NONPARAMETRIC ESTIMATION OF NONADDITIVE RANDOM FUNCTIONS

Rosa L. Matzkin*
Department of Economics
Northwestern University

April 2003

We present estimators for nonparametric functions that are nonadditive in unobservable random terms. The distribution of the unobservable random terms is assumed to be unknown. We show that when a nonadditive, nonparametric function is strictly monotone on an unobservable random term, and it satisfies some other properties that may be implied by economic theory, such as homogeneity of degree one or separability, the function and the distribution of the unobservable random term are identified. We also present convenient normalizations, to use when the properties of the functions are unknown. The estimators for the nonparametric function and for the distribution of the unobservable random terms are shown to be consistent and asymptotically normal. We extend the results to functions that depend on multivariate random terms. The results of a limited simulation study are presented.

KEYWORDS: nonparametric estimation, nonadditive random term, shape restrictions, conditional distributions, nonseparable models, conditional quantiles

*The research presented in this paper was partially supported by NSF; it was partially completed while I was visiting Universidad de San Andrés (Argentina). I am grateful to four referees and to Richard Blundell for their insightful comments, which helped generate a substantially improved paper. I have greatly benefitted from my interaction with Joseph Altonji, James Heckman, Daniel McFadden, Charles Manski, and Whitney Newey, the research assistance of Ulrich Doraszelski and Elie Tamer, and from the comments of José Sheinkman, Christopher Taber, and seminar participants at Hanyang University (Korea), McGill University, Michigan State University, New York University, Northwestern University, Universidad de San Andrés (Argentina), Universidad Di Tella (Argentina), University of Michigan, University of Wisconsin, the 1999 Summer Meeting of the Econometric Society, the 1999 Latin American Meeting of the Econometric Society, the Conference in Honor of Rolf Mantel (June, 2000), and the 2000 Midwest Econometrics Group Conference.

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 an 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, a utility value for some unobserved attributes, or some other relevant unobservable variable (see, for example, Heckman (1974), Heckman and Willis (1974), McFadden (1974), and Lancaster (1979)). 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 (McElroy (1981, 1987), Brown and Walker (1989, 1995), Lewbel (1996).) In general, unless one specifies very restrictive parametric structures for the functions in the economic model, the function by which the values of Y are determined from X and ε is nonadditive in ε.

Most nonparametric methods that are currently used to specify the relationship between a vector of observable exogenous variables, X, an unobservable term, and an 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 some other unobservable variables, the existent methods do not provide a way of studying this function, which has information about the important interactions between the observable and unobservable variables.

In this paper, we present a nonparametric method for estimating a nonparametric, not necessarily additive function of a vector of exogenous variables, X, and an unobservable vector of variables, ε. The value of a dependent variable, Y, is assumed to be determined by this nonparametric function. The distribution of ε is not parametrically specified and it is also estimated.

We first consider the model $Y = m(X,\varepsilon)$, where ε is a random variable, m is strictly increasing in ε, and both the function m and the distribution of ε are unknown. We characterize the set of functions that are observationally equivalent to m, when ε is independent of X, and provide three different specifications for the function m, which allow one to identify the distribution of ε and the function m. The first specification is just a convenient normalization. It specifies the value of $m(x, \varepsilon)$ at a particular value of x, or a subvector of x. The second specification imposes an homogeneity of degree one condition, along a given ray, on some coordinates of X and ε. This condition, together with the specification of the value of m at only one point of the ray, is shown to be sufficient to identify the distribution of ε and the function m. This second specification is particularly useful, for example, when the function m is either a cost or profit function, since economic theory implies that these functions are homogenous of degree one in some or all of their arguments. The third specification can be seen as a nonparametric generalization of semiparametric transformation models where neither the transformation function nor the distribution of the unobservable random term are parametrically specified. Instead of specifying that $Y = \Lambda(\beta' X + \varepsilon)$, where Λ is a strictly increasing, unknown function, and where both, the absolute value of one of the coordinates of β and the value of Λ at one point are given (see, for example, Horowitz (1996)), we specify that $Y = s(X_1, \varepsilon - X_2)$, for some unknown function s, which is strictly increasing in the last coordinate and whose value is given at one point. In the latter specification, $X = (X_1, X_2)$ and $X_2 \in R$.

For each of the three specifications, we extend the identification results to the case where ε is independent of only some coordinates of X, conditional on the other coordinates. A special case of this is, of course, when ε is independent of X, conditional on some vector Z, which is not an argument of m, since we can consider functions m that are constant as Z varies. We also extend the results to the case where the variable Y depends on a vector of unobservable variables, $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_K)$.

For each of the specifications and assumptions on the distribution of ε, we show that the estimator for the distribution of ε at a particular value, e, is obtained from an estimator for the conditional distribution function of Y given X, evaluated at particular values of X and Y. The estimator for the value of the function m at a particular vector, (x, e), is defined as an estimator for a quantile of the conditional distribution function of Y given $X = x$, where the quantile is the value of the estimator for the distribution of

ε, at $\varepsilon = e$. The estimator for the quantile is based on the quantile estimator of Nadaraya (1964) (see also Azzalini (1981)).

The estimators for the distribution of ε and for the function m are shown to be consistent and asymptotically normal. Each of these estimators is a nonlinear functional of a kernel estimator for the density function of (Y, X). We derive their asymptotic distributions using a Delta method of the type developed in Ait-Sahalia (1994) and Newey (1994). This method proceeds by first obtaining a first order Taylor expansion of each nonlinear functional around its true value, and then deriving the asymptotic distribution of the linear part of the expansion.

Some other papers that considered nonparametric models where the random terms do not enter in an additive form are Roehrig (1988), Brown and Matzkin (1996), Olley and Pakes (1996), Altonji and Ichimura (1997), Briesch, Chintagunta and Matzkin (1997), Heckman and Vytlacil (1999, 2001), Vytlacil (2000), Blundell and Powell (2000), and Altonji and Matzkin (2001). 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 form studied in Roehrig (1988). Olley and Pakes (1996) consider a dynamic model where a firm's investment at time t depends in a nonadditive, monotone way on an unobservable productivity variable. Altonji and Ichimura (1997) consider models with one dependent variable, and estimate an average derivative. Briesch, Chintagunta, and Matzkin (1997) consider estimation of discrete choice models where an unobserved heterogeneity variable enters the systematic subutilities in a nonadditive way. Heckman and Vytlacil (1999, 2001) and Vytlacil (2000) consider models where potential outcomes are nonadditive in unobservable random terms. Blundell and Powell (2000) consider a nonadditive structural function, and estimate its average. Altonji and Matzkin (2001) provide methods to estimate functions, distributions, and average derivatives in nonparametric, nonadditive models with endogenous regressors.

Recently, Bajari and Benkard (2001), Chesher (2002a,2002b), Heckman, Matzkin and Nesheim (2002), Hong and Shum (2001), Imbens and Newey (2001), and Matzkin (2002) have extended some of the identification ideas presented in this paper to develop estimation methods for hedonic prices (Bajari and Benkard), triangular equation models (Chesher and Imbens and Newey), hedonic equilibrium models (Heckman, Matzkin and Nesheim), and models where endogeneity is dealt with using functional restrictions (Matzkin).

In nonparametric models where the unobservable random term is additive, shape restrictions have been used in previous work to identify otherwise unidentified nonparametric functions and to estimate nonparametric models (see, for example, Matzkin (1992)). Matzkin (1994) provides a review of some of the existent literature for limited dependent variable models and nonparametric regression functions.

There is also a large literature in econometrics, which started with Heckman and Singer (1984a), on models that incorporate an unobservable random term, which is interpreted as an heterogeneity parameter, and whose distribution is nonparametric.

The outline of the paper is as follows. In the next section, we present the basic model and study its identification. In Section 3, we present estimators for the function m and the distribution of ε, together with their asymptotic properties. The results are extended to functions that depend on a multidimensional random term ε, in Appendix A. Section 4 presents the results of some simulations. A short summary is presented in Section 5. Appendix B contains the proofs of the main theorems.

2. THE MODEL

The building block for the models that we will study can be described by the basic model

$$(2.1) \quad Y = m(X, \varepsilon)$$

where $m : A \times E \rightarrow R$ is continuous in (X, ε) and strictly increasing in ε, $A \subset R^L$ is the support of X, $E \subset R$ is the support of ε, Y and X are observable, X has a continuous density f_X, and ε is an unobservable random term which is distributed, with a distribution F_ε, independently (or conditionally independently) of X. Many widely used type of models fall into this category. Models where ε represents unobserved heterogeneity or a technological shock may satisfy model (2.1). Models that are expressed in terms of an unobservable variable that is not independent of X may be rewritten as models with an unobservable random term that is independent of X. If $Y = r(X, \eta)$, where η is not independent of X, but $\eta = s(X, \varepsilon)$ where ε is independent of X, then $Y = r(X, s(X, \varepsilon)) = m(X, \varepsilon)$. Suppose, for example, that we represent the relationship $Y = m(X, \varepsilon)$ by $Y = v(X) + \eta$, where $v(X) = E(Y|X)$. Then, $\eta = Y - v(X)$ is mean independent of X, but will, in general, depend on X. The conditional expectation function v is useful to predict Y. However, this function is not as useful when one is interested in studying the structural random relationship between Y and X, which gives information about the interaction between the observable X and the unobservable ε. In fact, estimating $m(X, \varepsilon)$ is analogous to estimating the function $\eta = s(X, \varepsilon) = Y - v(X)$.

Some transformation models satisfy (2.1), such as the one presented in Box and Cox (1964) and the semiparametric generalized regression model in Han (1987), when the transformation is strictly increasing. All the transformation models studied in Horowitz (1996), of the type $Y = \Lambda^{-1}(\beta' X + \varepsilon)$, where Λ is an unknown, strictly increasing function and ε is distributed independently of X with an unknown distribution, satisfy model (2.1).

Duration models, where Y denotes time in a state and ε is the negative of the log-integrated hazard function, fall into the category of model (2.1), even when the hazard function is not separable in any of its arguments. In this case, ε is distributed extreme value, independently of X, and $m(X, \varepsilon) = \Lambda^{-1}(X, e^{-\varepsilon})$, where $\Lambda(X, Y)$ is the integrated hazard up to time Y, conditional on X, and $\Lambda^{-1}(X, \cdot)$ denotes the inverse of $\Lambda(X, Y)$ with respect to Y.

Duration models with unobserved heterogeneity also satisfy model (2.1), when the conditional hazard function is multiplicative in the unobserved heterogeneity variable. Let θ denote the unobserved heterogeneity variable, assumed to be distributed independently of X. Let $h(s|X, \theta)$ denote the conditional hazard function, and suppose that it can be written as $h(s|X, \theta) = r(s, X)e^{-\theta}$ for some unknown, nonnegative function r. Let $\varepsilon = u + \theta$, where u is the negative of the log of the integrated conditional hazard function. Then, u is distributed extreme value, independently of (X, θ), and, hence, ε is independent of X. In this model $m(X, \varepsilon) = \Lambda^{-1}(X, e^{-\varepsilon})$, with $\Lambda(X, Y) = \int_{s=0}^{Y} r(s, X)ds$. The identification of this model, with r possessing no particular structure, was studied in Heckman (1991). The case where $r(s, X) = r_1(s)r_2(X)$ was studied by Elbers and Ridders (1982), Heckman and Singer (1984), Barros and Honore (1988), and Ridders (1990). (See Barros (1986) for the case where $r(s, X)$ is a known function of $r_1(s)$ and $r_2(X)$.)

In many situations, the value of Y is determined by a vector, $(\varepsilon_1, ..., \varepsilon_K)$, of unobservable variables, instead of by a single variable. In Appendix A, we deal with this important case.

The first question that arises when specifying the model in (2.1) 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 belongs. 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_ε" 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), $F_{y,x}(y, x; m', F'_\varepsilon) = F_{y,x}(y, x; m^{"}, F^{"}_\varepsilon)$.

To analyze the identification of (m, F_ε) in model (2.1), we first note that, since m is strictly increasing in ε, there exists a function v such that for all $x \in A$, $\varepsilon \in E$, and $y \in m(A, E)$, $v(x, y) = \varepsilon$ if and only if $y = m(x, \varepsilon)$. Hence, the function v is the inverse of m, conditional on X. Clearly, (v, F_ε) is identified if and only if (m, F_ε) is identified. Let Γ denote a set of continuous, strictly increasing distribution functions. Let V denote a set of continuous functions to which v belongs. The next Lemma shows what properties V has to satisfy to guarantee the identification of (v, F_ε) in $V \times \Gamma$. If the function v were assumed to be differentiable, we could present a different proof for this lemma, using the results in Roehrig (1988).

LEMMA 1: $v, \widetilde{v} \in V$ *are observationally equivalent if and only if there exists a strictly increasing function* $g : R \to R$ *such that* $\widetilde{v} = g \circ v$.

PROOF OF LEMMA 1**:** Note that, by the definition of v and the independence between ε and X,

$$\Pr(Y \leq y|X=x) = \Pr(m(X,\varepsilon) \leq y|X=x) = \Pr(\varepsilon \leq v(x,y)|X=x) = F_\varepsilon(v(x,y)).$$

Hence, $F_{Y|X=x}(y) = F_\varepsilon(v(x,y))$.

If v and $\widetilde{v}$ are observationally equivalent, there exist $\widetilde{F}_\varepsilon$ and $\overline{F}_\varepsilon$ in Γ such that for all (x,y), $\overline{F}_\varepsilon(v(x,y)) = \widetilde{F}_\varepsilon(\widetilde{v}(x,y))$. Since $\widetilde{F}_\varepsilon$ is strictly increasing, $\widetilde{v}(x,y) = (\widetilde{F}_\varepsilon)^{-1} \circ \overline{F}_\varepsilon(v(x,y))$. Let $g = (\widetilde{F}_\varepsilon)^{-1} \circ \overline{F}_\varepsilon$. Then, g is strictly increasing and $\widetilde{v} = g \circ v$.

On the other side, suppose that $\widetilde{v} = g \circ v$ for some strictly increasing function g. Let $\widetilde{F}_\varepsilon = F_\varepsilon \circ g^{-1}$. It then follows that

$$F_{Y|X=x}(y; v, F_\varepsilon) = F_\varepsilon(v(x,y)) = \widetilde{F}_\varepsilon(\widetilde{v}(x,y)) = F_{Y|X=x}(y; \widetilde{v}, \widetilde{F}_\varepsilon)$$

Hence, v and $\widetilde{v}$ are observationally equivalent. This completes the proof.

