DEPARTAMENTO DE ECONOMÍA
DOCUMENTO DE TRABAJO

# "Nonparametric Estimation of Nonadditive Random Functions"

Rosa Matzkin

D.T.: N° 38 Septiembre 2001

Vito Dumas 284, (B1644BID) Victoria, Pcia. de Buenos Aires, Argentina
Tel: 4725-7053 Fax: 4725-7010
Email: economia@udesa.edu.ar

# NONPARAMETRIC ESTIMATION OF NONADDITIVE RANDOM FUNCTIONS


Rosa L. Matzkin*

Northwestern University
and
Universidad de San Andrés



First version: April 1999
This version: September 2001



**Abstract**

We present estimators for nonparametric functions that depend on unobservable random variables in nonadditive ways. The distributions of the unobservable random terms are assumed to be unknown. We show how properties that may be implied by economic theory, such as monotonicity, homogeneity of degree one, and separability can be used to identify the unknown, nonparametric functions and distributions. We also present convenient normalizations, to use when the properties of the functions are unknown. The estimators for the nonparametric distributions and for the nonparametric functions and their derivatives are shown to be consistent and asymptotically normal. The results of a limited simulation study are presented.



**Keywords:** nonparametric estimation, nonadditive random term, nonseparable models, shape restrictions, conditional distributions, kernel estimators


---


*The research presented in this paper was partially supported by NSF. I am grateful to three referees and to Richard Blundell for their insightful comments and suggestions, which help generate a substantially improved paper. I have greatly benefitted from my interaction with Joseph Altonji, James Heckman, and Whitney Newey, from the research assistance of Ulrich Doraszelski and Elie Tamer, and from the comments of Charles Manski, José Sheinkman, Christopher Taber, and seminar participants at New York University, Northwestern University, Universidad de San Andrés, Universidad Di Tella, 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, 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, $\varepsilon$, to the relationship. In the resulting model, $\varepsilon$ 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, $\varepsilon$ may represent an heterogeneity parameter in a utility function, some productivity shock in a production function, or some other relevant unobservable variable (see, for example, Heckman (1974), Heckman and Willis (1974), and Lancaster (1979)). When using this approach, the random term $\varepsilon$ 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 $\varepsilon$ is nonlinear in $\varepsilon$.

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, $\varepsilon$. The value of a dependent variable, $Y$, is assumed to be determined by this nonparametric function. The distribution of $\varepsilon$ is not parametrically specified and it is also estimated.

We first consider the model $Y = m(X,\varepsilon)$, where $\varepsilon$ is a random variable, $m$ is strictly increasing in $\varepsilon$, and both the function $m$ and the distribution of $\varepsilon$ are unknown. We characterize the set of functions that are observationally equivalent to $m$, when $\varepsilon$ is independent of $X$, and provide three different specifications for the function $m$, which allow one to identify the distribution of $\varepsilon$ 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 $x$. The second specification imposes an homogeneity of degree one condition, along a given ray, on some coordinates of $X$ and $\varepsilon$. 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 $\varepsilon$ 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 $\Lambda$ is a strictly increasing, unknown function, and where both, the absolute value of one of the coordinates of $\beta$ and the value of $\Lambda$ 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 $\varepsilon$ is independent of only some coordinates of $X$, conditional on the other coordinates. A special case of this is, of course, when $\varepsilon$ 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.

For each of the specifications and assumptions on the distribution of $\varepsilon$, we show that the estimator for the distribution of $\varepsilon$ 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 $\varepsilon$, 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 $\varepsilon$, the function $m$, and the derivatives of $m$ 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 consider nonparametric models where the random terms do not enter in an additive form are Roehrig (1988), Brown and Matzkin (1996), Altonji and Ichimura (1997), Altonji and Matzkin (1997), Heckman and Vytlacil (1999, 2001), Vytlacil (2000), Bajari and Benkard (2001), and Imbens and Newey (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). Altonji and Ichimura (1997) consider models with one dependent variable, and estimate an average derivative. Altonji and Matzkin (1997) consider the estimation of models for panel data. Heckman and Vytlacil (1999, 2001) and Vytlacil (2000) study models where potential outcomes are nonadditive in unobservable random terms. Bajari and Benkard consider the identification and estimation of nonadditive price functions in hedonic models. Imbens and Newey (2001) study the estimation of a triangular, nonseparable simultaneous equations model.

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 $\varepsilon$, together with their asymptotic properties. The results are extended to functions that depend on a multidimensional unobservable random term $\varepsilon$, in Section 4. Estimators for the derivatives of $m$ are studied in Section 5. Section 6 presents the results of some simulations. A short summary is presented in Section 7. Appendix A contains most of the proofs of the main theorems, while Appendix B presents, in a Lemma, previously obtained results, which are used in the proofs given in Appendix A.

## 2. The Model

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

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

where $m : A \times E \to R$ is continuous in $(X, \varepsilon)$ and strictly increasing in $\varepsilon$, $A \subset R^L$ is the support of $X$, $E \subset R$ is the support of $\varepsilon$, $Y$ and $X$ are observable, and $\varepsilon$ is an unobservable random term which is distributed, with a distribution $F_\varepsilon$, independently (or conditionally independently) of $X$. Many widely used type of models fall into this category. Models where $\varepsilon$ represents unobserved heterogeneity or a technological shock may satisfy model (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 $\eta$ is not independent of $X$, but $\eta = s(X, \varepsilon)$ where $\varepsilon$ is independent of $X$, then $Y = r(X, s(X, \varepsilon)) = m(X, \varepsilon)$.

Some transformation models satisfy (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 $\Lambda$ is an unknown, strictly increasing function and $\varepsilon$ is distributed independently of $X$ with an unknown distribution, satisfy model (1).

Duration models, where $Y$ denotes time in a state and $\varepsilon$ is the negative of the log-integrated hazard function, fall into the category of model (1), even when the hazard function is not separable in any of its arguments. In this case, $\varepsilon$ 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 (1), when the conditional hazard function is multiplicative in the unobserved heterogeneity variable. Let $\theta$ 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 $(X, \theta)$, and, hence, $\varepsilon$ 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 $r_2(X)$.)

The first question that arises when specifying the model in (1) is whether one can identify the function $m$ and the distribution of $\varepsilon$. Following the standard definition of identification, we say that $(m, F_\varepsilon)$ 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 $\Gamma$ denote a set to which $F_\varepsilon$ 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* $(m, F_\varepsilon)$ *is identified in the set* $(M \times \Gamma)$ *if*
(i) $(m, F_\varepsilon) \in (M \times \Gamma)$ *and (ii) for all* $(m', F'_\varepsilon)$ *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, $F'_\varepsilon$ and $F_\varepsilon$" in $\Gamma$ such that the pairs $(m', F'_\varepsilon)$ 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* $\Gamma$ *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 $(m, F_\varepsilon)$ in model (1), we first note that, since $m$ is strictly increasing in $\varepsilon$, 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, $(v, F_\varepsilon)$ is identified if and only if $(m, F_\varepsilon)$ is identified. Let $\Gamma$ 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 $(v, F_\varepsilon)$ 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:** $v, v' \in V$ are observationally equivalent if and only if there exists a strictly increasing function $g : R \to R$ such that $v' = g \circ v$.

**Proof of Lemma 1:** Note that, by the definition of $v$ and the independence between $\varepsilon$ and $X$,

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

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

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

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

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

Hence, $v$ and $v'$ are observationally equivalent. This completes the proof.

The lemma states that the function $v$ is identified up to a monotone transformation, $g$. For any such transformation, $(g \circ v, F_\varepsilon \circ g^{-1})$ and $(v, F_\varepsilon)$ generate the same distribution of $(Y, X)$. To see what this means in terms of the inverse function $m$, suppose that $m^*$ and $F^*_\varepsilon$ 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 $F^*_\varepsilon$ and $y = m^*(x, \varepsilon)$. Let $g$ be any strictly increasing transformation. Let $\varepsilon' = g(\varepsilon)$ and $v'(x, y) = g(v^*(x, y))$. The lemma implies that the model $\varepsilon' = g(\varepsilon) = g(v^*(x, y)) = v'(x, y)$ generates the same distribution of the observable variables as the model $\varepsilon = v^*(x, y)$. Let $m'$

denote the inverse function of $v'$, conditional on $x$. Then, for any value $e$, $m'(x,e)$ denotes the value of $y$ that satisfies $e = v'(x,y)$. Let then $\varepsilon'$ and $x$ be given. To find such a value of $y$, we note that since $\varepsilon' = v'(x,y) = g(v^*(x,y))$, $v^*(x,y) = g^{-1}(\varepsilon')$. Hence, since $m^*$ is the inverse of $v^*$, conditional on $x$, $y = m^*(x, g^{-1}(\varepsilon'))$. This shows that $m'(x,\varepsilon') = m^*(x, g^{-1}(\varepsilon'))$, or, since $\varepsilon' = g(\varepsilon)$, $m'(x, g(\varepsilon)) = m^*(x,\varepsilon)$. Hence, $m'$ and $m^*$ are observationally equivalent if and only if $m'$ equals $m^*$ with $\varepsilon$ substituted by $g(\varepsilon)$, for some strictly increasing function $g$, that is

$$m'(x, g(\varepsilon)) = m^*(x,\varepsilon).$$

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 accounts to fixing the values of the function $m$ at some value of the vector $X$. If for example, $m(\overline{x}, \varepsilon) = \varepsilon \cdot \overline{x}$, then this is satisfied for $\overline{x} = 1$. 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 $X_1 \neq \overline{x}_1$.

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 $\varepsilon$ 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 $\lambda \geq 0$

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

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 $\lambda \geq 0$

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

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.

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

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

This specification can be seen as a nonparametric, partially nonadditive generalization of the transformation model studied in Horowitz (1996), where $Y = \Lambda^{-1}(\beta' X + \varepsilon)$, $\Lambda^{-1}$ is unknown and strictly increasing, and the distribution of $\varepsilon$ is unknown. In Horowitz (1996), the value of $\Lambda$ 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 MdElroy (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 $X_{11} \neq \overline{x}_{11}$.

Other specifications could be developed by using, as it was exemplified in the last paragraphs, the result of Lemma 1.

# 3. Estimation of the Basic Model

To develop estimators for the function $m$ and the distribution of $\varepsilon$ 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 the 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**: $\varepsilon$ is independent of $X_1$, conditional on $X_0$.

**Assumption I.2:** For all $X$, $m$ is strictly increasing in $\varepsilon$.

Assumption I.1 guarantees that, conditional on $X_0$, the distribution of $\varepsilon$ 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 $\varepsilon$ 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

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

This is because $F_{\varepsilon|X_0=x_0}(e) = \Pr\left(\varepsilon \le e|X_0 = x_0\right) = P\left(\, \varepsilon \le e|X_0 = x_0, X_1 = x_1\right) = \Pr(\, m(X,\varepsilon) \le m(x,e) \mid X = x\,) = \Pr\left(Y \;\le\; m(x,e) \mid X = x\right) = F_{Y|X=x}\left(m(x,e)\right)$. The first equality follows by the definition of $F_\varepsilon$, the second follows by the conditional independence between $\varepsilon$ 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 $\varepsilon$ 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 $X$.

## 3.1. Specification I

Consider first the case where

(I.1) $m(x_0,\overline{x}_1,\varepsilon) = \varepsilon$ for some $\overline{x}_1$ and all $x_0$, and Assumptions I.1 and I.2 are satisfied.

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

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

Hence, the conditional distribution of $\varepsilon$ given $X_0$ is the value of 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, 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 $(x_0,x_1)$,

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

Suppose, next that

(I.2) $m(x_0,\overline{x}_1,\varepsilon) = \varepsilon$ for some $\overline{x}_1$ and all $x_0$, and Assumptions I.1' and I.2 are satisfied, where

**Assumption I.1':** $\varepsilon$ is independent of $(X_0, X_1)$.

then, we have that

$$(3.1') \quad F_\varepsilon(e) = F_{Y|X=x}\left(m(x,e)\right) \qquad \text{for all } e \in E \text{ and } x \in A,$$

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

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

Expression (3.1') follows because $F_\varepsilon(e) = P\left(\varepsilon \leq e | X = x\right) = \Pr\left(m(x,\varepsilon) \leq m(x,e) \mid X = x\right) = \Pr\left(Y \leq m(x,e) \mid X = x\right) = F_{Y|X=x}\left(m(x,e)\right)$. Expression (3.2') follows by, first, letting $X_1 = \overline{x}_1$ in (3.1'), so that for all $e$ and $x_0$,

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

Then, since $\int f(x_0|X_1 = \overline{x}_1)\, dx_0 = 1$, it follows that

$$\begin{aligned} F_\varepsilon(e) &= \int F_{Y|X=(x_0,\overline{x}_1)}\left(e\right)\, f(x_0|X_1 = \overline{x}_1)\, dx_0 \\ &= \int \int_{-\infty}^{e} \frac{f(s,x_0,\overline{x}_1)}{f(x_0,\overline{x}_1)} \frac{f(x_0,\overline{x}_1)}{f(\overline{x}_1)}\, ds\, dx_0 \\ &= \int_{-\infty}^{e} \frac{f(s,\overline{x}_1)}{f(\overline{x}_1)} ds \\ &= F_{Y|X_1=\overline{x}_1}\left(e\right). \end{aligned}$$

Expression (3.3') then follows from (3.1') and (3.2').

