widehat{F} - F\right\| \quad \text{and} \quad \left|R\Phi(F, \widehat{F} - F)\right| \leq C \left\|\widehat{F} - F\right\|^2.$$

Under Assumptions S.1-S.4, $\left\|\widehat{F} - F\right\| \to 0$ in probability. Hence, by (22) and (23), it follows that

(24) $\widehat{m}(x,e) \to m(x,e)$ in probability.

Hence, the estimator of $m(x,e)$ is consistent

Next, by (22) and (23), Φ is $L(2, 2K+2)-$Hadamard differentiable. By (8),

$$D\Phi(F,H) = B(F,H)$$

$$= \frac{1}{f_{y|x}(m(x,e))} \int\int \frac{\left[1(s \leq e) - F_{y|\overline{x}}(e)\right]}{f(\overline{x})} 1_{s,\overline{x}}(s,z)\ h(s,z)\ ds\ dz -$$

$$- \frac{1}{f_{y|x}(m(x,e))} \int\int \frac{\left[1(s \leq m(x,e)) - F_{y|x}(m(x,e))\right]}{f(x)} 1_{s,x}(s,z)\ h(s,z)\ ds\ dz.$$

Hence, by Ait-Sahalia (1994), it follows that

$$\sqrt{N}\,\sigma_N^{K/2}\left(\hat{m}(x,e)-m(x,e)\right)=\sqrt{N}\,\sigma_N^{K/2}\left(\Phi(\hat{F})-\Phi(F)\right)\to N(0,V_m)$$

where $V_m=\left\{\int\int\left[\int K(s,z)ds\right]^2 dz\right\}\left[\frac{1}{f_{Y|X=x}(m(x,e))}\right]^2 L$ and

$$L=\int\left[\frac{1(s<e)}{f(\overline{x})}-\frac{F_{y|\overline{x}}(e)}{f(\overline{x})}\right]^2 f(s,\overline{x})\,ds\;+\;\int\left[\frac{1(s<m(x,e))}{f(x)}-\frac{F_{y|x}(e)}{f(x)}\right]^2 f(s,x)\,ds$$

$$=\frac{1}{f(\overline{x})}F_{y|\overline{x}}(e)(1-F_{y|\overline{x}}(e))+\frac{1}{f(x)}F_{y|x}(m(x,e))(1-F_{y|x}(m(x,e)))$$

$$=\left[\frac{1}{f(\overline{x})}+\frac{1}{f(x)}\right]F_\varepsilon(e)(1-F_\varepsilon(e))$$

Suppose next that $m\in M1$. By substituting $\overline{x}$ in the above argument by $(e/\overline{\varepsilon})\overline{x}$ we obtain that $\widehat{m}(x,e)$ is a consistent estimator of $m(x,e)$ and $\sqrt{N}\,\sigma_N^{K/2}\left(\hat{m}(x,e)-m(x,e)\right)$ converges in distribution to $N(0,V_m)$ where

$$V_m=\left\{\int\int\left[\int K(s,z)ds\right]^2 dz\right\}\left[\frac{1}{f_{Y|X=x}(m(x,e))^2}\right]\left[\frac{1}{f((e/\overline{\varepsilon})\overline{x})}+\frac{1}{f(x)}\right]F_\varepsilon(e)(1-F_\varepsilon(e)).$$

Proof of Theorem 4: As in the proof of Theorem 1, we first note that for all $e\in E$ and $x\in A$,