The lemma states that the function v is identified up to a monotone transformation, g. One implication of this is that ratios of derivatives of v are identified, without requiring any normalization. Another implication is that for any monotone transformation g, $(g \circ v, F_\varepsilon \circ g^{-1})$ and (v, F_ε) generate the same distribution of (Y,X). To see what this means in terms of the inverse function m, suppose that m^* and F_ε^* are the true function and distribution, and let v^* denote the inverse function of m^*, conditional on x. Then, $\varepsilon = v^*(x,y)$ is distributed with F_ε^* and $y = m^*(x,\varepsilon)$. Let g be any strictly increasing transformation. Let $\widetilde{\varepsilon} = g(\varepsilon)$ and $\widetilde{v}(x,y) = g(v^*(x,y))$. The lemma implies that the model $\widetilde{\varepsilon} = g(\varepsilon) = g(v^*(x,y)) = \widetilde{v}(x,y)$ generates the same distribution of the observable variables as the model $\varepsilon = v^*(x,y)$. Let $\widetilde{m}$ denote the inverse function of $\widetilde{v}$, conditional on x. Then, for any value e, $\widetilde{m}(x,e)$ denotes the value of y that satisfies $e = \widetilde{v}(x,y)$. Let then $\widetilde{\varepsilon}$ and x be given. To find such a value of y, we note that since $\widetilde{\varepsilon} = \widetilde{v}(x,y) = g(v^*(x,y))$, $v^*(x,y) = g^{-1}(\widetilde{\varepsilon})$. Hence, since m^* is the inverse of v^*, conditional on x, $y = m^*(x, g^{-1}(\widetilde{\varepsilon}))$. This shows that $\widetilde{m}(x,\widetilde{\varepsilon}) = m^*(x, g^{-1}(\widetilde{\varepsilon}))$, or, since $\widetilde{\varepsilon} = g(\varepsilon)$, $\widetilde{m}(x, g(\varepsilon)) = m^*(x,\varepsilon)$. Hence, $\widetilde{m}$ and m^* are observationally equivalent if and only if $\widetilde{m}$ equals m^* with ε substituted by $g(\varepsilon)$, for some strictly increasing function g, that is $\widetilde{m}(x, g(\varepsilon)) = m^*(x,\varepsilon)$.

The discussion in the above paragraph shows that, for normalization purposes, we are free to choose the function g. One convenient normalization is given by the function g such that, for some given value $\overline{x}$ of X, $g(v(\overline{x},y)) = y$.

The function $m,$ which is the inverse function of $g \circ v$ is the function that satisfies $m(\overline{x}, \varepsilon) = \varepsilon$. Hence, this normalization amounts to fixing the values of the function m at some value of the vector X. Note that a normalization of this type is implicitly assumed when specifying a linear random coefficient model where $m(x, \varepsilon) = \varepsilon \cdot x$, in which case $\overline{x} = 1$. An implication of this is that the linear random coefficient model is too restrictive. There is no need to specify a multiplicative structure between ε and x. One only needs the property that $m(1, \varepsilon) = \varepsilon$ in order to identify the distribution of ε and the function m. Note also that the linear specification $m(x, \varepsilon) = \beta \cdot x + \varepsilon$ satisfies this normalization with $\overline{x} = 0$. Somewhat more generally, we could require that $m(x_0, \overline{x}_1, \varepsilon) = \varepsilon$, for all x_0 and some given $\overline{x}_1$, where $X = (X_0, X_1)$. If, for example, $m(x_0, \overline{x}_1, \varepsilon) = \varepsilon \cdot \overline{x}_1 + r(x_0, \overline{x}_1)$, where $r(x_0, x_1) = 0$ when $x_1 = \overline{x}_1$, then m would satisfy this. Note that the structure would not need to be maintained when $X_1 \neq \overline{x}_1$.

When using a normalization of the type $m(\overline{x}, \varepsilon) = \varepsilon$ when estimating a random demand or supply function, it may become important to know what is the implied normalization in the generating utility or production function. Suppose for example that $m(x, \varepsilon)$ represents the demand for a single input by a perfectly competitive firm, where $X = w$ is the input price, in terms of the output price, and ε is a productivity shock. Denote the random production function of the firm by $f(y, \varepsilon)$, where y denotes the quantity of the input. Then, the value $y = m(w, \varepsilon)$ that satisfies the first order condition for profit maximization is that for which $f_1(y, \varepsilon) = w$, where f_1 denotes the derivative of f with respect to its first coordinate. The condition that $m(\overline{w}, \varepsilon) = \varepsilon$ for some $\overline{w}$ can be restated in terms of the production function f, by requiring that for each t, $f_1(t, t) = \overline{w}$. This condition states that along the isoprofit defined by $\overline{w}$, the value of the productivity shock ε that corresponds to a firm whose production function is tangent to the isoprofit at $Y = y$ is $\varepsilon = y$. Examples of random production functions where the productivity shock enters in this way are those that, for some $\alpha \in (0, 1)$ and all $\varepsilon > 0$, concide on the ray where $y = \varepsilon$ with functions of the type $f(y, \varepsilon) = y^{\alpha} \varepsilon^{1-\alpha}$ or $f(y, \varepsilon) = (y^{\alpha} + \varepsilon^{\alpha})^{1/\alpha}$. For the first type, $\overline{w} = \alpha$, while for the second $\overline{w} = (2)^{(1-\alpha)/\alpha}$. (For the use of this normalization in hedonic models see Heckman, Matzkin and Nesheim (2002)).

An alternative route to choosing a normalization is to see whether the restrictions of economic theory that are implied on the function m could be used to restrict the set of functions v in such a way that no two different functions that satisfy those restrictions can be strictly transformations of each other. Suppose for example that the function m is homogeneous of degree one in ε and some other of its arguments, on some given ray from the origin. More specifically, suppose that, for some $X = (\overline{x}_0, \overline{x}_1)$, some $\alpha \in R$, some $\overline{\varepsilon}$, and all $\lambda \geq 0$

$$m(\overline{x}_0, \lambda\overline{x}_1, \lambda\overline{\varepsilon}) = \lambda\alpha \quad where \;\; m(\overline{x}_0, \overline{x}_1, \overline{\varepsilon}) = \alpha$$

Then, using arguments as those in Matzkin (1992, 1994), one can show that for any two conditional inverse functions $v,$ corresponding to two different functions m, it is not possible to write one of those v functions as a strictly increasing transformation of the other. One can obtain the same effect if the function m is such that for some $\overline{x}_1$, some $\alpha \in R$, all x_0 and all $\lambda \geq 0$

$$m(x_0, \lambda\overline{x}_1, \lambda\overline{\varepsilon}) = \lambda\alpha \quad where \;\; m(x_0, \overline{x}_1, \overline{\varepsilon}) = \alpha.$$

When m is a profit function or a cost function, m is homogeneous of degree one in all or some of its arguments. Hence, in these cases, identification requires only a location normalization, which can be imposed by fixing the value of the function at one point. Suppose, for example, that X_0 denotes a vector of observable characteristics of a typical firm, (X_1, ε) denotes the vector of output and input prices, and m denotes the profit of the firm. If the firm chooses its output and input quantities taking prices as given, m will be homogenous of degree one in (X_1, ε). As another example, suppose that X_0 denotes the output quantity of a typical firm, (X_1, ε) denotes a vector of input prices, and m denotes the cost function of the firm. Then, if the firm minimizes costs taking input prices as given, m will be homogenous of degree one in (X_1, ε).

If it is reasonable to assume that the v function is additive in one of its arguments, then, again one can show that no two different functions v can be written as strictly increasing transformations of each other (see Matzkin (1992,1994)). More explicitly, suppose that $X = (X_0, X_{11}, X_{12})$ is such that $X_{12} \in R$, and that

$$v(x_0, x_{11}, x_{12}, y) = r(x_0, x_{11}, y) + x_{12}$$

where for some $(\overline{x}_0, \overline{x}_{11}, \overline{y})$, $r(\overline{x}_0, \overline{x}_{11}, \overline{y}) = \alpha$. Then, the inverse function m has the form

$$m(x_0, x_{11}, x_{12}, \varepsilon) = s(x_0, x_{11}, \varepsilon - x_{12}) \qquad where \; s(\overline{x}_0, \overline{x}_{11}, \alpha) = \overline{y}.$$

This specification can be seen as a nonparametric, partially nonadditive generalization of the transformation model studied in Horowitz (1996), where $Y = \Lambda^{-1}(\beta' X + \varepsilon)$, Λ^{-1} is unknown and strictly increasing, and the distribution of ε is unknown. In Horowitz (1996), the value of Λ is specified at one point and the absolute value of the coefficient of one coordinate of X is set to 1. In our specification, we specify the value of s at the point $(\overline{x}_0, \overline{x}_{11}, \alpha)$ and set the coefficient of X_{12} equal to -1. (Note also the resemblance with the parametric, random production function specified in McElroy (1987)). The identification here can be also achieved if

$$m(x_0, x_{11}, x_{12}, \varepsilon) = s(x_0, x_{11}, \varepsilon - x_{12})$$

where for some $\overline{x}_{11}$ and all x_0, $s(x_0, \overline{x}_{11}, \alpha) = \overline{y}$. This would be satisfied, for example, if the function m were such that $m(x_0, \overline{x}_{11}, x_{12}, \varepsilon) = n_1(\overline{x}_{11}, x_{12} - \varepsilon) + n_2(x_0, \overline{x}_{11})$, for some unknown functions n_1 and n_2 such that $n_2(x_0, \overline{x}_{11}) = 0$ for all x_0. Note that this function need not be additively separable in the n_1 and n_2 functions when $X_{11} \neq \overline{x}_{11}$.

To see how this specification may arise, for example, in a demand function, suppose that the preferences of a typical consumer for commodities Z and Y are represented by a twice continuously differentiable, strictly increasing and strictly concave utility function $U(z-\varepsilon, y)$, with strictly positive cross partial derivative, U_{12}. Then, the solution to the maximization of U subject to the budget constraint $z + py = I$, which is obtained by maximizing $U(I - \varepsilon + py, y)$ over y, is given by a function of the form $Y = m(p, I - \varepsilon)$, which is strictly increasing in its last coordinate. If the utility function U depends also on some vector, w, of observable characteristics of the consumer, then we will have that $Y = m(w, p, I - \varepsilon)$.

An additional implication of Lemma 1 is that, instead of studying various specifications for the function m, one can achieve identification by specifying the distribution function F_ε. Suppose that the utility function of a typical consumer is a function $U(z, y, \varepsilon)$, which is strictly increasing and strictly concave with respect to its first two arguments and twice continuously differentiable with respect to its three arguments. The first and second order conditions for utility maximization subject the budget constraint $z + py = I$ together with the Implicit Function Theorem imply that the demand function $Y = m(p, I, \varepsilon)$ will be strictly increasing with respect to ε if for all p, $U_{13}p - U_{23} < 0$. This is satisfied, for example, if for some functions v and $\widetilde{v}$, $U(z, y, \varepsilon) = v(z, y) + \widetilde{v}(y, \varepsilon)$, where $\widetilde{v}_{12} > 0$. It follows by Lemma 1 that if the distribution of ε is specified, the demand function m will be identified. (See Brown and Matzkin (1996) and Heckman, Matzkin and Nesheim (2002) for methods to estimate the utility functions in a particular case and in hedonic models, respectively.)

In some cases, one may even know the distribution of ε. Consider, for example, a duration model with a nonparametric hazard function $\lambda(X,t) > 0$. Let $\varepsilon = \ln \int_{-\infty}^{T} \lambda(X,t)\, dt$. Then, it is well known that $\eta = -\varepsilon$ is distributed independently of X and, for all e, $F_\eta(e) = \exp\left(-\exp(-e)\right)$. Hence, $F_\varepsilon(e) = 1 - \exp\left(-\exp(e)\right)$. Let $y = T$. Then, using Lemma 1, we get the well known result that the function $v(x,y) = \ln \int_{-\infty}^{y} \lambda(x,t)\, dt$ is nonparametrically identified, since $F_{Y|X=x}(y) = 1 - \exp\left(-\exp\left(v(x,y)\right)\right)$.

3. Estimation of the Basic Model

To develop estimators for the function m and the distribution of ε in the basic model (2.1), we will derive expressions for these, in terms of the distribution of the vector of the observable variables. We will do this for the three basic specifications described in Section 2. Analogous expressions could be obtained for other specifications of the function m. Once the unknown functions and distributions are expressed in terms of the distribution of (Y,X), we will derive estimators for these unknown functions and distributions by substituting the distribution of the observable variables with a nonparametric estimator of it. While we could consider using any type of nonparametric estimator for this distribution, we present here the details and asymptotic properties for the case in which the conditional cdf's are estimated using the method of kernels. To express the unknown functions and distributions in terms of the distribution of the observable variables, let $X = (X_0, X_1)$. We will make the following assumptions:

ASSUMPTION I.1: *ε is independent of X_1, conditional on X_0.*

ASSUMPTION I.2: *For all values of X, m is strictly increasing in ε.*

Assumption I.1 guarantees that, conditional on X_0, the distribution of ε is the same for all values of X_1. Although we explicitly write X_0 as an argument of the function m, this is not necessary. The vector X_0 may be such that the function m is not a function of it. Assumption I.2 guarantees that the distribution of ε can be obtained from the conditional distribution of Y given X.

Under these assumptions, the mapping between the unknown functions m and $F_{\varepsilon|X}$ and the distribution of the observable variables $F_{Y,X}$ is given by

$$(3.1)\quad F_{\varepsilon|X_0=x_0}(e) = F_{Y|X=x}\left(m(x,e)\right) \qquad \text{for all } e \in E \text{ and } x = (x_0, x_1) \text{ such that } f_X(x) > 0,$$

This is because $F_{\varepsilon|X_0=x_0}(e) = \Pr\left(\varepsilon \leq e | X_0 = x_0\right) = P\left(\varepsilon \leq e | X_0 = x_0, X_1 = x_1\right)$ $= \Pr(m(X,\varepsilon) \leq m(x,e) \mid X = (x_0,x_1)\ ) = \Pr\left(Y \leq m(x,e) \mid X = x\right) =$ $F_{Y|X=x}\left(m(x,e)\right)$. The first equality follows by the definition of F_ε, the second follows by the conditional independence between ε and X_1, the third follows by the monotonicity of $m(x,\cdot)$ in its last argument, the fourth follows by the definition of Y, and the fifth equality follows by the definition of $F_{Y|X}$.

Equation (3.1) provides an easy interpretation of $m(x,e)$. From these equations it follows that $m(x,e)$ is the same quantile of the distribution of Y given $X = x$ as the quantile that e is of the distribution of ε conditional on X_0. In other words, let q be such that e is the q^{th} quantile of $F_{\varepsilon|X_0}$; then, by (3.1), $m(x,e)$ must be the q^{th} quantile of the conditional distribution, $F_{Y|X=x}$, of Y given $X = x$. The set $\{m(x,e)|x \in A\}$ then represents the set of the conditional q^{th} quantiles of the distribution of Y given X. So, for example, if the median of ε, conditional on X_0, is zero, then for all x, $m(x,0)$ is the median of Y conditional on X.