Finally, suppose that

(I.3) $m(\overline{x}_1, \varepsilon) = \varepsilon$ for some $\overline{x}_1$, and Assumptions I.1' and I.2 are satisfied.

Then, following arguments similar to the ones given above, we have that

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

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

## 3.2. Specification II

Consider next the case where

(II.1) Assumptions I.1' and I.2 are satisfied, and for some $\overline{x}_1, \overline{\varepsilon}$, some $\alpha \in R$, all $x_0$, and all $\lambda$ such that $\lambda\overline{\varepsilon} \in E$

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

Then, given any $\lambda$ and letting $x_1 = \lambda\overline{x}_1$ and $e = \lambda\overline{\varepsilon}$, we have, from (3.1), that for all $x_0$, $F_{\varepsilon|X_0=x_0}(\lambda\overline{\varepsilon}) = F_{Y|X=(x_0,\lambda\overline{x}_1)}\left(m(x_0, \lambda\overline{x}_1, \lambda\overline{\varepsilon})\right) = F_{Y|X=(x_0,\lambda\overline{x}_1)}(\lambda\alpha)$, 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 $e \in E$,

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

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

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

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

Next, suppose that

(II.2) Assumptions I.1 and I.2 are satisfied, and for some $\overline{x}_1$, $\overline{\varepsilon}$, some $\alpha \in R$, all $x_0$ and all $\lambda$ such that $\lambda\overline{\varepsilon} \in E$,

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

Then, using the same reasoning as used for the case where $m(x_0, \overline{x}_1, \varepsilon) = \varepsilon$, we will have that (3.1') is satisfied, as well as

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

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

When the specification is given by

(II.3) Assumptions I.1 and I.2 are satisfied, and for some $\overline{x}_0, \overline{x}_1, \overline{\varepsilon}$, $\alpha \in R$, and all $\lambda$ such that $\lambda\overline{\varepsilon} \in E$

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

we can show that

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

which together with the specification implies that

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

and for all $X = (x_0, x_1)$,

$$(3.5'')\ \ m(x,e) = F^{-1}_{Y|X=x}\left(F_{Y|X=(\overline{x}_0,(e/\overline{\varepsilon})\overline{x})}\left((e/\overline{\varepsilon})\alpha\right)\right).$$

## 3.3. Specification III

Finally, we consider the case where for some unknown function $s(\cdot)$,

(III.1) $m(x_0, x_{11}, x_{12}, \varepsilon) = s(x_0, x_{11}, \varepsilon - x_{12})$, for all $x_0$, $s(x_0, \overline{x}_{11}, \alpha) = \overline{y}$, and Assumptions I.3 and I.4 are satisfied,

where

**Assumption I.3:** $\varepsilon$ 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,

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

since $F_{\varepsilon|X_0=x_0}(e) = \Pr\left(\varepsilon \leq e|X_0 = x_0\right) = P\left(\ \varepsilon \leq e|(X_0, X_1) = (x_0, x_1)\right) = \Pr(\ \varepsilon - X_{12} \leq e - x_{12} \mid (X_0, X_1) = (x_0, x_1)\ ) = \Pr\left(s(X_0, X_{11}, \varepsilon - X_{12}) \leq s(x_0, x_{11}, \varepsilon - x_{12}) \mid X = x\right) = F_{Y|X=x}\left(s(x_0, x_{11}, e - x_{12})\right).$

Letting $X_{11} = \overline{x}_{11}$ and $X_{12} = e - \alpha$, in (3.6), we get that

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

Hence, the conditional distribution of $\varepsilon$ given $X_0$ is 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

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

If, we consider

**Assumption I.3':** $\varepsilon$ is independent of $X$

and the specification is

(III.2) $m(x_0, x_{11}, x_{12}, \varepsilon) = s(x_0, x_{11}, \varepsilon - x_{12})$, for all $x_0$, $s(x_0, \overline{x}_{11}, \alpha) = \overline{y}$, and Assumptions I.3' and I.4 are satisfied,

then, we can average out over $x_0$, using the conditional pdf of $X_0$ given $X_1$, to get

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

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

If the specification is

(III.3) $m(x_0, x_{11}, x_{12}, \varepsilon) = s(x_0, x_{11}, \varepsilon - x_{12})$, for some $(\overline{x}_0, \overline{x}_{11})$, $s(\overline{x}_0, \overline{x}_{11}, \alpha) = \overline{y}$, and Assumptions I.3' and I.4 are satisfied,

then,

$$(3.7'') \; F_\varepsilon(e) = F_{Y|X=(\overline{x}_0,\overline{x}_{11},e-\alpha)}(\overline{y}) \quad \text{and}$$

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

## 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,

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

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

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

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

where $K : R \times R^L \to R$ is a kernel function and $\sigma_N$ 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 $k_2 : R^L \to R$,

$$\hat{F}_{Y|X=x}(y) = \frac{\int_{-\infty}^{y} \hat{f}_N(s,x)\, ds}{\int_{-\infty}^{\infty} \hat{f}_N(s,x)\, ds} = \frac{\sum_{i=1}^N \widetilde{k}_1\left(\frac{y-Y^i}{\sigma}\right) k_2\left(\frac{x-X^i}{\sigma}\right)}{\sum_{i=1}^N k_2\left(\frac{x-X^i}{\sigma}\right)}$$

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 $\hat{F}_{Y|X=x}(y) = t$. 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″), (3.4), (3.5), (3.4′), (3.5′), (3.4″), (3.5″), (3.6), (3.7), (3.6′), (3.7′), (3.6″), and (3.7″). Hence, when (I.1) is satisfied

$$\widehat{F}_{\varepsilon|X_0=x_0}(e) = \widehat{F}_{Y|X=(x_0,\overline{x}_1)}(e) \quad \text{and}$$

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

when (I.2) is satisfied,

$$\widehat{F}_{\varepsilon}(e) = \widehat{F}_{Y|X_1=\overline{x}_1}(e) \quad \text{and}$$

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

and when (I.3) is satisfied

$$\widehat{F}_{\varepsilon}(e) = \widehat{F}_{Y|X_1=\overline{x}_1}(e) \quad \text{and}$$

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

When (II.1) is satisfied,

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

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

when (II.2) is satisfied

$$\widehat{F}_{\varepsilon}(e) = \widehat{F}_{Y|X_1=((e/\overline{\varepsilon})\overline{x}_1)}\left((e/\overline{\varepsilon})\alpha\right) \text{ and}$$

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

and when (II.3) is satisfied

$$\widehat{F}_{\varepsilon}(e) = \widehat{F}_{Y|X=(\overline{x}_0,(e/\overline{\varepsilon})\overline{x}_1)}\left((e/\overline{\varepsilon})\alpha\right), \text{ and}$$

$$\widehat{m}(x,e) = \widehat{F}^{-1}_{Y|X=x}\left(\widehat{F}_{Y|X=(\overline{x}_0,(e/\overline{\varepsilon})\overline{x})}\left((e/\overline{\varepsilon})\alpha\right)\right).$$

Finally, when (III.1) is satisfied

$$\widehat{F}_{\varepsilon|X_0=x_0}(e) = \widehat{F}_{Y|X=(x_0,\overline{x}_{11},e-\alpha)}(\overline{y}), \text{ and}$$

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

when (III.2) is satisfied

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

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

when (III.3) is satisfied

$$\widehat{F}_\varepsilon(e) = \widehat{F}_{Y|X=(\overline{x}_0,\overline{x}_{11},e-\alpha)}\left(\overline{y}\right) \quad \text{and}$$

$$\widehat{s}(x_0, x_{11}, e - x_{12}) = \widehat{F}^{-1}_{Y|X=x}\left(\widehat{F}_{Y|X_1=(\overline{x}_0,\overline{x}_{11},e-\alpha)}\left(\overline{y}\right)\right).$$

Note that when $\widehat{F}_{Y|X=x}$ is not strictly increasing, $\widehat{F}^{-1}_{Y|X=x}$ may contain more than one value. In that case we let the estimator be any of those values.

In all the above definitions, the value of the marginal or conditional distribution of $\varepsilon$ 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. Hence, the consistency and asymptotic normality of the estimator of the marginal or conditional distribution of $\varepsilon$ will follow from the consistency and asymptotic normality of the kernel estimator for the conditional distribution of $Y$ given that $W = w$. It follows that the asymptotic properties for each of the estimators for the distribution of $\varepsilon$ given above can be derived from the following result, by substituting the corresponding values of $w$ and $y$.

Let $d$ denote the dimension of $w$, and let $d' = d + 1$. Let $\int K(z)^2 = \int \left(\int K(s,z)\, ds\right)^2 dz$, where $s \in R$ and $z \in R^d$. 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 is continuously differentiable up to the order $s'$, for some $s' > 0$.

**Assumption C.3:** The kernel function $K(\cdot,\cdot)$ is Lipschitz, vanishes outside a compact set, integrates to 1, and is of order $s'$.

**Assumption C.4:** As $N \to \infty$, $\ln(N)/N\sigma_N^{d'} \to 0$ and $\sigma_N^{s'}\sqrt{N\sigma_N^{2d}} \to 0$.

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

Assumption C.2 requires that the pdf of $(Y, W)$ be sufficiently smooth. Note that this requires $\varepsilon$ 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, $\sigma_N$, 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* $s' \geq 2$*. Then,*

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

*and*

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

*where*

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

The proof is given in Appendix A.

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 $X$, or a 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 $X$, or a subvector, $\widetilde{W}$, of $X$ equals the value $\widetilde{w}$. The subvectors $W$ and $\widetilde{W}$, of $X$, are not required to have either the same dimension or common coordinates. 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, $\Phi$, 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 $\int K(z)^2 = \int\left(\int K(s, z)\, ds\right)^2 dz$, where $s \in R$ and $z \in R^d$. Our next theorem will make use of Assumptions C.1-C.3 and the following:

**Assumption C.4':** As $N \to \infty$, $\ln(N)/N\sigma_N^{d_j+1} \to 0$ and $\sigma_N^{s'}\sqrt{N\sigma_N^{2d_j}} \to 0$ $(j = 1, 2)$.

**Assumption C.5':** $0 < f(w), f(\widetilde{w}) < \infty$ and there exists $\delta, \xi > 0$ 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, for* $s' \geq 2$*. Then,*

$\widehat{n}(w, e)$ *converges in probability to* $n(w, e)$*,*

*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*

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

The proof is given in Appendix A.

## 3.4. Estimation when $m$ is additive separable

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

When the function $m$ is additively separable, we can develop estimators for $m$ with improved rates of convergence, using ideas similar to those presented in Linton and Nielsen (1995) for the estimation of additive regression functions. Suppose that

(3.9) $m(x_1,x_2,\varepsilon)=v_1(x_1,\varepsilon)+v_2(x_2)$,

for some unknown functions $v_1$ and $v_2$, and that the following is satisfied:

**Assumption S.1:** $E(v_1(X_1,\varepsilon)|X_2=x_2)=\int v_1(x_1,\varepsilon)\ f(x_1,\varepsilon|X_2=x_2)\ dx_1\, d\varepsilon=0.$

Assumption **S** is a location normalization, which is needed to guarantee the joint identification of the nonparametric functions $v_1$ and $v_2$. Making use of this assumption and (3.9), it follows that

$$v_2(x_2)=E(Y|X_2=x_2).$$

Hence, $v_2$ can be estimated by the kernel estimator for the conditional expectation of $Y$ given $X_2$ (Nadaraya (1969), Watson (1969)). The asymptotic properties of such an estimator are well known (Schuster (1972), Bierens (1987)). In particular, its asymptotic distribution does not depend on the dimensionality of $X_1$, only on that of $X_2$. Let

$$V=\int v_2(x_2)\,f(x_2)\,dx_2.$$

Since $Y=m(x_1,x_2,\varepsilon)$, (3.9) implies that

$$Y=\int Y\ f(x_1)\ dx_1=\int v_1(x_1,\varepsilon)\,f(x_2)\,dx_2+V=v_1(x_1,\varepsilon)+V$$

Thus, letting $X_1=(X_{10},X_{11})$, it follows from the arguments in the beginning of this section that if, for example, the following assumptions hold:

**Assumption S.2**: $\varepsilon$ is independent of $X_{11}$, conditional on $X_{10}$.

**Assumption S.3:** For all $X_1$, $v_1(X_1, \cdot)$ is strictly increasing in $\varepsilon$.

then, for all $(x_1, e)$,

$$(3.10)\quad F_{\varepsilon|X_{10}=x_{10}}(e) = F_{Y|X_1=x_1}\left(v_1(x_1, e) + V\right)$$