(1) $F_\varepsilon(e)=F_{Y|X=x}\left(m(x,e)\right).$

Suppose first that $m\in M3a$, we get from (1), letting $x_1=\overline{x}_1$ and $x_2=(e/\overline{\varepsilon})\overline{x}_2$, that $F_\varepsilon(e)=F_{Y|X=(\overline{x}_1,(e/\overline{\varepsilon})\overline{x}_2)}\left(m(x_1,x_2,e)\right)=F_{Y|X=(\overline{x}_1,(e/\overline{\varepsilon})\overline{x}_2)}(v_1(\overline{x}_1)+v_2((e/\overline{\varepsilon})\overline{x}_2,(e/\overline{\varepsilon})\overline{\varepsilon}))=F_{Y|X=(\overline{x}_1,(e/\overline{\varepsilon})\overline{x}_2)}\left(\gamma+(e/\overline{\varepsilon})\alpha\right)$. Hence, $F_e(\varepsilon)$ is identified from the conditional cdf of Y given X, when $x_1=\overline{x}_1$, $x_2=(e/\overline{\varepsilon})\overline{x}_2$, and $y=\gamma+(e/\overline{\varepsilon})\alpha$. Using this together with (1), it follows that for all (x_1,x_2,e), $F_{Y|X=(\overline{x}_1,(e/\overline{\varepsilon})\overline{x}_2)}(\gamma+(e/\overline{\varepsilon})\alpha)=F_{Y|X=x}\left(v_1(x_1)+v_2(x_2,e)\right)$. Letting $x_1=\overline{x}_1$ in this last expression, we get that $v_2(x_2,e)=F^{-1}_{Y|X=(\overline{x}_1,x_2)}F_{Y|X=(\overline{x}_1,(e/\overline{\varepsilon})\overline{x}_2)}(\gamma+(e/\overline{\varepsilon})\alpha)-\gamma$. Hence, v_2 is identified. Letting $x_2=\overline{x}_2$ and $e=\overline{\varepsilon}$ in that same expression, we get that $v_1(x_1)=F^{-1}_{Y|X=(x_1,\overline{x}_2)}\left(F_{Y|X=(\overline{x}_1,\overline{x}_2)}(\gamma+\overline{\varepsilon})\right)-\alpha$. Hence, v_1 is identified.

Universidad de San Andrés

Suppose next that $m \in M3b$. Then, letting $x_1 = \overline{x}_1$ and $x_2 = \overline{x}_2$, we get from (1) that $F_\varepsilon(e) = F_{Y|X=(\overline{x}_1,\overline{x}_2)}\left(m(x_1, x_2, e)\right) = F_{Y|X=(\overline{x}_1,\overline{x}_2)}\left(v_1(\overline{x}_1) + v_2(\overline{x}_2, e)\right) = F_{Y|X=(\overline{x}_1,\overline{x}_2)}(\gamma + e)$. Hence, F_ε is identified from the conditional cdf of Y given $X = (\overline{x}_1, \overline{x}_2)$, and for all (x_1, x_2, e), $F_{Y|X=(\overline{x}_1,\overline{x}_2)}(\gamma + e) = F_{Y|X=x}(v_1(x_1) + v_2(x_2, e))$. Letting $x_1 = \overline{x}_1$ in this last expression, we get that $v_2(x_2, \varepsilon) = F^{-1}_{Y|X=(\overline{x}_1,x_2)}(F_{Y|X=(\overline{x}_1,\overline{x}_2)}(\gamma + e) - \gamma)$. Hence, v_2 is identified. Letting $x_2 = \overline{x}_2$ in that same expression, we get that, for any e, $v_1(x_1) = F^{-1}_{Y|X=(x_1,\overline{x}_2)}(F_{Y|X=(\overline{x}_1,\overline{x}_2)}(\gamma + e)) - e$. Hence, v_1 is identified.

QUAERERE VERUM
Universidad de
SanAndrés

8 References

AIT-SAHALIA, Y. (1994) "The Delta and Bootstrap Methods for Nonlinear Functionals of Nonparametric Kernel Estimators Based on Dependent Multivariate Data," mimeo, Princeton University..

ALTONJI, J.G. and H. ICHIMURA (1997) "Estimating Derivatives in Nonseparable Models with Limited Dependent Variables", mimeo.

BROWN, D.J. and R.L. MATZKIN, (1996) "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," mimeo, Northwestern University.

BROWN, B.W. and M.B. WALKER (1989) "The Random Utility Hypothesis and Inference in Demand Systems", Econometrica, 57, 815-829.

______________ (1995) "Stochcastic Specification in Random Production Models of Cost-Minimizing Firms", Journal of Econometrics, 66, (1-2).

McELROY, M.B. (1981) "Duality and the Error Structure in Demand Systems," Discussion Paper #81-82, Economics Research Center/NORC.

McELROY, M.B. (1987) "Additive General Error Models for Production, Cost, and Derived Demand or Share Systems", Journal of Political Economy, 95, 737-757.

MATZKIN, R.L. (1994) "Restrictions of Economic Theory in Nonparametric Methods," in Handbook of Econometrics, Vol. IV, edited by R.F. Engel and D.L. McFadden.

MATZKIN, R.L. and W. NEWEY (1993) "Kernel Estimation of Nonparametric Limited Dependent Variable Models,", mimeo.

ROEHRIG, C.S. (1988) "Conditions for Identification in Nonparametric and Parametric Models", Econometrica, 56, 433-447.

QUAERERE VERUM
Universidad de
SanAndrés