3.1. *Specification I*

Consider first the case where for all $\varepsilon \in E$, some $\overline{x}_1$ and all x_0 such that $f_X(x_0,\overline{x}_1) > 0$,

$$(\text{I.1})\qquad \begin{array}{c} m(x_0,\overline{x}_1,\varepsilon) = \varepsilon, \\ \textit{and Assumptions I.1 and I.2 are satisfied} \end{array}$$

Letting $X_1 = \overline{x}_1$ in (3.1), it follows that for all x_0 such that $f_X(x_0,\overline{x}_1) > 0$,and all $e \in E$,

$$(3.2)\quad F_{\varepsilon|X_0=x_0}(e) = F_{Y|X=(x_0,\overline{x}_1)}\left(e\right).$$

Hence, the conditional distribution of ε given $X_0 = x_0$ equals the conditional distribution of Y when $X = (x_0, \overline{x}_1)$. To derive an expression for the function m, we note that since $Y = m(X, \varepsilon)$ and $m(x, \cdot)$ is strictly increasing on E, 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 $m(x, E)$. From (3.1) and (3.2), it then follows that for all (x_0, x_1) such that $f_X(x_0, x_1) > 0$,

$$(3.3)\quad m(x, e) = F^{-1}_{Y|X=(x_0,x_1)} \left(F_{Y|X=(x_0,\overline{x}_1)} (e) \right).$$

Suppose, next that for all $\varepsilon \in E$, some $\overline{x}_1$ and all x_0 such that $f_X(x_0, \overline{x}_1) >$ 0,

$$\text{(I.2)}\qquad \begin{array}{c} m(x_0, \overline{x}_1, \varepsilon) = \varepsilon, \\ \textit{and Assumptions I.1' and I.2 are satisfied} \end{array}$$

where

ASSUMPTION I.1': *ε is independent of* (X_0, X_1).

then, we have that for all $e \in E$ and x such that $f_X(x) > 0$,

$$(3.1')\quad F_\varepsilon(e) = F_{Y|X=x} \left(m(x, e) \right)$$

Expression (3.1') follows because $F_\varepsilon(e) = P(\ \varepsilon \leq e | X = x) = \Pr(\ m(x, \varepsilon)$ $\leq m(x, e) \mid X = x\ ) = \Pr\left(Y\ \leq\ m(x, e) \mid X = x\right) = F_{Y|X=x}\left(m(x, e)\right)$. This expression implies, in particular, that for all $e \in E$ and all $\widetilde{x}_0$ such that $f_X(\widetilde{x}_0, \overline{x}_1) > 0$,

$$(3.2')\quad F_\varepsilon(e) = F_{Y|X=(\widetilde{x}_0,\overline{x}_1)} (e) \quad \text{and}$$
$$(3.3')\quad m(x, e) = F^{-1}_{Y|X=(x_0,x_1)} \left(F_{Y|X=(\widetilde{x}_0,\overline{x}_1)} (e) \right)$$

The overidentification of $F_\varepsilon(e)$ and $m(x, e)$, in this case, is the result of strengthening the conditional independence assumption I.1 to the stronger independence assumption I.1'. Since $\int f_{X_0|X_1=\overline{x}_1}(\widetilde{x}_0)\, d\widetilde{x}_0 = 1$, it follows from $(3.2')$ that

$$F_\varepsilon(e) = \int F_\varepsilon(e)\ f_{X_0|X_1=\overline{x}_1}(\widetilde{x}_0)\ d\widetilde{x}_0$$

$= \int F_{Y|X=(\widetilde{x}_0,\overline{x}_1)}(e)\ f_{X_0|X_1=\overline{x}_1}(\widetilde{x}_0)\ d\widetilde{x}_0$

$= \int \int_{-\infty}^{e} \frac{f(s,\widetilde{x}_0,\overline{x}_1)}{f(\widetilde{x}_0,\overline{x}_1)} \frac{f(\widetilde{x}_0,\overline{x}_1)}{f(\overline{x}_1)}\ ds\ d\widetilde{x}_0$

$= \int_{-\infty}^{e} \frac{f(s,\overline{x}_1)}{f(\overline{x}_1)} ds$

$= F_{Y|X_1=\overline{x}_1}(e)\ .$

Hence, under (I.2) we also have that

$$(3.2'')\ \ F_{\varepsilon}(e) = F_{Y|X_1=\overline{x}_1}(e) \quad \text{and}$$

$$(3.3'')\ \ m(x,e) = F^{-1}_{Y|X=(x_0,x_1)}\left(F_{Y|X_1=\overline{x}_1}(e)\right).$$

When X_0 is not an argument of m, (3.3″) implies that

$$(3.3''')\ \ m(x,e) = F^{-1}_{Y|X_1=x_1}\left(F_{Y|X_1=\overline{x}_1}(e)\right).$$

3.2. *Specification II*

Consider next the case where for some $\overline{\varepsilon} \in E$, some $\alpha, \overline{y} \in R$, some $\overline{x}_1$, all x_0 such that $f_X(x_0,\overline{x}_1) > 0$, and all $\lambda \in R$ such that $\lambda\overline{\varepsilon} \in E$ and $f_X(x_0,\lambda\overline{x}_1) > 0$,

$$\text{(II.1)} \qquad \begin{array}{c} m(x_0,\overline{x}_1,\overline{\varepsilon}) = \alpha \\ m(x_0,\lambda\overline{x}_1,\lambda\overline{\varepsilon}) = \lambda\alpha \\ \textit{and Assumptions I.1 and I.2 are satisfied} \end{array}$$

Then, given any λ and letting $x_1 = \lambda\overline{x}_1$ and $e = \lambda\overline{\varepsilon}$, we have, from (3.1), that for all such x_0, $F_{\varepsilon|X_0=x_0}(\lambda\overline{\varepsilon}) = F_{Y|X=(x_0,\lambda\overline{x}_1)}\left(m(x_0,\lambda\overline{x}_1,\lambda\overline{\varepsilon})\right) = F_{Y|X=(x_0,\lambda\overline{x}_1)}(\lambda\alpha)$, where the second equality follows because $m(x_0,\lambda\overline{x}_1,\lambda\overline{\varepsilon}) = \lambda m(x_0,\overline{x}_1,\overline{\varepsilon}) = \lambda\alpha$. In particular, for any $e \in E$ such that $f_X(x_0,(e/\overline{\varepsilon})\overline{x}_1) > 0$,

$$(3.4)\ F_{\varepsilon|X_0=x_0}(e) = F_{Y|X=(x_0,(e/\overline{\varepsilon})\overline{x}_1)}\left((e/\overline{\varepsilon})\alpha\right),$$

by letting $\lambda = (e/\overline{\varepsilon})$. Hence, $F_{\varepsilon|X_0=x_0}(e)$ can be recovered from the conditional cdf of Y given X, when $y = (e/\overline{\varepsilon})\alpha$ and $x = (x_0,(e/\overline{\varepsilon})\overline{x}_1)$. Since the strict monotonicity of $m(x,\cdot)$ implies that $F_{Y|X}$ has an inverse on $m(x,E)$, it follows from (3.1) and (3.4) that

$$(3.5)\quad m(x,e)=F^{-1}_{Y|X=x}\left(F_{Y|X=(x_0,(e/\overline{\varepsilon})\overline{x})}\left((e/\overline{\varepsilon})\alpha\right)\right),$$

which provides the mapping between $m(x,e)$ and the distribution of the observable variables.

Next, suppose that for some $\overline{\varepsilon}\in E$, some $\alpha,\overline{y}\in R$, some $\overline{x}_1$, all x_0 such that $f_X(x_0,\overline{x}_1)>0$, and all $\lambda\in R$ such that $\lambda\overline{\varepsilon}\in E$ and $f_X(x_0,\lambda\overline{x}_1)>0$,

$$\text{(II.2)}\qquad \begin{array}{c} m(x_0,\overline{x}_1,\overline{\varepsilon})=\alpha \\ m(x_0,\lambda\overline{x}_1,\lambda\overline{\varepsilon})=\lambda\alpha \\ \textit{and Assumptions I.1' and I.2 are satisfied}\end{array}$$

Then, using the same reasoning as used for the case where $m(x_0,\overline{x}_1,\varepsilon)=\varepsilon$, we have that (3.1') is satisfied, and we obtain the overidentification result that for all $\widetilde{x}_0$ such that $f_X(\widetilde{x}_0,(e/\overline{\varepsilon})\overline{x}_1)>0$

$$(3.4')\quad F_\varepsilon(e)=F_{Y|X=(X_0,X_1)=(\widetilde{x}_0,(e/\overline{\varepsilon})\overline{x}_1)}\left((e/\overline{\varepsilon})\alpha\right)\quad\text{and}$$

$$(3.5')\quad m(x,e)=F^{-1}_{Y|X=(x_0,x_1)}\left(F_{Y|X=(X_0,X_1)=(\widetilde{x}_0,(e/\overline{\varepsilon})\overline{x}_1)}\left((e/\overline{\varepsilon})\alpha\right)\right)$$

Using, analogously to the derivation of (3.2″), the fact that $\int f_{X_0|X_1=(e/\overline{\varepsilon})\overline{x}_1}(\widetilde{x}_0)\,d\widetilde{x}_0=1$, we get that

$$(3.4'')\quad F_\varepsilon(e)=F_{Y|X_1=((e/\overline{\varepsilon})\overline{x}_1)}\left((e/\overline{\varepsilon})\alpha\right)\quad\text{and}$$

$$(3.5'')\quad m(x_0,x_1,e)=F^{-1}_{Y|X=(x_0,x_1)}\left(F_{Y|X_1=((e/\overline{\varepsilon})\overline{x}_1)}\left((e/\overline{\varepsilon})\alpha\right)\right).$$

As in specification (I.2), when X_0 is not argument of m, (3.5″) can be substituted by

$$(3.5''')\quad m(x_1,e)=F^{-1}_{Y|X_1=x_1}\left(F_{Y|X_1=((e/\overline{\varepsilon})\overline{x}_1)}\left((e/\overline{\varepsilon})\alpha\right)\right).$$

3.3. *Specification III*

Finally, we consider the case where for some unknown function $s(\cdot)$, all $\varepsilon \in E$, some $\alpha, \overline{y} \in R$, some $\overline{x}_{11}$ and all x_0, x_{12} such that $f_X(x_0, \overline{x}_{11}, x_{12}) > 0$

$$\text{(III.1)} \qquad \begin{gathered} m(x_0, x_{11}, x_{12}, \varepsilon) = s(x_0, x_{11}, \varepsilon - x_{12}) \\ s(x_0, \overline{x}_{11}, \alpha) = \overline{y} \\ \textit{and Assumptions I.3 and I.4 are satisfied} \end{gathered}$$

where

ASSUMPTION I.3: *ε is independent of $X_1 = (X_{11}, X_{12})$, conditional on X_0.*

ASSUMPTION I.4: *For all (x_0, x_{11}), $s(x_0, x_{11}, \cdot)$ is strictly increasing.*

Then, for all $e \in E$ and $x = (x_0, x_{11}, x_{12})$ such that $f_X(x) > 0$,

$$\text{(3.6)} \quad F_{\varepsilon|X_0=x_0}(e) = F_{Y|X=x}\left(s(x_0, x_{11}, e - x_{12})\right)$$

since $F_{\varepsilon|X_0=x_0}(e) = \Pr\left(\varepsilon \le e|X_0 = x_0\right) = P\left(\ \varepsilon \le e|(X_0, X_1) = (x_0, x_1)\right) = \Pr(\ \varepsilon - X_{12} \le e - x_{12} \mid (X_0, X_1) = (x_0, x_1)\ ) = \Pr\left(s(X_0, X_{11}, \varepsilon - X_{12}) \le s(x_0, x_{11}, \varepsilon - x_{12}) \mid X = x\right)$ $= F_{Y|X=x}\left(s(x_0, x_{11}, e - x_{12})\right)$.
Letting $X_{11} = \overline{x}_{11}$ and $X_{12} = e - \alpha$, in (3.6), we get that

$$\text{(3.7)} \quad F_{\varepsilon|X_0=x_0}(e) = F_{Y|X=(x_0, \overline{x}_{11}, e-\alpha)}\left(\overline{y}\right).$$

Hence, the value of the conditional distribution of ε given X_0, at $\varepsilon = e$, equals the value of the conditional distribution of Y at $\overline{y}$, when $(X_{11}, X_{12}) = (\overline{x}_{11}, e - \alpha)$. To derive an expression for the function s, we use (3.6) and (3.7) to get

$$\text{(3.8)} \quad s(x_0, x_{11}, e - x_{12}) = F_{Y|X=x}^{-1}\left(F_{Y|X=(x_0, \overline{x}_{11}, e-\alpha)}\left(\overline{y}\right)\right)$$

If we consider

ASSUMPTION I.3': *ε is independent of X.*

and the specification is that for some unknown function $s(\cdot)$, all $\varepsilon \in E$, some $\alpha, \overline{y} \in R$, some $\overline{x}_{11}$ and all x_0, x_{12} such that $f_X(x_0, \overline{x}_{11}, x_{12}) > 0$

$$\text{(III.1)} \qquad \begin{array}{c} m(x_0, x_{11}, x_{12}, \varepsilon) = s(x_0, x_{11}, \varepsilon - x_{12}) \\ s(x_0, \overline{x}_{11}, \alpha) = \overline{y} \\ \textit{and Assumptions I.3' and I.4 are satisfied} \end{array}$$

then, we get an overidentification result that for all for all $\widetilde{x}_0$ such that $f_X(\widetilde{x}_0, \overline{x}_1, e-\alpha) > 0$

$$(3.7') \quad F_\varepsilon(e) = F_{Y|(X_0,X_1)=(\widetilde{x}_0,\overline{x}_{11},e-\alpha)}\left(\overline{y}\right) \quad \text{and}$$

$$(3.8') \quad s(x_0, x_{11}, e - x_{12}) = F^{-1}_{Y|X=(x_0,x_{11},x_{12})}\left(F_{Y|(X_0,X_1)=(\widetilde{x}_0,\overline{x}_{11},e-\alpha)}\left(\overline{y}\right)\right)$$

which, averaging out over $\widetilde{x}_0$, using the conditional pdf of X_0 given X_1, gives that

$$(3.7'') \quad F_\varepsilon(e) = F_{Y|X_1=(\overline{x}_{11},e-\alpha)}\left(\overline{y}\right) \quad \text{and}$$

$$(3.8'') \quad s(x_0, x_{11}, e - x_{12}) = F^{-1}_{Y|X=x}\left(F_{Y|X_1=(\overline{x}_{11},e-\alpha)}\left(\overline{y}\right)\right)$$

As in (I.2) and (II.2), if X_0 is not argument of the function s, then (3.8″) may be substituted by

$$(3.8''') \quad s(x_{11}, e - x_{12}) = F^{-1}_{Y|(X_{11},X_{12})=(x_{11},x_{22})}\left(F_{Y|(X_{11},X_{12})=(\overline{x}_{11},e-\alpha)}\left(\overline{y}\right)\right).$$

3.4. *Estimation using specifications I, II, and III*

To develop the estimators, 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, for all $(y,x) \in R^{1+L}$,

$$\hat{f}(y,x) = \frac{1}{N\sigma_N^{L+1}} \sum_{i=1}^N K\left(\frac{y-Y^i}{\sigma}, \frac{x-X^i}{\sigma}\right), \; \hat{F}(y,x) = \int_{-\infty}^y \int_{-\infty}^x \hat{f}_N(s,z)\, ds\, dz,$$