Hence, when the function $v_1$ satisfies any of the specifications given above for the function $m$, we can obtain estimators for the conditional distribution of $\varepsilon$ given $X_{01}$ and for the function $v_1$. These estimators are functionals of the kernel estimator for the conditional distribution of $Y$ given $X_1$ and of an estimator for $V$. The rates of convergence of such estimators will be independent of the dimensionality of $X_2$. Suppose, for example, that we specify that $v_1$ does not depend on $X_{10}$, and for some $\overline{x}_{11}$ and all $\varepsilon$

$$v_1(\overline{x}_{11}, \varepsilon) = \varepsilon.$$

Let

$$\widehat{V} = \int \widehat{v}_2(x_2)\, \widehat{f}(x_2)\, dx_2.$$

Then, the estimators for the conditional distribution of $\varepsilon$ given $X_{10}$ and for the function $v_1$ are, respectively

$$\widehat{F}_{\varepsilon|X_{10}=x_{10}}(e) = \widehat{F}_{Y|X_1=(x_{10},\overline{x}_{11})}\left(e + \widehat{V}\right)$$

and

$$\widehat{v}_1(x_1, e) = \widehat{F}^{-1}_{Y|X_1=(x_{10},x_{11})}\left(\widehat{F}_{Y|X_1=(x_{10},\overline{x}_{11})}\left(e + \widehat{V}\right)\right) - \widehat{V}$$

Since, as it is easy to show, $\sqrt{N}(\widehat{V} - V)$ possess a limiting distribution that is Normal, slight modifications in the proofs of Theorems 1 and 2 yield the result that, when Assumptions C.1-C.5 are satisfied for $s' \geq 2$, and $d$ denotes the dimension of $X_1$,

$$\sup_e \left|\widehat{F}_{\varepsilon|X_{10}=x_{10}}(e) - F_{\varepsilon|X_{10}=x_{10}}(e)\right| \to 0 \quad \text{in probability}$$

and

$$\sqrt{N}\sigma^{(d/2)}\left(\widehat{F}_{\varepsilon|X_{10}=x_{10}}(e) - F_{\varepsilon|X_{10}=x_{10}}(e)\right) \to N\left(0, V_F\right),$$

where

$$V_F = \{\int K(z)^2\}\left[F_{\varepsilon|X_{10}=x_{10}}(e)(1 - F_{\varepsilon|X_{10}=x_{10}}(e))\right]\left[1/f(x_{10}, \overline{x}_{11})\right];$$

and when assumptions C.1-C.3, C.4' and C.5' are satisfied, for $s' \geq 2$

$\widehat{v}_1(x_1, e) \to v_1(x_1, e)$ in probability,

and

$$\sqrt{N}\sigma_N^{d/2}\left(\widehat{v}_1(x_1, e) \to v_1(x_1, e)\right) \to N\left(0, V_n\right) \text{ in distribution,}$$

where

$$V_n = \{\int K(z)^2\}\left[\frac{F_{\varepsilon|X_{10}=x_{10}}(e)(1-F_{\varepsilon|X_{10}=x_{10}}(e))}{f_{Y|X_1=(x_{10},x_{11})}(v_1(x_{11},e))^2}\right]\left[\frac{1}{f(x_{10},x_{11})} + \frac{1}{f(x_{10},\overline{x}_{11})}\right]$$

and

$$\int K(z)^2 = \int K(z)^2 \, dz, \text{ with } z \in R^d.$$

# 4. Multivariate Unobservable Random Term

Imposing some structure on the function $m$, we can use the basic model described in the previous section 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 $\varepsilon$ 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 $\varepsilon_k$'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 $\varepsilon$ and each of the $K$ random functions.

In particular, our results will allow the identification of each individual function in a summation, when only the value of the sum of the random functions is observed. They will 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, $\varepsilon$.

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

$$(4.1) \quad m(X, \varepsilon) = r(n_1(w_{0_1}, w_1, \varepsilon_1), ..., n_K(w_{0_K}, w_K, \varepsilon_K))$$

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 $\varepsilon$, 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 $(w_{0_k}, \varepsilon_k)$, $n_k(w_{0_k}, \overline{w}_k, \varepsilon_k) = \varepsilon_k$.

(A.iii) 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 $\varepsilon_k$,

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

(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.iii) impose on each function $n_k$ the specification (I.1). Assumption (A.iv) 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 $\varepsilon_k$ are independent across them, while Assumption (A.v') states that the $\varepsilon_k$ are unconditionally independent across them. These assumptions allow us to identify, respectively, the conditional and unconditional joint distribution of $\varepsilon$, 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 $\varepsilon_k$. By Assumption (A.vi), $\varepsilon$ is independent of $X_1$, conditional on $X_0$, while by Assumption (A.vi'), $\varepsilon$ 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, whose proof is given in Appendix A:

**Theorem 3** *:(3.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 $(w_0, w_k, e_k)$,*

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

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

*(3.II) If Assumptions (A.i)-(A.iv), (A.v') and (A.vi') are satisfied, then $F_{\varepsilon|X_0=w_0}$ and $m$ are identified. In particular, for all $k$ and all $(w_0, w_k, e_k)$,*

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

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

Since, in the statement of the above theorem, the random functions, $n_k$, and the marginal distributions of the $\varepsilon_k$'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 $\varepsilon_k$'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 $(w_0, \overline{w}^k)$,

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

where

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

and

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

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

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

where

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

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

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

## 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 the profit function of a typical firm because we are interested in the supply and demand functions of that firm. These can be derived by differentiating the profit function. In this section, we present estimators for the derivatives of the function $m$, for some of the specifications presented in Section 2, and show their consistency and asymptotic normality. So, for example, when $m(x, \varepsilon)$ represents the profit function of a particular firm, $x$ is the price of the observable prices and $\varepsilon$ is the unobservable price of some input, we can use the derivatives of $m$ with respect to $x$ to determine the supply of the output and the demand of the inputs, for which their prices are observed. We can use the derivative of $m$ with respect to $\varepsilon$ to determine the demand for the input whose price is unobserved. For simplicity, we will only consider the case where $\varepsilon$ is independent of $X$, and where the dimensionality of the subvectors $W$ and $\widetilde{W}$, conditional on which we have to calculate the conditional distribution of $Y$, is equal to the dimensionality of $X$. We next, then, provide estimators for the derivatives of the function $m$, when $\varepsilon$ is independent of $X$ and $m$ satisfies each of the following specifications:

(5.I) $m(\overline{x}, \varepsilon) = \varepsilon$

(5.II) $m(\overline{x}, \overline{\varepsilon}) = \alpha$ and $m(\lambda\overline{x}, \lambda\overline{\varepsilon}) = \lambda\alpha$ for all $\lambda$

(5.III) $m(x_1, x_2, \varepsilon) = s(x_1, \varepsilon - x_2)$ and $s(\overline{x}_1, \alpha) = \overline{y}$

Let $\widetilde{x}$ and $\widetilde{e}$ be such that , if $m$ satisfies specification (5.I),

$$\widetilde{x} = \overline{x} \text{ and } \widetilde{e} = e,$$

if $m$ satisfies specification (5.II),

$$\widetilde{x} = (e/\overline{\varepsilon})\overline{x} \text{ and } \widetilde{e} = (e/\overline{\varepsilon})\alpha,$$

and if $m$ satisfies specification (5.III),

$$\widetilde{x} = (\overline{x}_1, e - \alpha) \text{ and } \widetilde{e} = \overline{y}.$$

Then, by the definition of $m$ in each of these cases, it follows that

(5.1) $\int^{m(x,e)} f(s,x)\,ds = f(x)\,\frac{\int^{\widetilde{e}} f(s,\widetilde{x})\,ds}{f(\widetilde{x})}.$

To obtain an estimator for the derivative of $m$ with respect to $x$, we differentiate both sides of (5.1) with respect to $x$. This gives

$$\int^{m(x,e)} \frac{\partial f(s,x)}{\partial x}\,ds + f(m(x,e),x)\,\frac{\partial m(x,e)}{\partial x} = \frac{\partial f(x)}{\partial x}\,F_{Y|X=\widetilde{x}}(\widetilde{e}),$$

which implies that

$$\frac{\partial m(x,e)}{\partial x} = \frac{1}{f(m(x,e),x)}\frac{\partial f(x)}{\partial x}\,F_{Y|X=\widetilde{x}}(\widetilde{e}) - \frac{1}{f(m(x,e),x)}\int^{m(x,e)} \frac{\partial f(s,x)}{\partial x}\,ds.$$

The estimator for the derivative of $m$ with respect to $x$ is then defined by:

(5.2) $\widehat{\frac{\partial m(x,e)}{\partial x}} = \frac{1}{\widehat{f}(\widehat{m}(x,e),x)}\frac{\partial\widehat{f}(x)}{\partial x}\,\widehat{F}_{Y|X=\widetilde{x}}(\widetilde{e}) - \frac{1}{\widehat{f}(\widehat{m}(x,e),x)}\int^{\widehat{m}(x,e)} \frac{\partial\widehat{f}(s,x)}{\partial x}\,ds,$

where $\widehat{m}(x,e)$ is the estimator of $m(x,e)$ defined in Section 3, $\widehat{f}(\widehat{m}(x,e),x)$ is the kernel estimator for the joint pdf of $Y$ and $X$, evaluated at the vector $(\widehat{m}(x,e),x)$, $\widehat{F}_{Y|X=\widetilde{x}}(\widetilde{e})$ is the kernel estimator for the conditional cdf of $Y$ given $X=\widetilde{x}$ evaluated at $Y=\widetilde{e}$, and $\partial\widehat{f}(x)/\partial x$ is the derivative with respect to $x$ of the kernel estimator for the pdf of $X$ evaluated at $X=x$.

The following result establishes the consistency and asymptotic normality of $\partial m(\widehat{x,e})/\partial x$.

**Theorem 4** *: Suppose that Assumptions C.1-C.3, C.4' and C.5' are satisfied for* $\widetilde{w}=\widetilde{x}$, $w=x$, *and* $s'\geq 3$. *Then,*

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

*and*

*(ii)* $\sqrt{N}\sigma_N^{(L/2)+1}\left(\widehat{\frac{\partial m(x,e)}{\partial x}} - \frac{\partial m(x,e)}{\partial x}\right) \to N\left(0, V_{\partial x}\right)$ *in distribution,*

*where*

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

The asymptotic variance of $\partial m(\widehat{x,e})/\partial x$ depends on the derivative of the kernel and on the variance due to the variance of $\widehat{F}_{Y|X=x}(m(x,e))$. The asymptotic variance of $\widehat{F}_{Y|X=\widetilde{x}}(\widetilde{e})$ does not affect the asymptotic variance of $\partial m(\widehat{x,e})/\partial x$, because $\widehat{F}_{Y|X=\widetilde{x}}(\widetilde{e})$ does not depend on the value $x$. Since $f_{Y|X=x}(m(x,e)) = f_\varepsilon(e)/(\partial m(x,e)/\partial\varepsilon)$, the variance increases the smaller is the density of $\varepsilon$ at $e$ and the larger is the derivative of $m$ with respect to $\varepsilon$. Note that the rate of convergence of $\partial m(\widehat{x,e})/\partial x$ is slower than that of $\widehat{m}(x,e)$. This is because, in contrast to $\widehat{m}(x,e)$, $\partial m(\widehat{x,e})/\partial x$ depends on derivatives of the pdf of $X$ and of the joint pdf of $(Y,X)$, which converge at a rate $\sigma_N-$times slower than the estimators for those pdf's.

It is possible to derive also estimators for the derivatives of the function $m$ with respect to $\varepsilon$. In this case, however, the estimators have different forms depending on the specification of the function $m$. Suppose that $m$ satisfies specification (5.I). Then, the definition of $m$ implies that

$$\frac{\int^{m(x,e)} f(s,x)\, ds}{f(x)} = \frac{\int^{e} f(s,\overline{x})\, ds}{f(\overline{x})}.$$

Hence,

$$\frac{\partial m(x,e)}{\partial \varepsilon} = \frac{f_{Y|X=\overline{x}}(e)}{f_{Y|X=x}(m(x,e))}.$$

The estimator for the derivative of $m$ with respect to $\varepsilon$, in this case, is then defined by

$$(5.3)\quad \widehat{\frac{\partial m(x,e)}{\partial \varepsilon}} = \frac{\widehat{f}_{Y|X=\overline{x}}(e)}{\widehat{f}_{Y|X=x}(\widehat{m}(x,e))}.$$

Suppose now that $m$ satisfies specification (5.II). Then, by the definition of $m$ it follows that

$$(5.4)\quad \frac{\int^{m(x,e)} f(s,x)\, ds}{f(x)} = \frac{\int^{(e/\overline{\varepsilon})\alpha} f(s,(e/\overline{\varepsilon})\overline{x})\, ds}{f((e/\overline{\varepsilon})\overline{x})}.$$

Differentiating both sides of (5.4) with respect to $\varepsilon$ gives

$$f_{Y|X=x}(m(x,e))\frac{\partial m(x,e)}{\partial \varepsilon} = (\alpha/\overline{\varepsilon})f_{Y|X=(e/\overline{\varepsilon})\overline{x})}((e/\overline{\varepsilon})\alpha)$$
$$+\frac{\int^{(e/\overline{\varepsilon})\alpha}\left(\frac{\partial f(s,(e/\overline{\varepsilon})\overline{x})}{\partial x}\right)'(\overline{x}/\overline{\varepsilon})\, ds}{f((e/\overline{\varepsilon})\overline{x})} - \frac{\left(\frac{\partial f((e/\overline{\varepsilon})\overline{x})}{\partial x}\right)'(\overline{x}/\overline{\varepsilon})F_{Y|X=(e/\overline{\varepsilon})\overline{x})}((e/\overline{\varepsilon})\alpha)}{(f((e/\overline{\varepsilon})\overline{x}))}$$

which implies that

$$\frac{\partial m(x,e)}{\partial \varepsilon} = \frac{(\alpha/\overline{\varepsilon})f_{Y|X=(e/\overline{\varepsilon})\overline{x})}((e/\overline{\varepsilon})\alpha)}{f_{Y|X=x}(m(x,e))}$$
$$+\frac{\int^{(e/\overline{\varepsilon})\alpha}\left(\frac{\partial f(s,(e/\overline{\varepsilon})\overline{x})}{\partial x}\right)'(\overline{x}/\overline{\varepsilon})\, ds}{f_{Y|X=x}(m(x,e))\; f((e/\overline{\varepsilon})\overline{x})} - \frac{\left(\frac{\partial f((e/\overline{\varepsilon})\overline{x})}{\partial x}\right)'(\overline{x}/\overline{\varepsilon})F_{Y|X=(e/\overline{\varepsilon})\overline{x})}((e/\overline{\varepsilon})\alpha)}{f_{Y|X=x}(m(x,e))\;(f((e/\overline{\varepsilon})\overline{x}))}.$$

Hence, in this case, the estimator for $\partial m(x,e)/\partial \varepsilon$ is defined by

$$(5.5)\quad \widehat{\frac{\partial m(x,e)}{\partial \varepsilon}} = \frac{(\alpha/\overline{\varepsilon})\widehat{f}_{Y|X=(e/\overline{\varepsilon})\overline{x}}((e/\overline{\varepsilon})\alpha)}{\widehat{f}_{Y|X=x}(\widehat{m}(x,e))}$$
$$+\frac{\int^{(e/\overline{\varepsilon})\alpha}\left(\frac{\partial \widehat{f}(s,(e/\overline{\varepsilon})\overline{x})}{\partial x}\right)'(\overline{x}/\overline{\varepsilon})\, ds}{\widehat{f}_{Y|X=x}(\widehat{m}(x,e))\;\widehat{f}((e/\overline{\varepsilon})\overline{x})} - \frac{\left(\frac{\partial \widehat{f}((e/\overline{\varepsilon})\overline{x}}{\partial x}\right)'(\overline{x}/\overline{\varepsilon})}{\widehat{f}_{Y|X=x}(\widehat{m}(x,e))\;\widehat{f}((e/\overline{\varepsilon})\overline{x})}\widehat{F}_{Y|X=(e/\overline{\varepsilon})\overline{x}}\left((e/\overline{\varepsilon})\alpha\right).$$

Finally, if $m$ satisfies specification (5.III), it follows by the structure of $m$ that

$$\frac{\partial m(x_1,x_2,\varepsilon)}{\partial \varepsilon} = -\frac{\partial m(x_1,x_2,\varepsilon)}{\partial x_2}.$$

Hence,

$$(5.6)\quad \widehat{\frac{\partial m(x_1,x_2,\varepsilon)}{\partial \varepsilon}} = -\widehat{\frac{\partial m(x_1,x_2,\varepsilon)}{\partial x_2}}$$

The next theorem establishes the asymptotic properties of the estimators for $\partial m(x,e)/\partial\varepsilon$ defined in (5.3), (5.5) and (5.6).

**Theorem 5** *: Suppose that Assumptions C.1-C.3, C.4' and C.5' are satisfied with $\widetilde{w}=\widetilde{x}$, $w=x$, and $s'\geq 3$. Then,*

*(i)* $\frac{\partial \widehat{m(x,e)}}{\partial\varepsilon}$ *converges in probability to* $\frac{\partial m(x,e)}{\partial\varepsilon}$.

*If $m$ satisfies specification (5.I) and $m(x,e)\neq e$,*

$$\sqrt{N}\sigma_N^{(L+1)/2}\left(\frac{\partial \widehat{m(x,e)}}{\partial e}-\frac{\partial m(x,e)}{\partial e}\right)\rightarrow N(0,V_{I,\partial\varepsilon}) \text{ in distribution,}$$

*where*

$$V_{I,\partial\varepsilon}=\left(\int K(s,z)^2 ds dz\right)\left\{\frac{f_{Y|X=\overline{x}}(e)}{f_{Y|X=x}(m(x,e))^2\ f(\overline{x})}+\frac{f_{Y|X=\overline{x}}(e)^2}{f_{Y|X=x}(m(x,e))^3 f(x)}\right\}$$

*If $m$ satisfies specification (5.II),*

$$\sqrt{N}\sigma_N^{(L/2)+1}\left(\frac{\partial \widehat{m(x,e)}}{\partial\varepsilon}-\frac{\partial m(x,e)}{\partial\varepsilon}\right)\rightarrow N\left(0,V_{II,\partial\varepsilon}\right) \text{ in distribution,}$$

*where*

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

*And if $m$ satisfies specification (5.III),*

$$\sqrt{N}\sigma_N^{(L/2)+1}\left(\frac{\partial \widehat{m(x,e)}}{\partial\varepsilon}-\frac{\partial m(x,e)}{\partial\varepsilon}\right)\rightarrow N\left(0,V_{III,\partial\varepsilon}\right) \text{ in distribution,}$$

*where*

$$V_{III,\partial\varepsilon}=\frac{F_\varepsilon(e)\ (1-F_\varepsilon(e))}{f_{Y|X=x}(m(x,e))^2 f(x)}\left[\int\left(\int\frac{\partial K(s,z_1,z_2)}{\partial z_2}ds\right)^2 dz_1 dz_2\right]$$

When $m$ satisfies specification (5.II), the rate of convergence of $\partial \widehat{m(x,e)}/\partial\varepsilon$ is the same as that of $\partial \widehat{m(x,e)}/\partial x$ because the slowest converging terms of both estimators depend on derivatives with respect to $x$ of the pdf of $X$ and the joint pdf of $(Y,X)$. The asymptotic variance of $\partial \widehat{m(x,e)}/\partial\varepsilon$ is due to the variance of the estimator of $F_{Y|X=\widetilde{x}}(\widetilde{e})$. Although the value of $\varepsilon$ at which the function $m$ is evaluated affects both, $F_{Y|X=x}(m(x,e))$ and $F_{Y|X=\widetilde{x}}(\widetilde{e})$, the estimator of the derivative of $F_{Y|X=\widetilde{x}}(\widetilde{e})$ with respect to $\varepsilon$ converges at a slower rate than the estimator of the derivative of $F_{Y|X=x}(m(x,e))$ with respect to $\varepsilon$.

When $m$ satisfies specification (5.I), $F_{Y|X=\widetilde{x}}(\widetilde{e})$ does not depend on the value $\varepsilon$ at which the function $m$ is evaluated. Hence, the asymptotic variance of $\partial \widehat{m(x,e)}/\partial\varepsilon$ depends only on the asymptotic variance of the estimator of the derivative of $F_{Y|X=x}(m(x,e))$ with respect to $\varepsilon$. The rate of convergence of $\partial \widehat{m(x,e)}/\partial\varepsilon$ is higher when $m$ satisfies specification (5.I) than when it satisfies specification (5.II) because the derivative of $F_{Y|X=x}(m(x,e))$ with respect to $\varepsilon$ depends only on the values, and not the derivatives, of the pdf's of $X$ and of $(Y,X)$.

When $m$ satisfies either specification (5.I) or specification (5.II), the asymptotic covariance between $\partial \widehat{m(x,e)}/\partial x$ and $\partial \widehat{m(x,e)}/\partial\varepsilon$ is zero. When $m$ satisfies specification (5.II), this is because the asymptotic distributions of these estimators are driven by kernel estimators of the pdf of $(Y,X)$ evaluated at different points. When $m$ satisfies specification (5.I), the zero asymptotic covariance is a consequence of the different rates of convergence of the estimators.

# 6. Simulations

To provide an indication of how the new estimators perform in practice, we present below the results of simulations performed 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 $\varepsilon$ and its parametric form is known. Also, since the function satisfies specifications (5.I) and (5.II), in Section 5, it allows one to evaluate the effect of the two normalizations. This design was estimated using the first normalization with $\overline{x} = \overline{\varepsilon} = 1$, $\alpha = 2$, and using the second normalization with $\overline{x} = 0$. 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 write this function in terms of $-\varepsilon$ to transform it so that it is strictly increasing in $\varepsilon$. Alternatively, we could have calculated the estimators under the restriction that $m$ is strictly decreasing in $\varepsilon$. This would have only modified the estimator for $\widehat{F}_\varepsilon(e)$. Instead of deriving $\widehat{F}_\varepsilon(e)$ from the value of $\widehat{F}_{Y|X=x}(y)$ at a particular $y$ and $x$, we would have derived $\widehat{F}_\varepsilon(e)$ from $1-\widehat{F}_{Y|X=x}(y)$ at the same particular $y$ and $x$. The expression for $\widehat{m}$ would have been the same as for the strictly increasing case. We used this design with $\overline{x} = \overline{\varepsilon} = 6$ and $\alpha = 6\cdot 3^3/4^4$. The normal distributions, which were chosen for $X$ and $\varepsilon$ in these designs, violate the assumption that the support of the observable variables is compact, but, since we are dealing with a finite set of data, we could have obtained the same results if we specified the distributions of $X$ and $\varepsilon$ so that they are equal to the chosen distributions only on a large enough compact set.

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

| | $\sigma_Y$ | $\sigma_X$ |
|---|---|---|
| Design I | 0.4031 | 0.2928 |
| Design II | 0.0596 | 0.2619 |

The results obtained for Design I, Normalization I(where $m$ satisfies specification (II)) and for Design I, Normalization II (where $m$ satisfies specification (I)) are presented at the end of the paper, together with the results obtained when the function $m$ is estimated by Least Squares. The latter estimator is denoted by $\widetilde{m}_{LS}$. The estimators are evaluated at points where $x$ equals the 25th, 50th, and 75th quantile of the distribution of $X$, and where $\varepsilon$ equals the 10th, 50th, and 90th quantile of its distribution. We do not look at the same quantiles for $X$ and $\varepsilon$ to avoid considering too many points where the value of the function $m$ is known. For each point, the tables show the bias, variance, mean square error, and asymptotic variance of the estimator.

Comparing the results obtained from both normalizations, we can see that, at the evaluated points, the MSE's of the estimators obtained using Normalization I are, in general, larger than those obtained using Normalization II. The MSE's of $\widehat{m}$ and $\widehat{F}_{\varepsilon}$ obtained using normalization I can be up to 3 and 4 times larger, respectively, than when using Normalization II. The MSE's of $\widehat{\partial m/\partial x}$ can be up to 1.5 times larger, and that of $\widehat{\partial m/\partial \varepsilon}$ can be up to 18 times larger than when Normalization II is used. Comparing the results obtained from both normalization with those of the Least Squares (LS) estimators, we can see that the MSE of $\widehat{m}$ when using Normalization I can be up to 25 times larger than when using LS, while the MSE of $\widehat{m}$ when using Normalization II can be up to 7 times larger than that of the LS estimator. The MSE of $\widehat{\partial m/\partial x}$ can be up to 52 times (for Normalization I) and 45 times (for Normalization II) than the MSE of the LS estimator. The MSE of the LS estimator for $\widehat{\partial m/\partial \varepsilon}$ is zero. So, the nonparametric estimator for $\widehat{\partial m/\partial \varepsilon}$ obviously performs much worse than the LS estimator.

The superiority of the LS estimator gets reversed when the function $m$ is nonlinear and nonadditive in $\varepsilon$. Following the previous results, we present the results for the nonparametric estimator using Normalization I and the data generated by Design II, together with the corresponding results for the LS estimators.

From these results, we can see that the MSE of the LS estimator of $m$ can be 1656 times larger than that of the nonparametric estimator. The difference is lessened when we compare the estimators for the derivatives. The MSE of the LS estimator of $\partial m/\partial x$ is 101 times larger than that of the nonparametric estimator at the point in the southwest corner, but it is 16 times smaller than the nonparametric estimator at the southeast corner. The relative performance of the nonparametric estimator of $\partial m/\partial \varepsilon$ is better than the relative performance of the nonparametric estimator of $\partial m/\partial x$. The MSE of the LS estimator of $\partial m/\partial \varepsilon$ is between 139 and 255 times larger than that of the nonparametric estimator at the points in the north, while its MSE is between 15 and 26 times larger than the MSE of the nonparametric estimator at the points in the south. We should note that the nonparametric estimators for the derivatives of $m$ are very jagged, which suggest that the bandwidths used are too small. When the bandwidths are increased to twice their previous sizes, the MSE's of $\widehat{\partial m/\partial x}$ and $\widehat{\partial m/\partial \varepsilon}$ are, in general, smaller. We present these MSE's following the previous results.