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

where $K : R \times R^L \to R$ is a kernel function and σ_N is the bandwidth. The above estimator for $F(y,x)$ was proposed in Nadaraya (1964). When $K(s,z) = k_1(s)k_2(z)$ for some kernel functions $k_1 : R \to R$ and $k_2 : R^L \to R$,

$$\hat{F}_{Y|X=x}(y) = \frac{\sum_{i=1}^{N} \widetilde{k}_1(\frac{y-Y^i}{\sigma})\ k_2(\frac{x-X^i}{\sigma})}{\sum_{i=1}^{N} k_2(\frac{x-X^i}{\sigma})}$$

where $\widetilde{k}_1(u) = \int_{-\infty}^{u} k_1(s)\, ds$. 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 the kernel function k_1 is an everywhere positive density on a convex support, this set of values will contain a unique point. The estimators are obtained by substituting $F_{Y|X}$ and $F^{-1}_{Y|X}$ by $\widehat{F}_{Y|X}$ and $\widehat{F}^{-1}_{Y|X}$, at the corresponding values of Y and X, in equations (3.2), (3.3), (3.2′), (3.3′), (3.2″), (3.3″), (3.3‴), (3.4), (3.5), (3.4′), (3.5′), (3.4″), (3.5″), (3.5‴), (3.7), (3.8), (3.7′), (3.8′), (3.7″), (3.8″), and (3.8‴). Hence, for example, when (I.1) is satisfied

$$\widehat{F}_{\varepsilon|X_0=x_0}(e) = \widehat{F}_{Y|X=(x_0,\overline{x}_1)}\,(e) \quad \text{and} \quad \widehat{m}(x,e) = \widehat{F}^{-1}_{Y|X=(x_0,x_1)}\left(\widehat{F}_{Y|X=(x_0,\overline{x}_1)}\,(e)\right),$$

when (I.2) is satisfied,

$$\widehat{F}_{\varepsilon}(e) = \widehat{F}_{Y|X_1=\overline{x}_1}\,(e) \quad \text{and} \quad \widehat{m}(x,e) = \widehat{F}^{-1}_{Y|X=(x_0,x_1)}\left(\widehat{F}_{Y|X_1=\overline{x}_1}\,(e)\right), \quad \text{with}$$

$$\widehat{m}(x,e) = \widehat{F}^{-1}_{Y|X_1=x_1}\left(\widehat{F}_{Y|X_1=\overline{x}_1}\,(e)\right)$$

when X_0 is not an argument of m.

When (II.1) is satisfied,

$$\widehat{F}_{\varepsilon|X_0=x_0}(e) = \widehat{F}_{Y|X=(x_0,(e/\overline{\varepsilon})\overline{x}_1)}\,((e/\overline{\varepsilon})\alpha) \quad \text{and} \quad \widehat{m}(x,e) = \widehat{F}^{-1}_{Y|X=x}\left(\widehat{F}_{Y|X=(x_0,(e/\overline{\varepsilon})\overline{x})}\,((e/\overline{\varepsilon})\alpha)\right),$$

with analogous expressions for when (II.2) is satisfied; and when (III.1) is satisfied

$$\widehat{F}_{\varepsilon|X_0=x_0}(e) = \widehat{F}_{Y|X=(x_0,\overline{x}_{11},e-\alpha)}\,(\overline{y}) \quad \text{and} \quad \widehat{s}(x_0,x_{11},e-x_{12}) = \widehat{F}^{-1}_{Y|X=x}\left(\widehat{F}_{Y|X=(x_0,\overline{x}_{11},e-\alpha)}\,(\overline{y})\right),$$

with analogous expressions for when (III.2) is satisfied.

In all the above definitions, the value of the marginal or conditional distribution of ε at some given value e, is given by the value of the conditional distribution of Y, given that X, or, more generally, a subvector, W, of X, equals a given value, w. This conditional distribution of Y is evaluated at some given value y. The estimator is obtained by substituting the true conditional distribution of Y by its kernel estimator. Thus, the consistency and asymptotic normality of the estimator of the marginal or conditional distribution of ε will follow from the consistency and asymptotic normality of the kernel estimator for the conditional distribution of Y given that $W = w$. In particular, the asymptotic properties for each of the estimators for the distribution of ε given above can be derived from the result in Theorem 1, below, after substituting the corresponding values of w and y. Let W denote a subvector of X of dimension d. Let w be a particular value of W. Let $\int K(u)^2 = \int \left(\int K(u,v)\, dv\right)^2 du$, where $v \in R^{1+L-d}$ and $u \in R^d$. We make the following assumptions:

ASSUMPTION C.1**:** *The sequence* $\{Y^i, X^i\}$ *is i.i.d.*

ASSUMPTION C.2: $f(Y,X)$ *has compact support* $\Theta \subset R^{1+L}$ *and it has an extension to all of* R^{1+L} *that is continuously differentiable up to the order* s', *for some* $s' > 0$.

ASSUMPTION C.3: *The kernel function* $K(\cdot,\cdot)$ *is differentiable of order* $\widetilde{s}$, *the derivatives of* K *of order* $\widetilde{s}$ *are Lipschitz,* $K(\cdot)$ *vanishes outside a compact set, integrates to 1, and is of order* s'', *where* $\widetilde{s} + s'' \leq s'$.

ASSUMPTION C.4: *As* $N \to \infty$, $\sigma_N \to 0, \ln(N)/\left(N\sigma_N^{L+1}\right) \to 0, \sqrt{N}\sigma_N^{d/2} \to \infty$, $\sqrt{N}\sigma_N^{(d/2)+s''} \to 0$, and $\sqrt{N\sigma_N^d}\left(\sqrt{(\ln(N))\,/\left(N\sigma_N^{L+1}\right)} + \sigma^{s''}\right)^2 \to 0$.

ASSUMPTION C.5**:** $0 < f(w) < \infty$.

Assumption C.2 requires that the pdf of (Y,W) 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 C.3 restricts the kernel function that may be used. Assumption C.4 restricts the rate at which the bandwidth, σ_N, goes to zero.

THEOREM 1: *Let $\widehat{F}_{Y|W=w}(y)$ denote the kernel estimator for the conditional distribution of Y, conditional on $W = w$, evaluated at $Y = y$. Suppose that Assumptions C.1-C.5 are satisfied, for $\widetilde{s} \geq 0$ and $s'' \geq 2$. Then,*

$$\sup_{y \in R} \left| \widehat{F}_{Y|W=w}(y) - F_{Y|W=w}(y) \right| \to 0 \quad in \quad probability, \text{ and}$$

$$\sqrt{N}\sigma^{(d/2)} \left(\widehat{F}_{Y|W=w}(y) - F_{Y|W=w}(y) \right) \to N\left(0, V_F\right) \quad in\ distribution, \text{ where}$$

$$V_F = \left\{ \int K(u)^2 \right\} \left[F_{Y|W=w}(y) \ (1 - F_{Y|W=w}(y)\right] [1/f(w)]$$

The proof is given in Appendix B. This theorem shows that $\widehat{F}_\varepsilon$ converges to F_ε in the supremum norm, and $\widehat{F}_\varepsilon(e)$ is asymptotically normal with mean $F_\varepsilon(e)$ and variance equal to $\left\{\int K(u)^2\right\}$ $[F_\varepsilon(e)\ (1 - F_\varepsilon(e)]\ /\ \left[N\ f(w)\ \sigma^d\right]$, where w is the value of W on which we have to condition $\widehat{F}_{Y|W}$ to estimate $\widehat{F}_\varepsilon(e)$, and where d is the dimensionality of W.

To study the asymptotic properties of the estimator for the unknown function m, we note that the value of the function m at any given vector (w, e) is given by the composition of $F^{-1}_{Y|W=w}$ and $F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})$, for some particular vector values w and $\widetilde{w}$, and some particular value $\widetilde{e}$. By $F^{-1}_{Y|W=w}$ we denote the inverse of the conditional distribution of Y given that the subvector, W, of X, equals a value w; by $F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})$ we denote the conditional distribution of Y given that the subvector, $\widetilde{W}$, of X, equals the value $\widetilde{w}$. The subvectors W and $\widetilde{W}$, of X, are not required to have the same dimension. The estimator is obtained by substituting the true conditional distributions of Y by their kernel estimators. Hence, the consistency and asymptotic normality of the estimator of m will follow from the consistency and asymptotic normality of the functional, Φ, of the kernel estimator for the distribution of (Y, X), which is defined by $\Phi(\widehat{F}_{Y,X}) = \widehat{F}^{-1}_{Y|W=w}\left(\widehat{F}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)$. Let d_1 denote the number of coordinates of $\widetilde{W}$, d_2 denote the number of coordinates of W, and let $d = \max\{d_1, d_2\}$. Let $1[\cdot] = 1$ if the expression in $[\cdot]$ is true; $1[\cdot] = 0$ otherwise.

Let $\int K(u)^2 = \int \left(\int K(u,v)\, dv\right)^2 du$, where $v \in R^{1+L-d}$ and $u \in R^d$. Our next theorem will make use of Assumptions C.1-C.3 and the following:

ASSUMPTION C.4': *As* $N \to \infty,\ \sigma_N \to 0$, $\ln(N)/\left(N\sigma_N^{L+1}\right) \to 0$, for $j = 1, 2$, $\sqrt{N}\sigma_N^{d_j/2} \to \infty$, $\sqrt{N}\sigma_N^{(d_j/2)+s''} \to 0$, and $\sqrt{N\sigma_N^d}\left(\sqrt{(\ln(N))\,/\left(N\sigma_N^{L+1}\right)} + \sigma^{s''}\right)^2 \to 0$.

ASSUMPTION C.5': *The subvectors W and $\widetilde{W}$ have at least one coordinate in common, and the values, w and $\widetilde{w}$, are different at one such coordinate;* $0 < f(w), f(\widetilde{w}) < \infty$; *and there exist* $\delta, \xi > 0$ *such that* $\forall s \in N(m(w,e),\xi),\ f(s,w) \geq \delta$.

THEOREM 2: *Let* $\widehat{n}(w,e) = \widehat{F}^{-1}_{Y|W=w}\left(\widehat{F}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)$ *and* $n(w,e) = F^{-1}_{Y|W=w}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)$. *Suppose that Assumptions C.1-C.3, C.4' and C.5' are satisfied with* $s'' \geq 2$ *and* $s' \geq s''$. *Then,*

$$\widehat{n}(w,e) \to n(w,e) \text{ in probability and}$$

$$\sqrt{N}\sigma_N^{d/2}\left(\widehat{n}(w,e) - n(w,e)\right) \to N\left(0, V_n\right) \ \text{ in distribution, where}$$

$$V_n = \left\{\int K(u)^2\right\}\left[\frac{F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\ \left(1 - F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)}{f_{Y|W=w}(n(w,e))^2}\right]\left[\frac{1[d_1 = d]}{f(\widetilde{w})} + \frac{1[d_2 = d]}{f(w)}\right]$$

The proof is given in Appendix B. The statement of the theorem implies then that $\widehat{n}(w,e)$ is consistent and asymptotically normal with mean $n(x,e)$ and asymptotic variance equal to $\left\{\int K(u)^2\right\}\left(F_\varepsilon(e)\ \left(1 - F_\varepsilon(e)\right)\right)\left[\frac{1[d_1=d]}{f(\widetilde{w})} + \frac{1[d_2=d]}{f(w)}\right]$ / $\left[f_{Y|W=w}(n(w,e))^2\ N\ \sigma_N^d\right]$, where $\widetilde{w}$ is the value of the vector $\widetilde{W}$ on which we have to condition $\widehat{F}_{Y|\widetilde{W}}$ to estimate $\widehat{F}_\varepsilon(e)$, w is the value of the vector W that enters as a coordinate in the function n, and d is the maximum between the number of coordinates of $\widetilde{W}$ and W. Note that the value of the density of $\widetilde{w}$ influences the asymptotic variance only when the number of coordinates of $\widetilde{w}$ is at least as large as that of w. Also, if the distribution of ε is specified, instead of being estimated, so that $\widehat{n}(w,e) = \widehat{F}^{-1}_{Y|W=w}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)$ where $F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})$ is known, then $\sqrt{N}\sigma_N^{d/2}\left(\widehat{n}(w,e) - n(w,e)\right) \to N\left(0, V_n\right)$ in distribution where

$V_n = \left\{ \int K(u)^2 \right\} \left[\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) \ \left(1 - F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right) \right) / \left(f_{Y|W=w}(n(w,e))^2 \ f(w) \right) \right]$ with $d = d_2$.

While the kernel function used may be of any order larger than 2, $\widehat{F}^{-1}_{Y|W=w}$ will be a function only when the order is 2. With higher order kernels, $\widehat{F}^{-1}_{Y|W=w}$ will converge to a function, as the number of observations increases, but, for any given t, $\widehat{F}^{-1}_{Y|W=w}(t)$ may possess several values, when the number of observations is finite. Another issue that may be encountered in practice is that, with a finite number of observations, there may not exist a value n^* such that $n^* = \widehat{F}^{-1}_{Y|W=w}\left(\widehat{F}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)$. This may occur close to the endpoints of the support of $\widetilde{e}$, when the range of $\widehat{F}_{Y|W=w}$ does not include the range of $\widehat{F}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})$. To deal with this problem, one has to first find the minimum and maximum values, F_l and F^u, of the range of $\widehat{F}_{Y|W=w}$, and then define a function $\widehat{\widehat{F}}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})$ by: $\widehat{\widehat{F}}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) = F^u$ if $\widehat{F}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) > F^u$, $\widehat{\widehat{F}}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) = \widehat{F}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})$ if $F_l < \widehat{F}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) < F^u$, and $\widehat{\widehat{F}}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) = F_l$ if $\widehat{F}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) < F_l$. The estimator for $n(w,e)$ becomes then: $\widehat{n}(w,e) = \widehat{F}^{-1}_{Y|W=w}\left(\widehat{\widehat{F}}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)$.

To use the results of Theorems 1 and 2 in tests of hypotheses, it is necessary to replace V_F and V_n by consistent estimators. Under the assumptions of Theorem 1, a consistent estimator for V_F can be obtained by replacing $F_{Y|W=w}(y)$ and $f(w)$, in the equation for V_F, by their respective kernel estimators, $\widehat{F}_{Y|W=w}(y)$ and $\widehat{f}(w)$. Under the assumptions of Theorem 2, a consistent estimator for V_n can be obtained by replacing, in the equation for V_n, $F_{Y|\widetilde{W}=\widetilde{w}}\left(\widetilde{e}\right)$ by $\widehat{F}_{Y|\widetilde{W}=\widetilde{w}}\left(\widetilde{e}\right)$, $f\left(\widetilde{w}\right)$ by $\widehat{f}\left(\widetilde{w}\right)$, $f\left(w\right)$ by $\widehat{f}\left(w\right)$, and $f_{Y|W=w}\left(n(w,e)\right)$ by $\widehat{f}_{Y|W=w}\left(\widehat{n}(w,e)\right)$, where $\widehat{F}_{Y|\widetilde{W}=\widetilde{w}}$, $\widehat{f}$, and $\widehat{f}_{Y|W=w}$ are the kernel estimators for, respectively, $F_{Y|\widetilde{W}=\widetilde{w}}$, f, and $f_{Y|W=w}$, and where $\widehat{n}(w,e)$ is as defined in the statement of Theorem 2.