Design I / Normalization I

| $x$ | $e$ | $m$ | $Bias(\widehat{m})$ | $Var(\widehat{m})$ | $MSE(\widehat{m})$ | $Avar(\widehat{m})$ |
|---|---|---|---|---|---|---|
| -0.6745 | -1.2816 | -1.956041 | -0.055164 | 0.027445 | 0.030488 | 0.049802 |
| -0.6745 | 0.0000 | -0.674490 | 0.045053 | 0.013064 | 0.015093 | 0.017112 |
| -0.6745 | 1.2816 | 0.607062 | 0.158777 | 0.040833 | 0.066043 | 0.049802 |
| 0.0000 | -1.2816 | -1.281552 | -0.103554 | 0.035620 | 0.046344 | 0.046197 |
| 0.0000 | 0.0000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.015174 |
| 0.0000 | 1.2816 | 1.281552 | 0.115072 | 0.039464 | 0.052706 | 0.046197 |
| 0.6745 | -1.2816 | -0.607062 | -0.166698 | 0.037238 | 0.065027 | 0.049802 |
| 0.6745 | 0.0000 | 0.674490 | -0.054845 | 0.011865 | 0.014873 | 0.017112 |
| 0.6745 | 1.2816 | 1.956041 | 0.049250 | 0.029154 | 0.031580 | 0.049802 |

| $e$ | $F$ | $Bias(\widehat{F})$ | $Var(\widehat{F})$ | $MSE(\widehat{F})$ | $AVar(\widehat{F})$ |
|---|---|---|---|---|---|
| -1.6449 | 0.05 | 0.005751 | 0.000593 | 0.000626 | 0.000887 |
| -0.6745 | 0.25 | 0.008865 | 0.000766 | 0.000845 | 0.001137 |
| 0.0000 | 0.50 | -0.001386 | 0.000909 | 0.000911 | 0.001207 |
| 0.6745 | 0.75 | -0.007094 | 0.000712 | 0.000762 | 0.001137 |
| 1.6449 | 0.95 | -0.006719 | 0.000546 | 0.000591 | 0.000887 |

| x | e | $\partial m/\partial x$ | $Bias(\widehat{\partial m/\partial x})$ | $Var(\widehat{\partial m/\partial x})$ | $MSE(\widehat{\partial m/\partial x})$ | $Avar(\widehat{\partial m/\partial x})$ |
|---|---|---|---|---|---|---|
| -0.6745 | -1.2816 | 1.000000 | -0.096347 | 0.093556 | 0.102839 | 0.103337 |
| -0.6745 | 0.0000 | 1.000000 | -0.071001 | 0.060002 | 0.065043 | 0.055549 |
| -0.6745 | 1.2816 | 1.000000 | -0.057210 | 0.088162 | 0.091435 | 0.103337 |
| 0.0000 | -1.2816 | 1.000000 | -0.065292 | 0.079831 | 0.084094 | 0.082313 |
| 0.0000 | 0.0000 | 1.000000 | -0.064566 | 0.051048 | 0.055217 | 0.044248 |
| 0.0000 | 1.2816 | 1.000000 | -0.094097 | 0.071034 | 0.079888 | 0.082313 |
| 0.6745 | -1.2816 | 1.000000 | -0.087892 | 0.089573 | 0.097298 | 0.103337 |
| 0.6745 | 0.0000 | 1.000000 | -0.079195 | 0.060462 | 0.066734 | 0.055549 |
| 0.6745 | 1.2816 | 1.000000 | -0.098362 | 0.100482 | 0.110157 | 0.103337 |

| x | e | $\partial m/\partial \varepsilon$ | $Bias(\widehat{\partial m/\partial \varepsilon})$ | $Var(\widehat{\partial m/\partial \varepsilon})$ | $MSE(\widehat{\partial m/\partial \varepsilon})$ | $Avar(\widehat{\partial m/\partial \varepsilon})$ |
|---|---|---|---|---|---|---|
| -0.6745 | -1.2816 | 1.000000 | 0.141215 | 0.339550 | 0.359492 | 0.187114 |
| -0.6745 | 0.0000 | 1.000000 | 0.074860 | 0.086538 | 0.092142 | 0.044248 |
| -0.6745 | 1.2816 | 1.000000 | 0.112037 | 0.298026 | 0.310578 | 0.187114 |
| 0.0000 | -1.2816 | 1.000000 | 0.132217 | 0.325391 | 0.342872 | 0.187114 |
| 0.0000 | 0.0000 | 1.000000 | 0.064566 | 0.051048 | 0.055217 | 0.044248 |
| 0.0000 | 1.2816 | 1.000000 | 0.115924 | 0.304955 | 0.318394 | 0.187114 |
| 0.6745 | -1.2816 | 1.000000 | 0.134001 | 0.327781 | 0.345737 | 0.187114 |
| 0.6745 | 0.0000 | 1.000000 | 0.072430 | 0.085134 | 0.090380 | 0.044248 |
| 0.6745 | 1.2816 | 1.000000 | 0.116829 | 0.316786 | 0.330435 | 0.187114 |

| x | e | $m$ | $Bias(\widehat{m})$ | $Var(\widehat{m})$ | $MSE(\widehat{m})$ | $Avar(\widehat{m})$ |
|---|---|---|---|---|---|---|
| -0.6745 | -1.2816 | -1.956041 | 0.039292 | 0.018318 | 0.019862 | 0.031832 |
| -0.6745 | 0.0000 | -0.674490 | 0.045053 | 0.013064 | 0.015093 | 0.017112 |
| -0.6745 | 1.2816 | 0.607062 | 0.045520 | 0.018562 | 0.020634 | 0.031832 |
| 0.0000 | -1.2816 | -1.281552 | -0.000000 | 0.000000 | 0.000000 | 0.028227 |
| 0.0000 | 0.0000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.015174 |
| 0.0000 | 1.2816 | 1.281552 | 0.000000 | 0.000000 | 0.000000 | 0.028227 |
| 0.6745 | -1.2816 | -0.607062 | -0.062862 | 0.017918 | 0.021869 | 0.031832 |
| 0.6745 | 0.0000 | 0.674490 | -0.054845 | 0.011865 | 0.014873 | 0.017112 |
| 0.6745 | 1.2816 | 1.956041 | -0.053940 | 0.018461 | 0.021370 | 0.031832 |

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

| x | e | $\partial m/\partial x$ | $Bias(\widehat{\partial m/\partial x})$ | $Var(\widehat{\partial m/\partial x})$ | $MSE(\widehat{\partial m/\partial x})$ | $Avar(\widehat{\partial m/\partial x})$ |
|---|---|---|---|---|---|---|
| -0.6745 | -1.2816 | 1.000000 | -0.079104 | 0.087267 | 0.093525 | 0.103337 |
| -0.6745 | 0.0000 | 1.000000 | -0.071001 | 0.060002 | 0.065043 | 0.055549 |
| -0.6745 | 1.2816 | 1.000000 | -0.069602 | 0.081449 | 0.086293 | 0.103337 |
| 0.0000 | -1.2816 | 1.000000 | -0.053773 | 0.072362 | 0.075253 | 0.082313 |
| 0.0000 | 0.0000 | 1.000000 | -0.064566 | 0.051048 | 0.055217 | 0.044248 |
| 0.0000 | 1.2816 | 1.000000 | -0.081356 | 0.068593 | 0.075212 | 0.082313 |
| 0.6745 | -1.2816 | 1.000000 | -0.095052 | 0.079552 | 0.088587 | 0.103337 |
| 0.6745 | 0.0000 | 1.000000 | -0.079195 | 0.060462 | 0.066734 | 0.055549 |
| 0.6745 | 1.2816 | 1.000000 | -0.077366 | 0.089999 | 0.095985 | 0.103337 |

| x | e | $\partial m/\partial \varepsilon$ | $Bias(\widehat{\partial m/\partial \varepsilon})$ | $Var(\widehat{\partial m/\partial \varepsilon})$ | $MSE(\widehat{\partial m/\partial \varepsilon})$ | $Avar(\widehat{\partial m/\partial \varepsilon})$ |
|---|---|---|---|---|---|---|
| -0.6745 | -1.2816 | 1.000000 | 0.006771 | 0.023265 | 0.023311 | 0.059803 |
| -0.6745 | 0.0000 | 1.000000 | -0.001166 | 0.011048 | 0.011050 | 0.026308 |
| -0.6745 | 1.2816 | 1.000000 | 0.001031 | 0.021021 | 0.021022 | 0.059803 |
| 0.0000 | -1.2816 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.053030 |
| 0.0000 | 0.0000 | 1.000000 | -0.000000 | 0.000000 | 0.000000 | 0.023329 |
| 0.0000 | 1.2816 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.053030 |
| 0.6745 | -1.2816 | 1.000000 | 0.002536 | 0.019633 | 0.019640 | 0.059803 |
| 0.6745 | 0.0000 | 1.000000 | -0.002955 | 0.010270 | 0.010279 | 0.026308 |
| 0.6745 | 1.2816 | 1.000000 | 0.007917 | 0.023745 | 0.023807 | 0.059803 |

Design I / Least Squares

| x | e | $m$ | $Bias(\widetilde{m}_{LS})$ | $Var(\widetilde{m}_{LS})$ | $MSE(\widetilde{m}_{LS})$ |
|---|---|---|---|---|---|
| -0.6745 | -1.2816 | -1.956041 | -0.004112 | 0.002865 | 0.002882 |
| -0.6745 | 0.0000 | -0.674490 | -0.004112 | 0.002865 | 0.002882 |
| -0.6745 | 1.2816 | 0.607062 | -0.004112 | 0.002865 | 0.002882 |
| 0.0000 | -1.2816 | -1.281552 | -0.000076 | 0.002090 | 0.002090 |
| 0.0000 | 0.0000 | 0.000000 | -0.000076 | 0.002090 | 0.002090 |
| 0.0000 | 1.2816 | 1.281552 | -0.000076 | 0.002090 | 0.002090 |
| 0.6745 | -1.2816 | -0.607062 | 0.003961 | 0.003203 | 0.003219 |
| 0.6745 | 0.0000 | 0.674490 | 0.003961 | 0.003203 | 0.003219 |
| 0.6745 | 1.2816 | 1.956041 | 0.003961 | 0.003203 | 0.003219 |

| x | e | $\partial m/\partial x$ | $Bias(\partial\widetilde{m}_{LS}/\partial x)$ | $Var(\partial\widetilde{m}_{LS}/\partial x)$ | $MSE(\partial\widetilde{m}_{LS}/\partial x)$ |
|---|---|---|---|---|---|
| -0.6745 | -1.2816 | 1.000000 | 0.005984 | 0.002075 | 0.002111 |
| -0.6745 | 0.0000 | 1.000000 | 0.005984 | 0.002075 | 0.002111 |
| -0.6745 | 1.2816 | 1.000000 | 0.005984 | 0.002075 | 0.002111 |
| 0.0000 | -1.2816 | 1.000000 | 0.005984 | 0.002075 | 0.002111 |
| 0.0000 | 0.0000 | 1.000000 | 0.005984 | 0.002075 | 0.002111 |
| 0.0000 | 1.2816 | 1.000000 | 0.005984 | 0.002075 | 0.002111 |
| 0.6745 | -1.2816 | 1.000000 | 0.005984 | 0.002075 | 0.002111 |
| 0.6745 | 0.0000 | 1.000000 | 0.005984 | 0.002075 | 0.002111 |
| 0.6745 | 1.2816 | 1.000000 | 0.005984 | 0.002075 | 0.002111 |

| x | e | $\partial m/\partial \varepsilon$ | $Bias(\partial\widetilde{m}_{LS}/\partial \varepsilon)$ | $Var(\partial\widetilde{m}_{LS}/\partial \varepsilon)$ | $MSE(\partial\widetilde{m}_{LS}/\partial \varepsilon)$ |
|---|---|---|---|---|---|
| -0.6745 | -1.2816 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |
| -0.6745 | 0.0000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |
| -0.6745 | 1.2816 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.0000 | -1.2816 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.0000 | 0.0000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.0000 | 1.2816 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.6745 | -1.2816 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.6745 | 0.0000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |
| 0.6745 | 1.2816 | 1.000000 | 0.000000 | 0.000000 | 0.000000 |

Design II / Normalization I

| x | e | $m$ | $Bias(\widehat{m})$ | $Var(\widehat{m})$ | $MSE(\widehat{m})$ | $Avar(\widehat{m})$ |
|---|---|---|---|---|---|---|
| 5.3255 | -7.2816 | 0.219734 | 0.002511 | 0.000572 | 0.000579 | 0.000456 |
| 5.3255 | -6.0000 | 0.392749 | 0.015278 | 0.000788 | 0.001021 | 0.000738 |
| 5.3255 | -4.7184 | 0.807553 | -0.029261 | 0.008057 | 0.008914 | 0.014678 |
| 6.0000 | -7.2816 | 0.354044 | 0.006306 | 0.001094 | 0.001133 | 0.001099 |
| 6.0000 | -6.0000 | 0.632813 | 0.000000 | 0.000000 | 0.000000 | 0.001698 |
| 6.0000 | -4.7184 | 1.301161 | -0.094294 | 0.022210 | 0.031101 | 0.035347 |
| 6.6745 | -7.2816 | 0.542156 | 0.001258 | 0.001945 | 0.001947 | 0.002778 |
| 6.6745 | -6.0000 | 0.969041 | -0.025075 | 0.003811 | 0.004439 | 0.004491 |
| 6.6745 | -4.7184 | 1.992501 | -0.212119 | 0.046652 | 0.091647 | 0.089355 |

| $e$ | $F$ | $Bias(\widehat{F})$ | $Var(\widehat{F})$ | $MSE(\widehat{F})$ | $AVar(\widehat{F})$ |
|---|---|---|---|---|---|
| -7.6449 | 0.05 | 0.028041 | 0.000915 | 0.001701 | 0.000992 |
| -6.6745 | 0.25 | 0.030218 | 0.001118 | 0.002031 | 0.001271 |
| -6.0000 | 0.50 | -0.006567 | 0.001260 | 0.001303 | 0.001350 |
| -5.3255 | 0.75 | -0.037964 | 0.001188 | 0.002629 | 0.001271 |
| -4.3551 | 0.95 | -0.032881 | 0.001011 | 0.002092 | 0.000992 |

| x | e | $\partial m/\partial x$ | $Bias(\widehat{\partial m/\partial x})$ | $Var(\widehat{\partial m/\partial x})$ | $MSE(\widehat{\partial m/\partial x})$ | $Avar(\widehat{\partial m/\partial x})$ |
|---|---|---|---|---|---|---|
| 5.3255 | -7.2816 | 0.165042 | 0.006869 | 0.001542 | 0.001590 | 0.001183 |
| 5.3255 | -6.0000 | 0.294994 | -0.015741 | 0.004668 | 0.004915 | 0.002993 |
| 5.3255 | -4.7184 | 0.606554 | -0.070023 | 0.075588 | 0.080491 | 0.038067 |
| 6.0000 | -7.2816 | 0.236029 | 0.004099 | 0.003610 | 0.003627 | 0.002447 |
| 6.0000 | -6.0000 | 0.421875 | -0.015777 | 0.011761 | 0.012010 | 0.006190 |
| 6.0000 | -4.7184 | 0.867441 | -0.142476 | 0.120034 | 0.140333 | 0.078719 |
| 6.6745 | -7.2816 | 0.324912 | -0.014610 | 0.009300 | 0.009513 | 0.007205 |
| 6.6745 | -6.0000 | 0.580743 | -0.050031 | 0.037556 | 0.040059 | 0.018223 |
| 6.6745 | -4.7184 | 1.194099 | -0.192415 | 0.408242 | 0.445265 | 0.231742 |

| x | e | $\partial m/\partial \varepsilon$ | $Bias(\widehat{\partial m/\partial \varepsilon})$ | $Var(\widehat{\partial m/\partial \varepsilon})$ | $MSE(\widehat{\partial m/\partial \varepsilon})$ | $Avar(\widehat{\partial m/\partial \varepsilon})$ |
|---|---|---|---|---|---|---|
| 5.3255 | -7.2816 | 0.090530 | 0.025446 | 0.002902 | 0.003549 | 0.002143 |
| 5.3255 | -6.0000 | 0.196374 | -0.007714 | 0.002760 | 0.002819 | 0.002384 |
| 5.3255 | -4.7184 | 0.513444 | -0.069647 | 0.105880 | 0.110731 | 0.068928 |
| 6.0000 | -7.2816 | 0.145866 | 0.017416 | 0.006009 | 0.006312 | 0.005563 |
| 6.0000 | -6.0000 | 0.316406 | -0.015777 | 0.011761 | 0.012010 | 0.006190 |
| 6.0000 | -4.7184 | 0.827281 | -0.147779 | 0.215877 | 0.237716 | 0.178945 |
| 6.6745 | -7.2816 | 0.223368 | 0.001552 | 0.010970 | 0.010972 | 0.013045 |
| 6.6745 | -6.0000 | 0.484521 | -0.062615 | 0.022314 | 0.026234 | 0.014515 |
| 6.6745 | -4.7184 | 1.266836 | -0.252486 | 0.946042 | 1.009791 | 0.419617 |

Design II / Least Squares

| x | e | $m$ | $Bias(\widetilde{m}_{LS})$ | $Var(\widetilde{m}_{LS})$ | $MSE(\widetilde{m}_{LS})$ |
|---|---|---|---|---|---|
| 5.3255 | -7.2816 | 0.219734 | -0.978663 | 0.000804 | 0.958585 |
| 5.3255 | -6.0000 | 0.392749 | 0.129874 | 0.000804 | 0.017671 |
| 5.3255 | -4.7184 | 0.807553 | 0.996621 | 0.000804 | 0.994058 |
| 6.0000 | -7.2816 | 0.354044 | -0.734285 | 0.001560 | 0.540735 |
| 6.0000 | -6.0000 | 0.632813 | 0.268498 | 0.001560 | 0.073651 |
| 6.0000 | -4.7184 | 1.301161 | 0.881700 | 0.001560 | 0.778955 |
| 6.6745 | -7.2816 | 0.542156 | -0.543710 | 0.005129 | 0.300750 |
| 6.6745 | -6.0000 | 0.969041 | 0.310956 | 0.005129 | 0.101822 |
| 6.6745 | -4.7184 | 1.992501 | 0.569048 | 0.005129 | 0.328945 |

| x | e | $\partial m/\partial x$ | $Bias(\partial\widetilde{m}_{LS}/\partial x)$ | $Var(\partial\widetilde{m}_{LS}/\partial x)$ | $MSE(\partial\widetilde{m}_{LS}/\partial x)$ |
|---|---|---|---|---|---|
| 5.3255 | -7.2816 | 0.165042 | 0.396400 | 0.003091 | 0.160224 |
| 5.3255 | -6.0000 | 0.294994 | 0.266448 | 0.003091 | 0.074086 |
| 5.3255 | -4.7184 | 0.606554 | -0.045112 | 0.003091 | 0.005127 |
| 6.0000 | -7.2816 | 0.236029 | 0.325413 | 0.003091 | 0.108985 |
| 6.0000 | -6.0000 | 0.421875 | 0.139568 | 0.003091 | 0.022571 |
| 6.0000 | -4.7184 | 0.867441 | -0.305998 | 0.003091 | 0.096726 |
| 6.6745 | -7.2816 | 0.324912 | 0.236530 | 0.003091 | 0.059038 |
| 6.6745 | -6.0000 | 0.580743 | -0.019301 | 0.003091 | 0.003464 |
| 6.6745 | -4.7184 | 1.194099 | -0.632657 | 0.003091 | 0.403346 |

| x | e | $\partial m/\partial \varepsilon$ | $Bias(\partial\widetilde{m}_{LS}/\partial \varepsilon)$ | $Var(\partial\widetilde{m}_{LS}/\partial \varepsilon)$ | $MSE(\partial\widetilde{m}_{LS}/\partial \varepsilon)$ |
|---|---|---|---|---|---|
| 5.3255 | -7.2816 | 0.090530 | 0.909470 | 0.000000 | 0.827135 |
| 5.3255 | -6.0000 | 0.196374 | 0.803626 | 0.000000 | 0.645814 |
| 5.3255 | -4.7184 | 0.513444 | 0.486556 | 0.000000 | 0.236737 |
| 6.0000 | -7.2816 | 0.145866 | 0.854134 | 0.000000 | 0.729545 |
| 6.0000 | -6.0000 | 0.316406 | 0.683594 | 0.000000 | 0.467300 |
| 6.0000 | -4.7184 | 0.827281 | 0.172719 | 0.000000 | 0.029832 |
| 6.6745 | -7.2816 | 0.223368 | 0.776632 | 0.000000 | 0.603157 |
| 6.6745 | -6.0000 | 0.484521 | 0.515479 | 0.000000 | 0.265719 |
| 6.6745 | -4.7184 | 1.266836 | -0.266836 | 0.000000 | 0.071202 |

Design II / Normalization I / Large bandwidth

| x | e | $\partial m/\partial x$ | $MSE(\widehat{\partial m/\partial x})$ |
|---|---|---|---|
| 5.3255 | -7.2816 | 0.165042 | 0.000465 |
| 5.3255 | -6.0000 | 0.294994 | 0.003486 |
| 5.3255 | -4.7184 | 0.606554 | 0.062995 |
| 6.0000 | -7.2816 | 0.236029 | 0.000745 |
| 6.0000 | -6.0000 | 0.421875 | 0.011162 |
| 6.0000 | -4.7184 | 0.867441 | 0.158398 |
| 6.6745 | -7.2816 | 0.324912 | 0.002422 |
| 6.6745 | -6.0000 | 0.580743 | 0.032012 |
| 6.6745 | -4.7184 | 1.194099 | 0.343780 |

| x | e | $\partial m/\partial \varepsilon$ | $MSE(\widehat{\partial m/\partial \varepsilon})$ |
|---|---|---|---|
| 5.3255 | -7.2816 | 0.090530 | 0.001806 |
| 5.3255 | -6.0000 | 0.196374 | 0.001696 |
| 5.3255 | -4.7184 | 0.513444 | 0.085641 |
| 6.0000 | -7.2816 | 0.145866 | 0.000910 |
| 6.0000 | -6.0000 | 0.316406 | 0.011162 |
| 6.0000 | -4.7184 | 0.827281 | 0.269792 |
| 6.6745 | -7.2816 | 0.223368 | 0.001824 |
| 6.6745 | -6.0000 | 0.484521 | 0.039744 |
| 6.6745 | -4.7184 | 1.266836 | 0.711056 |

## 7. 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 an unobservable random vector. The estimators for the distribution of the unobservable random variable, the nonparametric function, and the derivatives of 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. To derive the asymptotic distributions of the estimators, we first linearized the functionals, by calculating their Hadamard-derivatives, and then applied a Delta method, as developed in Ait-Sahalia (1994) and Newey (1994).

The results of some simulations indicate that the method may outperform estimators that require specifying a parametric structure for the function to be estimated, when the specified structure is incorrect. Since one can rarely find a parametric specification that would perfectly fit the true function, there seems to be a benefit to using the new estimators.

# Appendix A

We present here the proofs of Theorems 1, 2, 5, and 6. To prove these theorems, we will use a Delta Method, like the ones developed in Ait-Sahalia (1994) and Newey (1994).

**Proof of Theorem 1:** Let $F$ denote the joint cdf of $(Y, X)$, $f(y,w)$ denote its probability density function (pdf), and $f(w)$ denote the marginal pdf of $w$. For any function $G : R^{1+L} \to R$, define $g(y,w) = \partial^{1+L}G(y,w)/\partial y\partial w$, $g(w) = \int g(y,w)\ dy$, $g(w) = \int g(y,w)\,dy$, $G_{y|W=w'}(y') = \left(\int_{-\infty}^{y'} g(y,w')ds\right)/g(w')$, and $\widetilde{G_Y}(y,w) = \int^{y} g(s,w)\ ds = \int 1[s \leq y]\ g(s,w)ds$ where $1[\cdot] = 1$ if $[\cdot]$ is true, and it equals zero otherwise. Let $\underline{\mathrm{C}}$ denote a compact set in $R^{1+L}$ that strictly includes $\Theta$, the compact support of $(Y \times X)$. Let D denote the set of all functions $G : \underline{R}^{1+L} \to R$ such that $g(y,w)$ exists and vanishes outside $\underline{\mathrm{C}}$. Let $D$ denote the set of all functions $\widetilde{G_Y}$ that are derived from some $G$ in D. Since there is a 1-1 relationship between functions in D and functions in $D$, we can define a functional on D or on $D$ without altering its definition. Define then 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 $\Lambda$ on $y$ and $w$, for brevity of exposition.

Let $\|G\|$ denote the sup norm of $g(y,w)$. Then, 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,

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

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

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

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

It follows that for some $c < \infty$,

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

Note the this implies that the functional $\Lambda$ is continuous.
Let $H = \widehat{F} - F$. From (1) and (2), $\left|\widehat{F}_{Y|W=w}(y) - F_{Y|W=w}(y)\right| \leq \frac{c}{f(w)}\left\|\widehat{F} - F\right\| + \frac{c}{f(w)^2}\left\|\widehat{F} - F\right\|^2$. By Assumptions (A.1)-(A.4) and Lemma B.3 in Newey (1994), $\left\|\widehat{F} - F\right\| \to 0$ in probability. Hence, $\sup_{y \in R}\left|\widehat{F}_{Y|W=w}(y) - F_{Y|W=w}(y)\right| \to 0$ in probability.