3.5. *Estimation of Derivatives*

In many cases in economics, we estimate a function because we are interested in its derivatives. For example, we might estimate a demand function because we want to study some price effect. As another example, we might want to estimate the demand and supply functions of a firm, when no observations are available on the demanded and supplied quantities, but there are observations on the profit of the firm, besides observations on prices. Then, we can estimate the profit function, and obtain the demand and supply functions by differentiating the estimated profit function with respect to prices. In particular, if ε represents an unobserved price in a profit function $m(x, \varepsilon)$, the derivative of m with respect to ε determines the demand for the input whose price is ε. Matzkin (1999) presents estimators for the derivatives of the function m with respect to x and ε, for some of the specifications presented in Section 2, and shows their consistency and asymptotic normality.

4. Simulations

To provide an indication of how the new estimators may perform in practice, we run a small simulation experiment, using the following two designs:

- *DESIGN I:* $Y = X + \epsilon$,

  where $X \sim N(0, 1)$ and $\epsilon \sim N(0, 1)$.

- *DESIGN II:* $Y = \frac{3^3}{4^4}\ X^4\ (-\varepsilon)^{-3}$

  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. Design II 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 wrote this function in terms of $-\varepsilon$ to transform it to be strictly increasing in ε.

For each design, we run 100 simulations of 250 and 500 observations. For each simulation, we estimated the functions m and F_ε at 100 fixed points, which were drawn from a uniform distribution with support $[-2, 2] \times [-2, 2]$ for design I and support $[4, 8] \times [-8, -4]$ for design II. Besides using our Nonparametric Nonadditive (NPNA) estimator, we also used, for comparison, a Nadaraya-Watson estimator (NW) and a Linear Least Squares Estimator (LS). When using the Nadaraya-Watson estimator, we estimated the model $y = m(x, \varepsilon) = v(x) + \varepsilon$, with v nonparametric and ε possibly dependent on

X. When using the Linear Least Squares estimator, we estimated the model $y = \beta \cdot x + \varepsilon$ with ε independent of X.

To estimate the functions m and F_ε using our new estimators, we specified $\overline{x} = \overline{\varepsilon} = 1$ and $\alpha = 2$ for Design I, and $\overline{x} = \overline{\varepsilon} = 6$ and $\alpha = 6 \cdot 3^3/4^4$ for Design II. The estimators were obtained using a multiplicative Gaussian kernel. The bandwidths that were used are presented in Table I. (Details about the simulations, including bandwidths selection, can be obtained from the author's web page. Matzkin (1999) presents the results obtained from using the same designs to estimate the derivatives of m with respect to x and ε.)

For each of the three estimation methods, we estimated the functions m and F_ε at each of the 100 fixed points that were drawn from a uniform distribution. For each point and estimated function, we used the simulations for which the estimated densities, and multiplications of densities, that appear in the denominator of the estimator, were above .025. $>$From these simulations, we calculated the absolute value of the bias, variance, and mean squared error. The averages of these results, over the 100 points, are reported in Tables II and III for design I, and in Tables IV and V for design II.

The graphs show the average behavior, over 500 simulations, of the NPNA estimators for design II with N=500. They show the average of $\widehat{m}$ over the simulations, and the mean (in a $--$ line), the median (in a $-.-$ 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). For graphs corresponding to Design I, see Matzkin (1999).

5. SUMMARY

We have presented estimators for models in which the value of a dependent variable is determined by a nonparametric function that is not necessarily additive in unobservable random terms. The estimators for the distribution of the unobservable random terms and the nonparametric function were derived and were shown to be consistent and asymptotically normal. The estimators were defined as nonlinear functionals of a kernel estimator for the distribution of the observable variables. The results of some simulations indicate that the method may outperform alternative parametric and nonparametric estimators.

Appendix A: Multivariate Unobservable Random Term

Imposing some structure on the function m, we can use the basic model described in Section 3 to identify and estimate random functions that depend on a vector of unobservable random terms. Let $X = (X_0, X_1)$ be such that $X_0 = w_0$, and $X_1 = (w_1, ..., w_K)$. Let $\varepsilon = (\varepsilon_1, ..., \varepsilon_K)$. Assume that ε is distributed independently of X_1 conditional on X_0. Assume, further, that the joint distribution of $(\varepsilon_1, ..., \varepsilon_K)$, conditional on X_0, is the multiplication of the marginal distributions of the ε_k's, conditional on X_0. For each k, let w_{0_k} denote a subvector of w_0. Then, if the function m can be expressed as a known function of K basic functions, each of which satisfies model (2.1), it is possible, under some restrictions, to identify the distribution of ε and each of the K random functions.

Our results allow the identification of each individual function in a summation, when only the value of the sum of the random functions is observed. They also allow the identification of each individual function in a multiplication, when only the total value of the multiplication of the random functions is observed. The summation case would be important, for example, if we were interested in identifying individual random behavior from observations on only the aggregate value of a dependent variable. The multiplicative case would be important, for example, if we were interested in estimating a multiplicative production function for some product, when the product is produced using some intermediate inputs. If these intermediate products were unobserved and were produced by some observable, more basic products, according to some unknown random production functions, then, using the results below, we can determine that the random production functions of the unobservable intermediate inputs are identified, as well as the distributions of the unobservable random terms, ε.

We present the results for the case in which each of the K basic functions satisfies specification (I.1). Analogous results can be obtained by using the other possible specifications. Suppose that

$$m(X, \varepsilon) = r(n_1(w_{0_1}, w_1, \varepsilon_1), ..., n_K(w_{0_K}, w_K, \varepsilon_K))$$

for some known, continuously differentiable function $r : R^K \to R$ and some unknown, nonparametric functions $n_1, ..., n_K$. Note that in this specification, each subvector w_k enters as an argument only in the function n_k. Some, or all, of the coordinates of w_0 may enter as arguments in some, or all, of the functions n_k. Let $F_{\varepsilon|X_0}$ denote the unknown distribution of ε, conditional on X_0. Let $\alpha_1, ..., \alpha_K$ be known numbers. We will make the following assumptions:

(A.i) At $(\alpha_1, ..., \alpha_K)$, *the function* r *is strictly increasing in each of its arguments.*

(A.ii) For each k, *there exists a value* $\overline{w}_k$ *of* w_k *such that for all values of* (w_{0_k}, ε_k),
$n_k(w_{0_k}, \overline{w}_k, \varepsilon_k) = \varepsilon_k$.

(A.iii) For each k, *there exists a value* $\widetilde{w}_k$ *of* w_k *such that for all values of* (w_{0_k}, ε_k),
$n_k(w_{0_k}, \widetilde{w}_k, \varepsilon_k) = \alpha_k$,

(A.iv) For each k, *and each* $(w_{0_k}, w_k, \varepsilon_k)$ *such that* $w_k \neq \widetilde{w}_k$, $n_k(w_{0_k}, w_k, \varepsilon_k)$ *is strictly increasing*
in ε_k,

(A.v) For all $e_1, ..., e_K$, $f_{(\varepsilon_1,...,\varepsilon_K)|X_0=w_0}(e_1, ..., e_K) = \prod_{k=1}^K f_{\varepsilon_k|X_0=w_0}(e_k)$

(A.v') For all $e_1, ..., e_K$, $f_{(\varepsilon_1,...,\varepsilon_K)}(e_1, ..., e_K) = \prod_{k=1}^K f_{\varepsilon_k}(e_k)$

(A.vi) $f_{(\varepsilon_1,...,\varepsilon_K)|X}(e_1, ..., e_K) = f_{(\varepsilon_1,...,\varepsilon_K)|X_0}(e_1, ..., e_K)$,

(A.vi') $f_{(\varepsilon_1,...,\varepsilon_K)|X}(e_1, ..., e_K) = f_{(\varepsilon_1,...,\varepsilon_K)}(e_1, ..., e_K)$, *and*

Assumptions (A.ii) and (A.iv) impose on each function n_k the specification (I.1). Assumption (A.iii) is used to find values of the vector X for which the conditional distribution of Y coincides with the conditional distribution of n_k. A very simple example of a function m that satisfies assumptions (A.i)-(A.iv) is $m(X, \varepsilon) = \sum_{k=1}^K \varepsilon_k w_k$, where $w_k \in R$. In this case, $\overline{w}_k = 1$ and for $\alpha_k = 0$, $\widetilde{w}_k = 0$. Assumption (A.v) states that, conditional on X_0, the ε_k are independent across them, while Assumption (A.v') states that the ε_k are unconditionally independent across them. These assumptions allow us to identify, respectively, the conditional and unconditional joint distribution of ε, from the marginal distributions. If these conditions are not satisfied, we will only be able to show the identification of the marginal distributions of the

ε_k. By Assumption (A.vi), ε is independent of X_1, conditional on X_0, while by Assumption (A.vi'), ε is independent of $X = (X_0, X_1)$. For each k, let w^k denote the value of X_1 when $w_j = \widetilde{w}_j$ for $j \neq k$; let $\overline{w}^k$ denote the value of X_1 when $w_k = \overline{w}_k$ and $w_j = \widetilde{w}_j$ for $j \neq k$; let $X^k = (w_{0_k}, X_1)$, and, for each k, define the function $r_k : R \to R$ by $r_k(t) = r(\alpha_1, ..., \alpha_{k-1}, t, \alpha_{k+1}, ..., \alpha_K)$. We can now state the following result, which is proved in Appendix B:

Theorem 1 *3:(5.I) If Assumptions (A.i)-(A.vi) are satisfied, then $F_{\varepsilon|X_0=w_0}$ and m are identified. In particular, for all k and all (w_0, w_k, e_k),*

$$F_{\varepsilon_k|X_0=w_0}(e_k) = F_{Y|X=(w_0,\overline{w}^k)}(r_k(e_k)), \quad \textit{and}$$

$$n_k(w_{0_k}, w_k, e_k) = r_k^{-1}\left(F^{-1}_{Y|X^k=(w_{0_k},w^k)}\left(F_{Y|X=(w_0,\overline{w}^k)}(r_k(e_k))\right)\right)$$

(5.II) If Assumptions (A.i)-(A.iv), (A.v') and (A.vi') are satisfied, then F_ε and m are identified. In particular, for all k and all (w_0, w_k, e_k),

$$F_{\varepsilon_k}(e_k) = F_{Y|X_1=\overline{w}^k}\left(r_k(e_k)\right), \quad \textit{and}$$

$$n_k(w_{0_k}, w_k, e_k) = r_k^{-1}\left(F^{-1}_{Y|X^k=(w_{0_k},w^k)}\left(F_{Y|X_1=\overline{w}^k}\left(r_k(e_k)\right)\right)\right)$$

Since, in the statement of the above theorem, the random functions, n_k, and the marginal distributions of the ε_k's are expressed in terms of functionals of the distribution of the observable variables, we can define estimators for these functions and distributions by substituting the true distribution of (Y, X) by its kernel estimator, in a similar way as that followed in Section 3. The asymptotic properties of the estimators for the marginal distributions of the ε_k's can be determined using the results of Theorem 1. The consistency of the estimators for the n_k functions follows by the convergence in probability of $\widehat{F}^{-1}_{Y|X^k=(w_{0_k},w^k)}\left(\widehat{F}_{Y|X=(w_0,\overline{w}^k)}(r_k(e_k)\right)$ to $F^{-1}_{Y|X^k=(w_{0_k},w^k)}\left(F_{Y|X=(w_0,\overline{w}^k)}(r_k(e_k)\right)$ and the convergence in probability of $\widehat{F}^{-1}_{Y|X^k=(w_{0_k},w^k)}\left(\widehat{F}_{Y|X_1=\overline{w}^k}\left(r_k(e_k)\right)\right)$ to $F^{-1}_{Y|X^k=(w_{0_k},w^k)}\left(F_{Y|X_1=\overline{w}^k}\left(r_k(e_k)\right)\right)$, which can be established using the results of Theorem 2, and the continuity of the function r. The asymptotic distribution of the estimators for the n_k functions follow from the results of Theorem 2 and by the standard Delta method, using the continuous differentiability of the function r. Hence, under the assumptions of Theorem 2, we get that, when (A.i)-(A.vi) are satisfied, and d equals the dimension of $(w_0, \overline{w}^k)$,

$$\sqrt{N}\sigma_N^{d/2}\left(\widehat{n}_k(w_{0_k}, w_k, e_k) - n_k(w_{0_k}, w_k, e_k)\right) \to N\left(0, V_k\right) \quad \textit{in distribution}, \text{ where}$$

$$V_k = \left\{\int K(u)^2\right\}\left[\frac{F_{Y|X=(w_0,\overline{w}^k)}(r_k(e_k))\left(1 - F_{Y|X=(w_0,\overline{w}^k)}(r_k(e_k))\right)}{f_{Y|X^k=(w_{0_k},w^k)}(n_k(w_{0_k}, w_k, e_k))^2}\right]\left[\frac{1}{f(w_0,\overline{w}^k)} + \frac{1[w_{0_k} = w_0]}{f(w_{0_k}, w^k)}\right](\Delta_k)^2$$

$$\mathit{\Delta}_k = \left(\frac{\partial r_k^{-1}\left(F^{-1}_{Y|X^k=(w_{0_k},w^k)}\left(F_{Y|X_1=\overline{w}^k}\left(r_k(e_k)\right)\right)\right)}{\partial t}\right) = \frac{1}{\left(\frac{\partial r_k\left(n_k(w_{0_k},w_k,e_k)\right)}{\partial t}\right)}$$

When (A.i)-(A.iv), (A.v') and (A.vi') are satisfied,

$$\sqrt{N}\sigma_N^{d/2}\left(\widehat{n}_k(w_{0_k}, w_k, e_k) - n_k(w_{0_k}, w_k, e_k)\right) \to N\left(0, V_k'\right) \quad \textit{in distribution}, \text{ where}$$

$$V_k' = \left\{\int K(u)^2\right\}\left[\frac{F_{Y|X_1=\overline{w}^k}\left(r_k(e_k)\right)\left(1 - F_{Y|X_1=\overline{w}^k}\left(r_k(e_k)\right)\right)}{f_{Y|X^k=(w_{0_k},w^k)}(n_k(w_{0_k}, w_k, e_k))^2}\right]\left[\frac{1[d_1 = d]}{f(X_1 = \overline{w}^k)} + \frac{1[d_2 = d]}{f(w_{0_k}, w^k)}\right](\Delta_k')^2$$

$$\mathit{\Delta}_k' = \left(\frac{\partial r_k^{-1}\left(F_{Y|X_1=\overline{w}^k}\left(r_k(e_k)\right)\left(1 - F_{Y|X_1=\overline{w}^k}\left(r_k(e_k)\right)\right)\right)}{\partial t}\right) = \frac{1}{\left(\frac{\partial r_k\left(n_k(w_{0_k},w_k,e_k)\right)}{\partial t}\right)},$$

and d_1 *denotes the dimension of* $\overline{w}^k$, d_2 *denotes the dimension of* (w_{0_k}, w^k), *and* $d = \max\{d_1, d_2\}$.