To prove the result about the asymptotic distribution, we note that by (1) and (2), $\Lambda$ is Hadamard differentiable at $F$. It then follows by Theorem 3.9.4 in van der Vaart and Wellner (1996) and the Lemma in Appendix B that

$$\sqrt{N\sigma_N^{2L}}\left(\Lambda(\widehat{\widetilde{F}}_Y) - \Lambda(\widetilde{F}_Y)\right) - D\Lambda\left(F, \sqrt{N\sigma_N^{2L}}(\widehat{\widetilde{F}}_Y - \widetilde{F}_Y)\right)$$

converges in outer probability to 0. Since

$$D\Lambda\left((F,\sqrt{N\sigma_N^{2L}}(\widehat{\widetilde{F}}_Y-\widetilde{F}_Y)\right)=\sqrt{N\sigma_N^{2L}}\left(\int\frac{(1[s\leq y]-F_{Y|W=w}(y))}{f(w)}\left(\widehat{f}(s,w)-f(s,w)\right)ds\right),$$

it follows by the Lemma in Appendix B that

$$D\Lambda\left((F,\sqrt{N\sigma_N^{2L}}(\widehat{\widetilde{F}}_Y-\widetilde{F}_Y)\right)\to N(0,V_F)$$

where

$$V_F=\left\{\int\left(\int K(s,z)ds\right)^2dz\right\}\left\{\left(\tfrac{1}{f(w)^2}\right)\int\left(1\left[s\leq y\right]-F_{Y|W=w}(y)\right)^2f(s,w)ds\right\}$$
$$=\left\{\int\left(\int K(s,z)ds\right)^2dz\right\}\left(\tfrac{1}{f(w)^2}\right)\left[F_{Y|W=w}(y)\left(1-F_{Y|W=w}(y)\right)\right]$$

Hence,

$$\sqrt{N\sigma_N^{2L}}\left(\widehat{F}_{Y|W=w}(y)-F_{Y|W=w}(y)\right)$$
$$=\sqrt{N\sigma_N^{2L}}\left(\Lambda(\widehat{\widetilde{F}}_Y)-\Lambda(\widetilde{F}_Y)\right)$$
$$\to N(0,V_F).$$

**Proof of Theorem 2:** Let $F(y,w)$ denote the distribution function (cdf) of the vector of observable variables $(Y,X)$, $f(y,w)$ denote its probability density function (pdf), and $f(x)$ denote the marginal pdf of $X$, and $F_{Y|X=x}$ denote the conditional cdf of $Y$ given $X=x$. For any subvector $W$ of $X$, let $f(w)$ denote the marginal pdf of $W$ when $W=w$, $f(y|w)$ denote the conditional pdf of $Y$, conditional on $W=w$, and $F_{Y|W=w}(y)$ denote the conditional cdf of $Y$, conditional on $W=w$. For any function $G:R^{1+L}\to R$, define $g(y,w)=\partial^L G(y,w)/\partial y\partial w$, $g(w)=\int g(y,w)$ $dy$, $g(w)=\int g(y,w)\,dy$, $G_{Y|W=w'}(y')=\left(\int_{-\infty}^{y'}g(y,w')ds\right)/g(w')$, and $\widetilde{G_Y}(y,w)=\int^y g(s,w)ds=\int 1[s\leq y]\;g(s,w)\;ds$ where $1[\cdot]=1$ if $[\cdot]$ is true, and it equals zero otherwise. Further, for any subvector $W$ of $X$, define $g(w)=g(y,w)=\int g(y,w,z)\;dz$, $g(w)=\int g(y,w,z)\;dy\,dz$, $G_{Y|W=w'}(y')$ $=\left(\int_{-\infty}^{y'}g(y,w')ds\right)/g(w')$, and $\widetilde{G_Y}(y,w)=\int^y g(s,w)ds=\int 1[s\leq y]\;g(s,w)\;ds$ where $1[\cdot]=1$ if $[\cdot]$ is true, and it equals zero otherwise, and where $z$ denotes the value of the coordinates of $X$ that are not included in $W$. Let $\underline{\mathrm{C}}$ denote a compact set in $R^L$ that strictly includes $\Theta$. Let D denote the set of all functions $G:R^L\to R$ such that $g(y,x)$ vanishes outside $\underline{\mathrm{C}}$. Let $D$ denote the set of all functions $\widetilde{G_Y}$ that are derived from some $G$ in D. Since there is a 1-1 relationship between functions in D and functions in $D$, we can define a functional on D or on $D$ without altering its definition. Let $W$ and $\widetilde{W}$ be two subvectors of $X$, not necessarily corresponding to the same coordinates 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, $\Phi(F)=\Phi(\widetilde{F}_Y)=n(w,e)$ and $\Phi(\widehat{F})$ $=\Phi(\widehat{\widetilde{F_Y}})=\widehat{n}(w,e)$.

Define the functionals $\eta$ and $\nu$ 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 $\|G\|$ denote the sup norm of $g(y,x)$. Then, if $H \in$D, there exists $\rho_1 > 0$ such that if $\|H\| \leq \rho_1$ then, for some $0 < a, b < \infty$, all $y$ and all $s \in N(m(w,e),\xi)$,

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

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

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

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

$$|r - r^*| \leq \left|\frac{1}{f_{Y|W=w}(F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))}\right|\frac{d\|H\|}{f(w)} + \frac{j_1 d^2\|H\|^2}{f(w)^2}.$$

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

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

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

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

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

$$\left|(F+H)_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})-F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right|\leq\frac{d\|H\|}{f(w)}$$

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,

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

$$=\frac{\partial(F+H)_{Y|W=w}}{\partial y}(r_2)\ \left[(F+H)^{-1}_{Y|W=w}(t)-\ F^{-1}_{Y|W=w}(t)\right]$$

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 $\eta$,

$$\eta(F+H)-\eta(F)=(F+H)_{Y|W=w}(\Phi(F+H))-F_{Y|W=w}(\Phi(F))$$

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

By the Mean Value Theorem, there exist $r_f$ and $r_h$ between $\Phi(F)$ and $\Phi(F+H)$ such that

$$\int^{\Phi(F+H)} f(s,w)\ ds - \int^{\Phi(F)} f(s,w)\ ds = f(r_f, w)\ (\Phi(F+H) - \Phi(F)) \text{ and}$$

$$\int^{\Phi(F+H)} h(s,w)\ ds - \int^{\Phi(F)} h(s,w)\ ds = h(r_h, w)\ (\Phi(F+H) - \Phi(F)).$$

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 $\nu$,

$$\nu(F+H) - \nu(F) = (F+H)_{Y|X=\widetilde{w}}(\widetilde{e}) - F_{Y|X=\widetilde{w}}(\widetilde{e})$$

$$= \frac{\int^{\widetilde{e}} f(s,\widetilde{w})\ ds + \int^{\widetilde{e}} h(s,\widetilde{w})\ ds}{f(\widetilde{w})+h(\widetilde{w})} - \frac{\int^{\widetilde{e}} f(s,\widetilde{w})\ ds}{f(\widetilde{w})}$$

$$= \frac{f(\widetilde{w})\int^{\widetilde{e}} h(s,\widetilde{w})\ ds - h(\widetilde{w})\int^{\widetilde{e}} f(s,\widetilde{w})\ ds}{f(\widetilde{w})\left(f(\widetilde{w})+h(\widetilde{w})\right)}$$

Let

$$A\widetilde{w} = f(\widetilde{w})\int^{\widetilde{e}} h(s,\widetilde{w})ds - h(\widetilde{w})\int^{\widetilde{e}} f(s,\widetilde{w})ds \text{ and}$$

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

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

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

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

By the Mean Value Theorem, there exist $r'_f$, between $\Phi(F)$ and $r_f$, such that

$f(r_f,w) - f(\Phi(F),w) = \frac{\partial f(r'_f,w)}{\partial y}\ (r_f - \Phi(F))$. Hence,

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

Let

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

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

Then,

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

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

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

Since, by the definition of $r_f$ and by (8),

$$|r_f-\Phi(F)|\le|\Phi(F+H)-\Phi(F)|\le c_1\|H\|,$$

it follows by (1) that, for some $a_4<\infty$,

$$(11)\ |R\Phi(F,H)|\le a_4\|H\|^2.$$

Moreover, by the definition of $D\Phi(F,H)$, there exists $a_5<\infty$ such that

$$(12)\ |D\Phi(F,H|\le a_6\|H\|.$$

Let $H=\widehat{F}-F$. Then,

$$(13)\ \widehat{n}(w,e)-n(w,e)=\Phi(\widehat{F})-\Phi(F)$$
$$=D\Phi(F,\widehat{F}-F)+R\Phi(F,\widehat{F}-F),$$
$$(14)\ \left|D\Phi(F,\widehat{F}-F)\right|\le a_6\left\|\widehat{F}-F\right\| \quad \text{and} \quad \left|R\Phi(F,\widehat{F}-F)\right|\le a_5\left\|\widehat{F}-F\right\|^2.$$

By Assumptions C.1-C.4 and Lemma B.3 in Newey (1994), $\left\|\widehat{F}-F\right\|\to 0$ in probability. Hence, by (13) and (14), it follows that $\widehat{n}(w,e)\to n(w,e)$ in probability. Hence, the estimator of $n(w,e)$ is consistent

Next, to derive the asymptotic distribution of $\widehat{n}(w,e)$, we note that by(10)-(12), $\Phi$ is Hadamard differentiable at $F$. It then follows by Theorem 3.9.4 in van der Vaart and Wellner (1996) that

$$\sqrt{N\sigma_N^{2\widetilde{L}}}\left(\Phi(\widehat{\widetilde{F}}_Y)-\Phi(\widetilde{F}_Y)\right)-D\Phi\left(\widetilde{F}_Y,\sqrt{N\sigma_N^{2\widetilde{L}}}(\widehat{\widetilde{F}}_Y-\widetilde{F}_Y)\right)$$

converges in outer probability to 0. Since

$$D\Phi(\widetilde{F}_Y,(\widehat{\widetilde{F}}_Y-\widetilde{F}_Y))=\frac{1}{f_{Y|W=w}(m(w,e))}\int\frac{\left[1(s\leq\widetilde{e})-F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})\right]}{f(\widetilde{w})}\left(\widehat{f}(s,\widetilde{w})-\widehat{f}(s,\widetilde{w})\right)ds$$
$$-\frac{1}{f_{Y|W=w}(m(w,e))}\int\frac{\left[1(s\leq m(w,e))-F_{Y|W=w}(m(w,e))\right]}{f(w)}\left(\widehat{f}(s,w)-\widehat{f}(s,w)\right)ds,$$

it follows by the definition of $n(w,e)$ and the Lemma in Appendix B that, if $w\neq\widetilde{w}$,

$$D\Phi(\widetilde{F}_Y,\sqrt{N\sigma_N^{2\widetilde{L}}}(\widehat{\widetilde{F}}_Y-\widetilde{F}_Y))\rightarrow N(0,V_n)$$

where $V_n=\{\int K(z)^2\}\left[\frac{1}{f_{Y|W=w}(m(w,e))}\right]^2\widetilde{L}$ and

$$\widetilde{L}=1[d_1=d]\int\left[\frac{1(s<\widetilde{e})}{f(\widetilde{w})}-\frac{F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})}{f(\widetilde{w})}\right]^2 f(s,\widetilde{w})\,ds$$
$$+1[d_2=d]\int\left[\frac{1(s<m(w,e))}{f(w)}-\frac{F_{Y|W=w}(m(w,e))}{f(w)}\right]^2 f(s,w)\,ds$$
$$=\frac{1[d_1=d]}{f(\widetilde{w})}F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})(1-F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))+\frac{1[d_2=d]}{f(w)}F_{Y|W=w}(m(w,e))(1-F_{Y|W=w}(m(w,e)))$$
$$=\left[\frac{1[d_1=d]}{f(\widetilde{w})}+\frac{1[d_2=d]}{f(w)}\right]F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e})(1-F_{Y|\widetilde{W}=\widetilde{w}}(\widetilde{e}))$$

where the last equality follows by the definition of $n(w,e)$. Hence,

$$\sqrt{N}\,\sigma_N^{\widetilde{L}/2}\,(\hat{m}(w,e)-m(w,e))=\sqrt{N}\,\sigma_N^{\widetilde{L}/2}\left(\Phi(\hat{F})-\Phi(F)\right)\rightarrow N(0,V_n)$$

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

**Proof of Theorem 3**: Consider first the case where Assumptions (A.i)-(A.vi) are satisfied. Without loss of generality, we will show the identification of the distribution of $\varepsilon_1$, 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 $\varepsilon_1$, 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 $\varepsilon_k$, 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 $\varepsilon$ 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. Without loss of generality, we show this for $k = 1$. Note that when $X = (w_0, w_1, \widetilde{w}_2, ..., \widetilde{w}_K) = (w_0, w^k)$, $Y = m(X, \varepsilon) = r_1 (n_1(w_{0_1}, w_1, \varepsilon_1))$. Hence, using the conditional independence between $\varepsilon$ and $X_1$, and the strict monotonicity of $n_1$ in $\varepsilon_1$ it follows that

$$\Pr(\varepsilon_1 \leq \eta | X_0 = w_0)$$

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

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

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

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

Since, as we have shown above,

$$\Pr(\varepsilon_1 \leq \eta | W_0 = w_0) = \Pr\left(Y \leq r_1(\eta) | X = (w_0, \overline{w}^k)\right)$$

it follows that

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

Partition $w_0$ as $w_0 = (w_{0_1}, w_{0_{-1}})$. Note that

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

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

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

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

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

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

Hence, since

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

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 first part of the theorem.