Appendix B: Proofs of Theorems

In this Appendix, we provide the proofs of Theorems 1, 2, and 3. Theorems 1 and 2 present the asymptotic properties of our estimators for the distribution of ε and the function m. Since all these estimators are functionals of kernel estimators for the distributions of the observable variables, we develop their asymptotic properties using a Delta method, as developed in Newey (1994) and Ait-Sahalia (1996). We present the general "Delta-method" result in the Lemma below.

To deal with the situation that the estimators are conditioned on vectors that may possess only some coordinates in common, we partition $X \in R^L$ into $X = (W_0, W_1, W_2, W_3)$, where, after relabeling the axes accordingly, $X = (Z_1, X_{-1}) = (Z_2, X_{-2})$, $Z_1 = (W_0, W_1)$, $Z_2 = (W_0, W_2)$, $X_{-1} = (W_2, W_3)$, and $X_{-2} = (W_1, W_3)$. Hence, W_0 denotes the subvector of coordinates of X that Z_1 and Z_2 share, W_3 denotes the subvector of X which is not included in either Z_1 or Z_2, and W_1 and W_2 denote the subvectors of X that are included in one but not the other of Z_1 and Z_2. Let $d_1 = \dim(Z_1)$ and $d_2 = \dim(Z_2)$. For any function $G : R^{1+L} \to R$, define $g(y,x) = \partial^{1+L}G(y,x)/\partial y \partial x$, $g(z_1) = \int g(y, z_1, x_{-1})\, dy\, dx_{-1}$, $g(z_2) = \int g(y, z_2, x_{-2})\, dy\, dx_{-2}$, $g(y, z_1) = \int g(y, z_1, x_{-1})\, dx_{-1}$, $g(y, z_2) = \int g(y, z_2, x_{-2})\, dx_{-2}$, $G_{y|Z_1=z_1}(y') = \left(\int_{-\infty}^{y'} g(y, z_1) ds\right) / g(z_1)$, and $G_{y|Z_2=z_2}(y') = \left(\int_{-\infty}^{y'} g(y, z_2) ds\right) / g(z_2)$. Let C denote a compact set in R^{1+L} that strictly includes Θ, the compact support of $(Y \times X)$. Let D denote the set of all functions $G : R^{1+L} \to R$ such that $g(y,x)$ exists and vanishes outside C. Denote the norm $\|G\|$ by

$$\|G\| = \sup_{(y,x)\in\Theta} |g(y,x)|$$

Let $\Omega(\cdot)$ denote a functional from the set D into an Euclidean space.

LEMMA**:** *Suppose that*

(i) *there exists a linear functional,* $D\Omega(\cdot)$ *and a reminder functional* $R\Omega(\cdot)$ *such that*

$(i.a)$ *for all* $H \in D$, $\Omega(F+H) - \Omega(F) = D\Omega(F,H) + R\Omega(F,H)$

$(i.b)$ *for* $0 < a_1, a_2 < \infty$ *and all* $H \in D$ *for which* $\|H\|$ *is sufficiently small,*

$|D\Omega(F,H)| \leq a_1 \|H\|$ *and* $|R\Omega(F,H)| \leq a_2 \|H\|^2$

$(i.c)$ *for values* z^1 *and* z^2 *of subvectors* Z_1 *and* Z_2 *of* X, *which possess at least one common coordinate of* X *with distinct values, and for real valued functions* $r^1(y, z^1, x_{-1})$ *and* $r^2(y, z^2, x_{-2})$, *which are bounded and* *continuous a.e. and vanish outside a compact set*

$$D\Omega\left(F,H\right)=\sum_{q=1}^{2}\left[\int r^{q}(s,z^{q},x_{-q})\ h(s,z^{q},x_{-q})\ d(s,x_{-q})\right]$$

where for some $q\in\{0,1\}$, $\int r^{q}(s,z^{q},x_{-q})\ h(s,z^{q},x_{-q})\ d(s,x_{-q})\neq 0$.

For each q, *let* $j_q=-1$ *if, for all* h, $\int r^{q}(s,z^{q},x_{-q})\,h(s,z^{q},x_{-q})\ d(s,x_{-q})\neq 0$; *let* $j_q=0$ *otherwise. Let* $\widetilde{d}=\max\{d_q|q \text{ such that } j_q\geq 0\}$.

(ii) Assumptions C.1, C.2, and C.3 are satisfied with $s''\geq 2$, $s'\geq\widetilde{s}+s''$ *and* $\widetilde{s}\geq 0$.

(iii) As $N\to\infty$, $\sigma_N\to 0$, $\ln(N)/\left(N\sigma_N^{L+1}\right)\to 0$, $\sqrt{N\sigma_N^{d}}\left(\sqrt{\left(\ln(N)\right)/\left(N\sigma_N^{L+1}\right)}+\sigma^{s''}\right)^2\to 0$ *and for all* q *such that* $j_q\geq 0$, $\sqrt{N}\sigma_N^{d_q/2}\to\infty$ *and* $\sqrt{N}\sigma_N^{(d_q/2)+s''}\to 0$.

Then,

$$\Omega\left(\widehat{F}\right)\to\Omega\left(F\right)\qquad\text{in probability, and}$$

$$\sqrt{N\sigma_N^{\widetilde{d}}}\left(\Omega\left(\widehat{F}\right)\to\Omega\left(F\right)\right)\to N\left(0,V\right)\qquad\text{in distribution, where}$$

$$V=\left[\sum_{q=1}^{2}1\left[j_q\geq 0\right]1\left[d_q=\widetilde{d}\right]\left[\int K(s,z_q,x_{-q})dx_{-q}\right]^2\left[\int\left[r^{q}(s,z^{q},x_{-q})\right]^2\ f(s,z^{q},x_{-q})d(s,x_{-q})\right]\right]$$

PROOF: To show convergence in probability, we note that by $(ii),(iii)$ and Lemma B.3 in Newey (1994), $\left\|\widehat{F}-F\right\|\to 0$ in probability. Let $H=\widehat{F}-F$ be such that $\|H\|$ is sufficiently small. Since by $(i.a)$ and $(i.b)$, $\left|\Omega\left(\widehat{F}\right)-\Omega\left(F\right)\right|\leq a_1\left\|\widehat{F}-F\right\|+a_2\left\|\widehat{F}-F\right\|^2$, and, by above, $\left\|\widehat{F}-F\right\|\to 0$ in probability, the result follows.

To show the convergence in distribution result, for each q such that $j_q\geq 0$, let $D\Omega\left(F,H;z^q\right)=\left[\int r_k^{q}(s,z^{q},x_{-q})\ h(s,z^{q},x_{-q})\ d(s,x_{-q})\right]$. Let $H=\widehat{F}-F$. Then, by $(i.c)$, (ii), (iii), and Lemma 5.3 in Newey (1994),

$$\sqrt{N\sigma_N^{d_q}}D\Omega\left(F,\widehat{F}-F;z^q\right)\to N\left(0,V_q\right)\qquad\text{in distribution, where}$$

$$V_q = \left[\int K(s, z_q, x_{-q}) dx_{-q} \right]^2 \left[\int \left[r^q(s, z^q, x_{-q}) \right]^2 \; f(s, z^q, x_{-q}) d(s, x_{-q}) \right].$$

By $(i.a)$, $(i.b)$, (iii), and Lemma B.3 in Newey (1994), $\sqrt{N\sigma_N^{\widetilde{d}}} R\Omega\left(F, H\right) \rightarrow 0$ in probability. Hence,

$$\sqrt{N\sigma_N^{\widetilde{d}}} D\Omega\left(F, H\right) = \sqrt{N\sigma_N^{\widetilde{d}}} \left[\sum_{q=1}^{2} 1\left[j_q \geq 0\right] 1\left[d_q = \widetilde{d}\right] D\Omega\left(F, \widehat{F} - F; z^q\right) \right] + o_p(1)$$

The result will then follow once we show that $\sqrt{N\sigma_N^{\widetilde{d}}} \left[1\left[j_1 \geq 0\right] 1\left[d_1 = \widetilde{d}\right] D\Omega\left(F, \widehat{F} - F; z^1\right) \right]$ and $\sqrt{N\sigma_N^{\widetilde{d}}} \left[1\left[j_2 \geq 0\right] 1\left[d_2 = \widetilde{d}\right] D\Omega\left(F, \widehat{F} - F; z^2\right) \right]$ have asymptotic covariance equal to 0. Denote z^1 and z^2 by $z^1 = (w_0^1, w_1^1)$ and $z^2 = (w_0^2, w_2^2)$, where w_0^1 are w_0^2 are the values of the coordinates that z^1 and z^2 have in common. For each q and i, let $v_i^q = \left(\sigma^{L+1}\right)^{-1} \int r^q(s, z^q, x_{-q}) K\left(\frac{y_i - s}{\sigma}, \frac{(z_q)_i - z^q}{\sigma} \frac{(x_{-q})_i - x_{-q}}{\sigma}\right) ds$ dx_{-q}. Then, as is well known (see, for example, the proof of Lemma 5.3 in Newey (1994))

$$E(v_i^q) = \int r^q(s, z^q, x_{-q}) f(s, z^q, x_{-q}) ds dx_{-q} + O(\sigma^{s''}) \tag{1}$$

By the definition of $D\Omega\left(F, \widehat{F} - F; z^q\right)$, the covariance between $\sqrt{N\sigma_N^{d_1}} \left(D\Omega\left(F, \widehat{F} - F; z^1\right)\right)$ and $\sqrt{N\sigma_N^{d_2}} \left(D\Omega\left(F, \widehat{F} - F; z^2\right)\right)$equals

$$\frac{\sigma^{(d_1+d_2)/2}}{\sigma^{2(L+1)}} \left\{ E\left[\left(\int r^1 K^1\right) \left(\int r^2 K^2\right) \right] - E\left(\int r^1 K^1\right) E\left(\int r^2 \partial K^2\right) \right\} \text{ where for } q = 1, 2$$

$$\left(\int r^q K^q\right) = \int r^q(s, z^q, x_{-q}) K\left(\frac{s_i - s}{\sigma}, \frac{(z_q)_i - z^q}{\sigma}, \frac{(x_{-q})_i - (x_{-q})}{\sigma}\right) d(s, x_{-q})$$

Note that

$$E\left[\left(\int r^1 K^1\right) \left(\int r^2 K^2\right) \right] = \sigma^{2L+2-d_1-d_2} \int \left(\int \widetilde{r}^1 \widetilde{K}^1\right) \left(\int \widetilde{r}^2 \widetilde{K}^2\right) f(s_i, x^i) d\left(s_i, x^i\right)$$

where $\int \widetilde{r}^1 \widetilde{K}^1 = \int r^1(s_i - \sigma\widetilde{s}, z^1, w_2^i - \sigma\widetilde{w}_2, w_3^i - \sigma\widetilde{w}_3) K\left(\widetilde{s}, \frac{w_o^i - w_0^1}{\sigma}, \frac{w_1^i - w_1^1}{\sigma}, \widetilde{w}_2, \widetilde{w}_3\right) d\widetilde{s}$ $d\widetilde{w}_2\; d\widetilde{w}_3$

$\left(\int \widetilde{r}^2 \widetilde{K}^2\right) = \int r^2(s_i - \sigma\widetilde{s}, z^2, w_1^i - \sigma\widetilde{w}_1, w_3^i - \sigma\widetilde{w}_3) K\left(\widetilde{s}, \frac{w_o^i - w_0^2}{\sigma}, \widetilde{w}_1, \frac{w_2^i - w_2^2}{\sigma}, \widetilde{w}_3\right) d\widetilde{s}$ $d\widetilde{w}_1\; d\widetilde{w}_3$

and $x^i = (w_o^i, w_1^i, w_2^i, w_3^i)$. Let $t_u = \dim(w_u)$ for $u = 0, 1, 2, 3$. Then, $d_1 = t_0 + t_1$, $d_2 = t_0 + t_2$, $L = t_0 + t_1 + t_2 + t_3$ and

$$E\left[\left(\int r^1 K^1\right)\left(\int r^2 K^2\right)\right] =$$

$$= \sigma^{(t_1+t_2)/2} \int \left(\int \widehat{r}^1 \widehat{K}^1\right)\left(\int \widehat{r}^2 \widehat{K}^2\right) f(s_i, w_o^1 + \sigma\widehat{w}_0, w_1^1 + \sigma\widehat{w}_1, w_2^2 + \sigma\widehat{w}_2, w_3^i) d\left(s_i, \widehat{w}_o, \widehat{w}_1, \widehat{w}_2, w_3^i\right)$$

where $\int \widehat{r}^1 \widehat{K}^1 = \int r^1(s_i - \sigma\widetilde{s}, z^1, w_2^2 - \sigma(\widehat{w}_2 - \widetilde{w}_2), w_3^i - \sigma\widetilde{w}_3)\; K\left(\widetilde{s}, \widehat{w}_0, \widehat{w}_1, \widetilde{w}_2, \widetilde{w}_3\right)\; d\widetilde{s}\, d\widetilde{w}_2\, d\widetilde{w}_3$

and $\int \widehat{r}^2 \widehat{K}^2 = \int r^2(s_i - \sigma\widetilde{s}, z^2, w_1^1 - \sigma(\widehat{w}_1 - \widetilde{w}_1), w_3^i - \sigma\widetilde{w}_3) K\left(\widetilde{s}, \widetilde{w}_0 + \frac{w_0^1 - w_0^2}{\sigma}, \widetilde{w}_1, \widehat{w}_2, \widetilde{w}_3\right) d\widetilde{s}\, d\widetilde{w}_1\, d\widetilde{w}_3.$

It then follows by bounded convergence, (1), and (iii) that the covariance converges to 0.

This completes the proof of the Lemma.

PROOF OF THEOREM 1**:** Define the functional $\Lambda(\cdot)$ by $\Lambda(G) = G_{Y|W=w}(y)$. Then, $\Lambda(\widehat{F}) = \widehat{F}_{Y|W=w}(y)$ and $\Lambda(F) = F_{Y|W=w}(y)$. (We omit writing explicitly the dependence of Λ on y and w, for brevity of exposition.) For any H such that H vanishes outside a compact set and $\|H\|$ is sufficiently small, we have that, $|h(w)| \leq a\,\|H\|$, $\left|\int_{-\infty}^{y} h(s,w)ds\right| \leq a\,\|H\|$, and $|f(w) + h(w)| \geq b\,|f(w)|$ for some $0 < a, b < \infty$. Moreover,

(1) $\Lambda(F + H) - \Lambda(F) = (F + H)_{Y|W=w}(y) - F_{Y|W=w}(y) = D\Lambda(F, H) + R\Lambda(F, H)$, where

$$D\Lambda(F, H) = \frac{\int_{-\infty}^{y} h(s,w)ds - h(w)\; F_{Y|W=w}(y)}{f(w)} \text{ and } R\Lambda(F, H) = \left[\frac{\int_{-\infty}^{y} h(s,w)ds - h(w)\; F_{Y|W=w}(y)}{f(w)}\right]\left[\frac{h(w)}{f(w)+h(w)}\right]$$

Hence, for some $c < \infty$,

(2) $|D\Lambda(F, H)| \leq \frac{c}{f(w)}\,\|H\|$ and $|R\Lambda(F, H)| \leq \frac{c}{f(w)^2}\,\|H\|^2$.

Letting $z^1 = w$ and $r^2 \equiv 0$, it follows by the assumptions of the Theorem and the Lemma that $F_{Y|W=w}(y) = \Lambda(\widehat{F}) \to \Lambda(F) = F_{Y|W=w}(y)$ in probability and $\sqrt{N\sigma_N^L}\left(\widehat{F}_{Y|W=w}(y) - F_{Y|W=w}(y)\right) = \sqrt{N\sigma_N^{2L}}\left(\Lambda(\widehat{F}) - \Lambda(F)\right) \to N(0, V_F)$, where

$$V_F = \left\{\int \left(\int K(u,v)dv\right)^2 du\right\}\left\{\left(\frac{1}{f(w)^2}\right)\right\}\left\{\int \left[1\left[s \le y\right] - F_{Y|W=w}(y)\right]^2 f(s,w)ds\right\}$$
$$= \left\{\int \left(\int K(u,v)dv\right)^2 du\right\}\left(\frac{1}{f(w)^2}\right)\left[F_{Y|W=w}(y)\left(1 - F_{Y|W=w}(y)\right)\right],\ u \in$$

R^d and $v \in R^{1+L-d}$

PROOF OF THEOREM 2**:** Let W and $\widetilde{W}$ be two subvectors of X. Define the functional $\Phi(\cdot)$ by $\Phi(G) = G^{-1}_{Y|W=w}\left(G_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)$, where $G^{-1}_{Y|W=w}$ denotes an arbitrary element of the set $G^{-1}_{Y|W=w}$, if $G^{-1}_{Y|W=w}$ is not a singleton. Then, $\Phi(F) = n(w,e)$ and $\Phi(\widehat{F}) = \widehat{n}(w,e)$. Define the functionals η and ν by $\eta(G) = G_{Y|W=w}(\Phi(G))$, and $\nu(G) = G_{Y|W=\widetilde{w}}(\widetilde{e})$. Then, $\Phi(F)$ satisfies the equation: $\eta(F) = \nu(F)$ and, for any H, $\Phi(F+H)$ satisfies the equation: $\eta(F+H) = (F+H)_{Y|W=w}(\Phi(F+H)) = (F+H)_{Y|W=\widetilde{w}}(\widetilde{e}) = \nu(F+H)$. Let $\rho_1 > 0$ be such that if $\|H\| \le \rho_1$ then, for some $0 < a, b < \infty$, all y and all $s \in N(m(w,e),\xi)$,

(1) $|h(w)| \le a\,\|H\|\,,\ \left|\int_{-\infty}^{y} h(s,w)ds\right| \le a\,\|H\|\,,$

$|f(w) + h(w)| \ge b\,|f(w)|\,,$ and $\ f(s,w) + h(s,w) \ge b\,|f(s,w)|\,,$

and, by (1) and (2) in the proof of Theorem 1, for some $d < \infty$ and all w' such that $0 < f(w') < \infty$,

(2) $\sup_{y\in R}\left|(F+H)_{Y|W=w'}(y) - F_{Y|W=w'}(y)\right| \le \frac{d\|H\|}{f(w')}.$

Using arguments similar to those used in Matzkin and Newey (1993), we will show that there exist $\rho \le \rho_1$ such that if $\|H\| \le \rho$ then

(3) $(F+H)^{-1}_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})) \in N(m(w,e),\xi).$

To show (3), we let $r^* = F^{-1}_{Y|W=w}(F^{-1}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))$, $r = (F+H)^{-1}_{Y|W=w}(F^{-1}_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))$, and $s = F_{Y|W=w}(r)$, so that $r = F^{-1}_{Y|W=w}(s)$. Then,

$$r - r^* = (F+H)^{-1}_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})) - F^{-1}_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))$$

$$= F^{-1}_{Y|W=w}(s) - F^{-1}_{Y|X=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))$$

$$= \left(\frac{1}{f_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))}\right)(s - F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})) + \operatorname{Re} m_1$$

where, for some $j_1 < \infty$, $|\operatorname{Re} m_1| \leq j_1 \left| s - F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right|^2$, and where the last equality follows from Taylor's Theorem. Since $(s - F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})) = (F_{Y|W=w}(r) - (F+H)_{Y|W=w}(r)$, it follows from (2) that $|r - r^*| \leq \left|1/\left(f_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))\right)\right| d\,\|H\|/f(w) + \left(j_1 d^2 \|H\|^2\right)/f(w)^2$. Hence, if $\|H\|$ is sufficiently small, $|r - r^*| < \xi$, which implies that $(F+H)^{-1}_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})) \in N(m(w,e),\xi)$.

Consider then the $H's$ such that $\|H\| \leq \rho$. We will show, again using arguments similar to those used in Matzkin and Newey (1993) that for some $c_1 < \infty$,

$$(4)\ |\Phi(F+H) - \Phi(F)| \leq c_1 \|H\|.$$

For this we note that

$$(5)\quad \Phi(F+H) - \Phi(F)$$

$$= (F+H)^{-1}_{Y|W=w}\left((F+H)_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right) - F^{-1}_{Y|W=w}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)$$

$$= \left\{(F+H)^{-1}_{Y|W=w}\left((F+H)_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right) - (F+H)^{-1}_{Y|W=w}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)\right\}$$

$$+ \left\{(F+H)^{-1}_{Y|W=w}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right) - F^{-1}_{Y|W=w}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)\right\}$$

To obtain an expression for the difference in the first brackets of (5), we note that by Taylor's Theorem,

$$(F+H)^{-1}_{Y|W=w}\left((F+H)_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right) - (F+H)^{-1}_{Y|W=w}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)$$

$$= \frac{\partial (F+H)^{-1}_{Y|W=w}}{\partial r}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)\left[(F+H)_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) - F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right] + \operatorname{Re} m_1$$

where, for some $j_2 < \infty$, $|\mathrm{Re}\, m_2| \leq \left|(F+H)_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) - F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right|^2$. Hence, since

$$\left|\frac{\partial (F+H)^{-1}_{Y|W=w}}{\partial y}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)\right| = \left|\frac{1}{(f+h)_{Y|W=w}((F+H)^{-1}_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})))}\right|$$
$$= \left|\frac{f(w)+h(w)}{f((F+H)^{-1}_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})),w)+h((F+H)^{-1}_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})),w)}\right|$$

is bounded by (1) and (3), and, by (2), $\left|(F+H)_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}) - F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right| \leq d\,\|H\| \;/\; f(w)$,it follows that for some $a_2 < \infty$,

(6) $\left|(F+H)^{-1}_{Y|W=w}\left((F+H)_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right) - (F+H)^{-1}_{Y|W=w}\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)\right| \leq a_2\,\|H\|$.

To obtain an expression for the difference in the second brackets of (5), we note that by (1) and the Mean Value Theorem, $(F+H)_{Y|W=w}\left((F+H)^{-1}_{Y|W=w}(t)\right) - (F+H)_{Y|W=w}\left(F^{-1}_{Y|W=w}(t)\right) = \partial(F+H)_{Y|W=w}/\partial y(r_2)\left[(F+H)^{-1}_{Y|W=w}(t) - F^{-1}_{Y|W=w}(t)\right]$, where r_2 is between $(F+H)^{-1}_{Y|W=w}(t)$ and $F^{-1}_{Y|W=w}(t)$ and where $t = F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})$. Hence, since $(F+H)_{Y|W=w}\left((F+H)^{-1}_{Y|W=w}(t)\right) = t = F_{Y|W=w}\left(F^{-1}_{Y|W=w}(t)\right)$, it follows by (3) that

$$(F+H)^{-1}_{Y|W=w}(t) - F^{-1}_{Y|W=w}(t) = \frac{F_{Y|W=w}\left(F^{-1}_{Y|W=w}(t)\right) - (F+H)_{Y|W=w}\left(F^{-1}_{Y|W=w}(t)\right)}{(f+h)_{Y|W=w}(r_2)}.$$

It then follows by (2) that for some $a_3 < \infty$,

(7) $\left|(F+H)^{-1}_{Y|W=w}(t = F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})) - F^{-1}_{Y|W=w}(t = F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))\right| \leq a_3\,\|H\|$.

Hence, (4) follows by (5)-(7).

Next, we will obtain a first order Taylor expansion for $\Phi(F+H)$, using the fact that $\eta(F+H) - \eta(F) = \nu(F+H) - \nu(F)$. Let $\int^t$ denote $\int_{-\infty}^t$. By the definition of η,

$$\eta(F+H) - \eta(F) = (F+H)_{Y|W=w}(\Phi(F+H)) - F_{Y|W=w}(\Phi(F))$$
$$= \frac{\int^{\Phi(F+H)} f(s,w)\, ds + \int^{\Phi(F+H)} h(s,w)\, ds}{f(w)+h(w)} - \frac{\int^{\Phi(F)} f(s,w)\, ds}{f(w)}.$$

By the Mean Value Theorem, there exist r_f and r_h between $\Phi(F)$ and $\Phi(F+H)$ such that $\int^{\Phi(F+H)} f(s,w)\ ds - \int^{\Phi(F)} f(s,w)\ ds = f(r_f,w)\ (\Phi(F+H) - \Phi(F))$ and $\int^{\Phi(F+H)} h(s,w)\ ds - \int^{\Phi(F)} h(s,w)\ ds = h(r_h,w)\ (\Phi(F+H) - \Phi(F))$. Let $\Delta\Phi = \Phi(F+H) - \Phi(F)$. Then,

$$\eta(F+H)-\eta(F) = \frac{f(w)f(r_f,w)\Delta\Phi + f(w)h(r_f,w)\Delta\Phi + f(w)\int^{\Phi(F)} h(s,w)ds - h(w)\int^{\Phi(F)} f(s,w)ds}{f(w)(f((w)+h(w))}.$$

where, by (1), $f(w)+h(w)>0$. By the definition of ν,

$$\begin{aligned}\nu(F+H)-\nu(F) &= (F+H)_{Y|X=\widetilde{w}}(\widetilde{e}) - F_{Y|X=\widetilde{w}}(\widetilde{e})\\ &= \frac{\int^{\widetilde{e}} f(s,\widetilde{w})\ ds + \int^{\widetilde{e}} h(s,\widetilde{w})\ ds}{f(\widetilde{w})+h(\widetilde{w})} - \frac{\int^{\widetilde{e}} f(s,\widetilde{w})\ ds}{f(\widetilde{w})}\\ &= \frac{f(\widetilde{w})\int^{\widetilde{e}} h(s,\widetilde{w})\ ds - h(\widetilde{w})\int^{\widetilde{e}} f(s,\widetilde{w})\ ds}{f(\widetilde{w})(f(\widetilde{w})+h(\widetilde{w}))}.\end{aligned}$$ Let

$A\widetilde{w} = f(\widetilde{w})\int^{\widetilde{e}} h(s,\widetilde{w})ds - h(\widetilde{w})\int^{\widetilde{e}} f(s,\widetilde{w})ds$ and $Aw = f(w)\int^{\Phi(F)} h(s,w)ds - h(w)\int^{\Phi(F)} f(s,w)ds$. Then,

(8) $\eta(F+H)-\eta(F) = \left[\frac{f(r_f,w)+h(r_f,w)}{f(w)+h(w)}\right]\Delta\Phi + \frac{Aw}{f(w)(f((w)+h(w))}$, and

(9) $\nu(F+H)-\nu(F) = \frac{A\widetilde{w}}{f(\widetilde{w})(f(\widetilde{w})+h(\widetilde{w}))}$.

Since $\eta(F+H)-\eta(F) = \nu(F+H)-\nu(F)$, it follows from (8) and (9) that

$$\Delta\Phi = \frac{(f(w)+h(w))A\widetilde{w}}{f(\widetilde{w})(f(\widetilde{w})+h(\widetilde{w}))\big(f(r_f,w)+h(r_f,w)\big)} - \frac{Aw}{f(w)\big(f(r_f,w)+h(r_f,w)\big)}.$$

By the Mean Value Theorem, there exist r'_f, between $\Phi(F)$ and r_f, such that $f(r_f,w) - f(\Phi(F),w) = \partial f(r'_f,w)/\partial y\ \ (r_f - \Phi(F))$. Hence,

$$\Delta\Phi = \frac{(f(w)+h(w))A\widetilde{w}}{f(\widetilde{w})(f(\widetilde{w})+h(\widetilde{w}))\Big(f(\Phi(F),w)+\frac{\partial f(r'_f,w)}{\partial y}\big(r_f-\Phi(F)\big)+h(r_f,w)\Big)} - \frac{Aw}{f(w)\Big(f(\Phi(F),w)+\frac{\partial f(r'_f,w)}{\partial y}\big(r_f-\Phi(F)\big)+h(r_f,w)\Big)}.$$

Let

$D\Phi(F,H) = \frac{f(w)}{f(\widetilde{w})^2 f(\Phi(F),w)} A\widetilde{w} + \frac{f(w)}{f(w)^2 f(\Phi(F),w)} Aw$, and

$$R\Phi(F,H)=\left[\frac{(f(w)+h(w))}{f(\widetilde{w})(f(\widetilde{w})+h(\widetilde{w}))\Big(f(\Phi(F),w)+\frac{\partial f(r_f',w)}{\partial y}\big(r_f-\Phi(F)\big)+h(r_f,w)\Big)}-\frac{f(w)}{f(\widetilde{w})^2 f(\Phi(F),w)}\right]A\widetilde{w}$$
$$-\left[\frac{1}{f(w)\Big(f(\Phi(F),w)+\frac{\partial f(r_f',w)}{\partial y}\big(r_f-\Phi(F)\big)+h(r_f,w)\Big)}-\frac{1}{f(w)f(\Phi(F),w)}\right]Aw.$$

Then,

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

By the definition of $R\Phi(F,H)$,

$$R\Phi(F,H)=\left[\frac{f(\widetilde{w})^2 f(\Phi(F),w)h(w)-f(w)f(\widetilde{w})^2\frac{\partial f(r_f',w)}{\partial y}(r_f-\Phi(F))-f(w)f(\widetilde{w})^2 h(r_h,w)}{f(\widetilde{w})^2(f(\widetilde{w})+h(\widetilde{w}))f(\Phi(F),w)\big(f(r_f,w)+h(r_f,w)\big)}\right]A\widetilde{w}$$
$$-\left[\frac{f(w)f(\widetilde{w})h(\widetilde{w})f(\Phi(F),w)+f(w)f(\widetilde{w})h(\widetilde{w})\frac{\partial f(r_f',w)}{\partial y}(r_f-\Phi(F))+f(w)f(\widetilde{w})h(\widetilde{w})h(r_h,w)}{f(\widetilde{w})^2(f(\widetilde{w})+h(\widetilde{w}))f(\Phi(F),w)\big(f(r_f,w)+h(r_f,w)\big)}\right]A\widetilde{w}$$
$$+\left[\frac{\frac{\partial f(r_f',w)}{\partial y}\big(r_f-\Phi(F)\big)+h(r_f,w)}{f(w)f(\Phi(F),w)\big(f(r_f,w)+h(r_f,w)\big)}\right]Aw.$$

Since, by the definition of r_f and by (8), $|r_f-\Phi(F)|\leq|\Phi(F+H)-\Phi(F)|\leq c_1\|H\|$, it follows by (1) that, for some $a_4<\infty$, $|R\Phi(F,H)|\leq a_4\|H\|^2$. Moreover, by the definition of $D\Phi(F,H)$, there exists $a_5<\infty$ such that $|D\Phi(F,H|\leq a_5\|H\|$. It then follows by the assumptions of the Theorem and the Lemma that $\hat{m}(w,e)-m(w,e)=\Phi(\hat{F})-\Phi(F)\to 0$ in probability and $\sqrt{N}\,\sigma_N^{\widetilde{d}/2}\left(\hat{m}(w,e)-m(w,e)\right)=\sqrt{N}\,\sigma_N^{\widetilde{d}/2}\left(\Phi(\hat{F})-\Phi(F)\right)\to N(0,V_n)$ where $\widetilde{d}=\max\{d_1,d_2\}$ and $V_n=\left\{\int K(u)^2\right\}\left[\left(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\ \left(1-F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right)\right)/f_{Y|W=w}(n(w,e))^2\right]$ $\left[1[d_1=\widetilde{d}]/f(\widetilde{w})+1[d_2=\widetilde{d}]/f(w)\right]$.

PROOF OF THEOREM 3: We consider the case where Assumptions (A.i)-(A.vi) are satisfied. The case where Assumptions (A.i)-(A.iv), (A.v') and (A.vi') are satisfied can be analyzed in a similar way. Without loss of generality, we will show the identification of the distribution of ε_1, conditional on $X_0=w_0$. Given $\eta\in R$, let $y=r_1(\eta)$. Note that when $X=(w_0,\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)=(w_0,\overline{w}^k)$, $Y=m(X,\varepsilon)=r_1(\varepsilon_1)$. Hence,

$$\Pr\left(Y \leq y | X = (w_0, \overline{w}^k)\right) = \Pr\left(r_1(\varepsilon_1) \leq r_1(\eta) | X = (w_0, \overline{w}^k)\right) = \Pr\left(\varepsilon_1 \leq \eta | X_0 = w_0\right)$$

where the last equality follows by Assumption (A.vi). Hence, the marginal distribution of ε_1, conditional on X_0, is identified from the conditional distribution of Y, when $X = (w_0, \overline{w}_k)$. Using similar arguments, we can conclude that the marginal distribution of each ε_k, conditional on W_0, is identified from the conditional distribution of Y when $X = (w_0, w_1, w_2, ..., w_K)$ is such that $w_k = \overline{w}_k$ and $w_j = \widetilde{w}_j$ for $j \neq k$. By Assumption (A.v), the distribution of ε conditional on X_0 is the multiplication of the marginal distributions, conditional on X_0. Hence, $F_{\varepsilon|X_0}$ is identified..

Next, we show that the functions n_k are identified. Again, without loss of generality, we take $k = 1$. Note that when $X = (w_0, w_1, \widetilde{w}_2, ..., \widetilde{w}_K) = (w_0, w^k)$, $Y = m(X, \varepsilon) = r_1\left(n_1(w_{0_1}, w_1, \varepsilon_1)\right)$. Hence, using the conditional independence between ε and X_1, and the strict monotonicity of n_1 in ε_1 it follows that

$\Pr(\varepsilon_1 \leq \eta | X_0 = w_0) = \Pr(\varepsilon_1 \leq \eta | X = (w_0, w^k))$

$= \Pr(n_1(w_{0_1}, w_1, \varepsilon_1) \leq n_1(w_{0_1}, w_1, \eta) | X = (w_0, w^k))$

$= \Pr\left(r_1\left(n_1(w_{0_1}, w_1, \varepsilon_1)\right) \leq r_1\left(n_1(w_{0_1}, w_1, \eta)\right) | X = (w_0, w^k)\right)$

$= \Pr\left(Y \leq r_1\left(n_1(w_{0_1}, w_1, \eta)\right) | X = (w_0, w^k)\right)$.

Since, as we have shown above, $\Pr\left(\varepsilon_1 \leq \eta | W_0 = w_0\right) = \Pr\left(Y \leq r_1(\eta) | X = (w_0, \overline{w}^k)\right)$ it follows that $F_{Y|X=(w_0,\overline{w}^k)}\left(r_1(\eta)\right) = F_{Y|X=(w_0,w^k)}\left(r_1\left(n_1(w_{0_1}, w_1, \eta)\right)\right)$. Partition w_0 as $w_0 = (w_{0_1}, w_{0_{-1}})$. Note that

$$F_{Y|X=(w_0,w^k)}\left(r_1\left(n_1(w_{0_1}, w_1, \eta)\right)\right)$$

$$= \int^{r_1\left(n_1(w_{0_1},w_1,\eta)\right)} f(s, w_0, w_1, \widetilde{w}_2, ..., \widetilde{w}_K) \ / \ f(w_0, w_1, \widetilde{w}_2, ..., \widetilde{w}_K) ds$$

$$= \int^{r_1\left(n_1(w_{0_1},w_1,\eta)\right)} \frac{f(s,w_0,w_1,\widetilde{w}_2,...,\widetilde{w}_K)}{f(w_0,w_1,\widetilde{w}_2,...,\widetilde{w}_K)} \left[\int \frac{f(w_{0_1},w_{0_{-1}},w_1,\widetilde{w}_2,...,\widetilde{w}_K)}{f(w_{0_1},w_1,\widetilde{w}_2,...,\widetilde{w}_K)} dw_{0_{-1}}\right] ds$$

$$= \int^{r_1\left(n_1(w_{0_1},w_1,\eta)\right)} \int \frac{f(s,w_{0_1},w_{0_{-1}},w_1,\widetilde{w}_2,...,\widetilde{w}_K)}{f(w_{0_1},w_1,\widetilde{w}_2,...,\widetilde{w}_K)} dw_{0_{-1}} ds$$

$$= \int^{r_1\left(n_1(w_{0_1},w_1,\eta)\right)} \frac{f(s,w_{0_1},w_1,\widetilde{w}_2,...,\widetilde{w}_K)}{f(w_{0_1},w_1,\widetilde{w}_2,...,\widetilde{w}_K)} ds$$

$$= F_{Y|X=(w_{0_1},w^k)}\left(r_1\left(n_1(w_{0_1}, w_1, \eta)\right)\right)$$

Hence, since $F_{Y|X=(w_0,\overline{w}^k)}(r_1(\eta)) = F_{Y|X=(w_{0_1},w^k)}\left(r_1\left(n_1(w_{0_1},w_1,\eta)\right)\right)$ it follows that $n_1(w_{0_1},w_1,\eta) = F^{-1}_{Y|X=(w_{0_1},w^k)}\left(F_{Y|X=(w_0,\overline{w}^k)}\left(r_1(\eta)\right)\right)$. This completes the proof of the first part of the theorem.

REFERENCES

AÏT-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.

ALTONJI, J.G. and R.L. MATZKIN (1997): "Panel Data Estimators for Nonseparable Models with Endogenous Regressors," mimeo, Northwestern University.

AZZALINI, A. (1981): "A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method," *Biometrika*, 68, 1, 326-328.

BAJARI, P. and l. BENKARD (2001): "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," mimeo, Stanford University.

BARROS, R. (1986): "The Identifiability of Hazard Functions", working paper, University of Chicago.

BARROS, R. and B. HONORÉ (1988): "Identification of Duration Models with Unobserved Heterogeneity", working paper, Northwestern University.

BIERENS, H.J. (1987): "Kernel Estimators for Regression Functions," in *Advances of Econometrics*, ed. by T.F. Bewley. Cambridge: Cambridge University Press, Vol. I, 99-144.

BLUNDELL, R. and J.L. POWELL (2000): “Endogeneity in Nonparametric and Semiparametric Regression Models,” in *Advances in Economics and Econometrics: Theory and Applications, Eigth World Congress, Vol, II*, ed. by M. Dewatripont, L.P. Hansen, and S.J. Turnovsky, Cambridge: Cambridge University Press.

BOX, G.E.P. and D.R. COX (1964): “An Analysis of Transformations,” *Journal of the Royal Statistical Society*, Series B, 26, 211-243.

BRIESCH, R., P. CHINTAGUNTA, and R.L. MATZKIN (1997): “Nonparametric Discrete Choice Models with Unobserved Heterogeneity,” mimeo, Northwestern University.

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.

BROWN, B.W. and M.B. WALKER (1995): “Stochastic Specification in Random Production Models of Cost-Minimizing Firms”, *Journal of Econometrics*, 66, (1-2).

CHESHER, A. (2002a): “Local Identification in Nonseparable Models,” cemmap working paper # CWP05/02.

CHESHER, A. (2002b): “Instrumental Values,” cemmap working paper # CWP17/02.

ELBERS, C. and G. RIDDER (1982): “True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model,” *Review of Economic Studies*, July 1982, 49, 403-409.

HAN, A.K. (1987a): “Nonparametric Analysis of a Generalized Regression Model,” *Journal of Econometrics*, 35, 303-316.

HAN, A.K. (1987b): "A Nonparametric Analysis of Transformations", *Journal of Econometrics*, 35, 191-209.

HECKMAN, J.J. (1974): "Effects of Day-Care Programs on Women's Work Effort," *Journal of Political Economy.*

HECKMAN, J.J. (1991): "Identifying the Hand of Past: Distinguishing State Dependence from Heterogeneity," *The American Economic Review*, Vol. 81, No. 2, Papers and Proceeding of the Hundred and Third Annual Meeting of the American Economic Association, 75-79.

HECKMAN, J., R.L. MATZKIN and L. NESHEIM (2002): "Nonparametric Estimation of Nonadditive Hedonic Models," mimeo, Northwestern University.

HECKMAN, J. and B. SINGER (1984a): "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271-320.

HECKMAN, J. and B. SINGER (1984b): "The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, July 1984, 51, 231-243.

HECKMAN, J.J. and E.J. VYTLACIL (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Science*, Vol 96.

HECKMAN, J.J. and E.J. VYTLACIL (2001): "Local Instrumental Variables," in *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. Powell, Cambridge: Cambridge University Press.

HECKMAN, J.J. and R. WILLIS (1977): "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women," *Journal of Political Economy.*

HONG H. and M. SHUM (2001): "A Semiparametric Estimator for Dynamic Optimization Models, with an Application to a Milk Quota Market," mimeo, Princeton University.

HOROWITZ, J.L. (1996): “Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable,” *Econometrica*, 64, 103-137.

IMBENS, G.W. AND W.K. NEWEY (2001): “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” mimeo, UCLA.

LANCASTER, T. (1979): “Econometric Methods for the Analysis of Unemployment,” *Econometrica,* 47, 939-956.

LEWBEL, A. (1996): “Demand Systems With and Without Errors: Reconciling Econometric, Random Utility, and GARP Models,” mimeo, Brandeis University, September 1996.

LINTON, O. and J.P. NIELSEN (1995): “A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration,” *Biometrika,* 82, 1, 93-100.

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.

McFADDEN, D. (1974): “Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka, ed. *Frontiers in Econometrics,* pp105-142, New York: Academic Press.

MATZKIN, R.L. (1992): “Nonparametric and Distribution-Free Estimation of the Threshold Crossing and Binary Choice Models,” *Econometrica*, 60, 239-270.

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. (1999): "Nonparametric Estimation of Nonadditive Random Functions," mimeo, Northwestern University.

MATZKIN, R.L. (2002): "Estimation of Nonadditive Functions with Endogenous Regressors Using Functional Restrictions," mimeo, Northwestern University.

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

NADARAYA, E.A. (1964): "Some New Estimates for Distribution Functions," *Theory Prob. Applic.,* 15, 497-500.

NADARAYA, E.A. (1964): "On Estimating Regression," *Theory Prob. Applic.*, 9, 141-142.

NEWEY, W.K. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233-253.

OLLEY, G.S. and A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, Vol. 64, 6, 1263-1297.

RIDDER, G. (1990): "The Nonparametric Identification of Generalized Accelerated Failure-Time Models", *Review of Economic Studies,* May 1990, 57, 167-181.

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

SCHUSTER, E.F. (1972): "Joint Asymptotic Distribution of the Estimated Regression Function at a Finite Number of Distinct Points," *Annals of Mathematical Statistics* 43, 84-88.

van der VAART, A.W. and J.A. WELLNER (1996): *Weak Convergence and Empirical Processes, with Applications to Statistics*, Springer Series in Statistics.

VYTLACIL, E. (2001): "Semiparametric Identification of the Average Treatment Effect in Nonseparable Model," mimeo, Stanford University.

WATSON, G.S. (1964): "Smooth Regression Analysis," *Sankhya A,* 26, 359-72.

TABLE I

BANDWIDTHS

	N=250		N=500	
	σ_Y	σ_X	σ_Y	σ_X
Design I	.4497	.3239	.4031	.2928
Design II	.0650	.3050	.0596	.2619

TABLE II

DESIGN I, N=250

	$NPNA$			NW			LS		
	$abs(bias)$	var	mse	$abs(bias)$	var	mse	$abs(bias)$	var	mse
m	.1284	.0691	.0961	.0981	.0191	.0324	.0048	.0049	.0050
F_ε	.0072	.0011	.0012	.0220	.0005	.0010	.0166	.0003	.0006

TABLE III

DESIGN I, N=500

	$NPNA$			NW			LS		
	$abs(bias)$	var	mse	$abs(bias)$	var	mse	$abs(bias)$	var	mse
m	.0986	.0409	.0564	.0668	.0105	.0171	.0056	.0024	.0025
F_ε	.0075	.0007	.0007	.0186	.0003	.0007	.0137	.0002	.0004

TABLE IV

DESIGN II, N=250

	$NPNA$			NW			LS		
	$abs(bias)$	var	mse	$abs(bias)$	var	mse	$abs(bias)$	var	mse
m	.1048	.1077	.1606	.6035	.0160	.5877	.8417	.0050	1.1455
F_{ε}	.0404	.0019	.0037	.1081	.0016	.0213	.0379	.0001	.0020

TABLE V

DESIGN II, N=500

	$NPNA$			NW			LS		
	$abs(bias)$	var	mse	$abs(bias)$	var	mse	$abs(bias)$	var	mse
m	.0800	.1022	.1285	.6030	.0155	.5816	.8408	.0025	1.1433
F_{ε}	.0324	.0012	.0023	.1104	.0013	.0215	.0293	.0002	.0014