Suppose now that Assumptions (A.i)-(A.iv), (A.v') and (A.vi') are satisfied. Then, the results in the first part of the theorem, together with the added conditions that $\varepsilon$ is independent of $X$, imply that, say, for $k = 1$ and $y = r_1(\eta)$,

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

Using the fact that

$$\begin{aligned}
&\Pr\left(Y \leq y | X = (w_0, \overline{w}^k)\right) \\
&\quad = \int F_{Y|X=(w_0,\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}(y)\ f_{w_0|w=(\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}(w_0)\ dw_0 \\
&\quad = \int \left[\int^y \frac{f(s,w_0,\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}{f(w_0,\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}\ ds\right] \frac{f(w_0,\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}{f(\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}\ dw_0 \\
&\quad = \int^y \int \frac{f(s,w_0,\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}{f(\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}\ dw_0\ ds \\
&\quad = \int^y \frac{f(s,\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}{f(\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}\ ds \\
&\quad = \int^y \frac{f(s,\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}{f(\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}\ ds \\
&\quad = F_{Y|w=(\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}(y),
\end{aligned}$$

it follows that

$$F_{\varepsilon_1}(\eta) = \Pr\left(\varepsilon_1 \leq \eta\right) = F_{Y|w=(\overline{w}_1,\widetilde{w}_2,...,\widetilde{w}_K)}(r_1(\eta)).$$

Since, $\Pr\left(\varepsilon_1 \leq \eta\right) = \Pr\left(\varepsilon_1 \leq \eta | X_0 = w_0\right)$ and, as shown in the first part of the proof,

$$\Pr\left(\varepsilon_1 \leq \eta | X_0 = w_0\right) = F_{Y|X=(w_{0_1},w_1,\widetilde{w}_2,...,\widetilde{w}_K)}\left(r_1\left(n_1(w_{0_1}, w_1, \eta)\right)\right),$$

$$\begin{aligned}
n_1(w_{0_1}, w_1, \eta) &= r_1^{-1}\left(F^{-1}_{Y|X=(w_{0_1},w_1,\widetilde{w}_2,...,\widetilde{w}_K)}\left(F_{\varepsilon_1}(\eta)\right)\right) \\
&= r^{-1}\left(F^{-1}_{Y|X=(w_{0_1},w^k)}\left(F_{Y|w=\overline{w}^k}\left(r_1(\eta)\right)\right)\right)
\end{aligned}$$

The argument for $k \neq 1$ is analogous. Hence, for each $k$, $F_{\varepsilon_k}$ and $n_k$ are identified. Since the $\varepsilon_k$'s are independent across $k$, this implies that the joint distribution of $(\varepsilon_1, ..., \varepsilon_K)$ is identified. This completes the proof.

**Proof of Theorem 5:** Let $\widetilde{x}$ and $\widetilde{e}$ be as defined in Section 4. Let the functional $\Lambda$ be defined as in the proof of Theorem 1 for $W = X$, $y = \widetilde{e}$ and $w = \widetilde{x}$. Let the functional $\Phi$ be defined as in the proof of Theorem 2. Hence, $\Lambda(G) = G_{Y|X=\widetilde{x}}(\widetilde{e})$ and $\Phi(G) = G^{-1}_{Y|X=x}\left(G_{Y|X=\widetilde{x}}(\widetilde{e})\right)$, where $G : R^{1+L} \to R$, and for all $y, x, z$, $G_{Y|X=x}(y) = \int_{-\infty}^{y} g(s,x)ds/g(x)$, $g(s,z) = \partial^{L+1}G(s,z)/\partial s\partial z_1 \cdots \partial z_L$ and $g(x) = \int_{-\infty}^{\infty} g(s,x)ds$. For all such $G$, define the functional $\Xi$ by

$$\Xi(G) = \frac{1}{g(\Phi(G),x)} \frac{\partial g(x)}{\partial x} G_{Y|X=\widetilde{x}}(\widetilde{e}) - \frac{1}{g(\Phi(G),x)} \int^{\Phi(G)} \frac{\partial g(s,x)}{\partial x} ds.$$

Then, $\Xi(F) = \frac{\partial m(x,e)}{\partial x}$ and $\Xi(\widehat{F}) = \frac{\partial \widehat{m(x,e)}}{\partial x}$ . Define the functionals $\mu$ , $\beta$, and $\gamma$ by

$$\mu(G) = \frac{1}{g(\Phi(G),x)}, \ \beta(G) = \frac{\partial g(x)}{\partial x}, \text{ and } \gamma(G) = \int^{\Phi(G)} \frac{\partial g(s,x)}{\partial x} ds.$$

Then,

(1) $\Xi(G) = \mu(G)\, \beta(G)\, \Lambda(G) - \mu(G)\, \gamma(G)$.

Let $\|G\|$ denote the sup norm of $\frac{\partial g(s,x)}{\partial x}$. Let $H$ be such that $H$ vanishes outside a compact set, $H$ is differentiable up to the $2L+1$. Let $\rho > 0$ be such that if $\|H\| \leq \rho$, then for some $a, b, d < \infty$

(2) $|h(x)| \leq a \|H\|$, $\left|\int_{-\infty}^{y} h(s,x)ds\right| \leq a \|H\|$,

$|f(x) + h(x)| \geq b\,|f(x)|$, $\quad f(s,x) + h(s,x) \geq b\,|f(s,x)|$, and

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

The existence of such a $\rho$ is guaranteed by (1) and (3) in Theorem 2. Consider the $H$'s that satisfy $\|H\| \leq \rho$.

Analogously to the main arguments in the proofs of Theorems 1 and 2, we will derive the asymptotic behavior of $\Xi(\widehat{F})$ by first obtaining a first order Taylor expansion of $\Xi(F+H)$, and then letting $H = \widehat{F} - F$. With this aim, we will first obtain first order Taylor expansions for $\mu\,(F+H)$, $\beta(F+H)$, $\Lambda(F+H)$, and $\gamma(F+H)$.

First note that, by the proof of Theorem 2,

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

where

$$(4)\ D\Phi(F,H) = \frac{1}{f_{Y|X=x}(m(x,e))} \int\int \frac{\left[1(s \leq \widetilde{e}) - F_{Y|X=\widetilde{x}}(\widetilde{e})\right]}{f(\widetilde{x})} 1_{(s,\widetilde{x})}(s,z)\, h(s,z)\, ds\, dz$$
$$- \frac{1}{f_{Y|X=x}(m(x,e))} \int\int \frac{\left[1(s \leq m(x,e)) - F_{Y|X=x}(m(x,e))\right]}{f(x)} 1_{(s,x)}(s,z)\, h(s,z)\, ds\, dz,$$

and for some $c_1 < \infty$,

(5) $|D\Phi(F,H)| \leq c_1 \|H\|$ and $|R\Phi(F,H)| \leq c_1 \|H\|^2$.

By (1) and (2) in the proof of Theorem 1,

(6) $\Lambda(F+H) - \Lambda(F) = D\Lambda(F,H) + R\Lambda(F,H)$, where

$$(7)\ D\Lambda(F,H) = \frac{\int_{-\infty}^{\widetilde{e}} h(s,\widetilde{x})ds - h(\widetilde{x})\ F_{Y|X=\widetilde{x}}(\widetilde{e})}{f(\widetilde{x})},$$

and for some $c_2 < \infty$,

$$(8)\ |D\Lambda(F,H)| \leq c_2 \|H\| \ \text{ and } \ |R\Lambda(F,H)| \leq c_2 \|H\|^2 .$$

To obtain a first order Taylor expansion for $\mu(F+H)$, note that

$$\mu(F+H) - \mu(F) = \frac{1}{f(\Phi(F+H),x) + h(\Phi(F+H),x)} - \frac{1}{f(\Phi(F+H),x)},$$

$$= \frac{f(\Phi(F),x) - f(\Phi(F+H),x) - h(\Phi(F+H),x)}{[f(\Phi(F+H),x) + h(\Phi(F+H),x)]f(\Phi(F+H),x)}.$$

where $\mu(F+H)$ is well defined by (2). Let $\Delta\Phi = \Phi(F+H) - \Phi(F)$. By Taylor's Theorem, there exist $d_1, d_2 < \infty,$ such that

$$f(\Phi(F+H),x) - f(\Phi(F),x) = \frac{\partial f(\Phi(F)),x)}{\partial y}\Delta\Phi + \operatorname{Re} m_1, \text{ and}$$

$$h(\Phi(F+H),x) - h(\Phi(F),x) = \frac{\partial h(\Phi(F)),x)}{\partial y}\Delta\Phi + \operatorname{Re} m_2, \text{ where}$$

$\operatorname{Re} m_1 \leq d_1 |\Delta\Phi|^2$ and $\operatorname{Re} m_2 \leq d_2 |\Delta\Phi|^2$ . Then,

$$f(\Phi(F),x) - f(\Phi(F+H),x) - h(\Phi(F+H),x)$$

$$= -\frac{\partial f(\Phi(F)),x)}{\partial y}\Delta\Phi - \operatorname{Re} m_1 - h(\Phi(F),x) - \frac{\partial h(\Phi(F)),x)}{\partial y}\Delta\Phi - \operatorname{Re} m_2$$

and

$$\frac{1}{[f(\Phi(F+H),x) + h(\Phi(F+H),x)]f(\Phi(F+H),x)}$$

$$= \frac{1}{f(\Phi(F),x)^2} + \left[\frac{1}{\left[f(\Phi(F),x) + \frac{\partial f(\Phi(F)),x)}{\partial y}\Delta\Phi + \operatorname{Re} m_1 + h(\Phi(F),x) + \frac{\partial h(\Phi(F)),x)}{\partial y}\Delta\Phi + \operatorname{Re} m_2\right]f(\Phi(F),x)} - \frac{1}{f(\Phi(F),x)^2}\right]$$

$$= \frac{1}{f(\Phi(F),x)^2} + \left[\frac{-\frac{\partial f(\Phi(F)),x)}{\partial y}\Delta\Phi - \operatorname{Re} m_1 - h(\Phi(F),x) - \frac{\partial h(\Phi(F)),x)}{\partial y}\Delta\Phi - \operatorname{Re} m_2}{\left[f(\Phi(F),x) + \frac{\partial f(\Phi(F)),x)}{\partial y}\Delta\Phi + \operatorname{Re} m_1 + h(\Phi(F),x) + \frac{\partial h(\Phi(F)),x)}{\partial y}\Delta\Phi + \operatorname{Re} m_2\right]f(\Phi(F),x)^3}\right]$$

Let

$$(9)\ D\mu(F,H) = \frac{-\frac{\partial f(\Phi(F)),x)}{\partial y}D\Phi(F,H) - h(\Phi(F),x)}{f(\Phi(F),x)^2}$$

and

$$R\mu(F,H) = \frac{-\frac{\partial f(\Phi(F)),x)}{\partial y}R\Phi(F,H) - \operatorname{Re} m_1 - \frac{\partial h(\Phi(F)),x)}{\partial y}\Delta\Phi - \operatorname{Re} m_2}{f(\Phi(F),x)^2}$$

$$+ \frac{\left[-\frac{\partial f(\Phi(F)),x)}{\partial y}\Delta\Phi - \operatorname{Re} m_1 - h(\Phi(F),x) - \frac{\partial h(\Phi(F)),x)}{\partial y}\Delta\Phi - \operatorname{Re} m_2\right]^2}{\left[f(\Phi(F),x) + \frac{\partial f(\Phi(F)),x)}{\partial y}\Delta\Phi + \operatorname{Re} m_1 + h(\Phi(F),x) + \frac{\partial h(\Phi(F)),x)}{\partial y}\Delta\Phi + \operatorname{Re} m_2\right]f(\Phi(F),x)^3}.$$

Then,

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

and for some $c_3<\infty$,

(11) $|D\mu|\leq c_3\|H\|$ and $|R\mu|\leq c_3\|H\|^2.$

Next, letting

(12) $D\beta(F,H)=\frac{\partial h(x)}{\partial x},$

it follows that

(13) $\beta(F+H)-\beta(F)=D\beta(F,H)$

and for some $c_4<\infty$,

(14) $|D\beta(F,H)|=c_4\|H\|.$

To obtain a first order Taylor expansion for $\gamma(F+H)$, we note that

$$\gamma(F+H)-\gamma(F)=\int^{\Phi(F+H)}\frac{\partial f(s,x)}{\partial x}ds+\int^{\Phi(F+H)}\frac{\partial h(s,x)}{\partial x}ds-\int^{\Phi(F)}\frac{\partial f(s,x)}{\partial x}ds$$

By Taylor's Theorem, there exists $d_3$ and $d_4$ such that

$$\int_{\Phi(F)}^{\Phi(F+H)}\frac{\partial f(s,x)}{\partial x}\,ds=\frac{\partial f(\Phi(F),x)}{\partial x}D\Phi+\operatorname{Re}m_3,\text{ and}$$

$$\int_{\Phi(F)}^{\Phi(F+H)}\frac{\partial h(s,x)}{\partial x}\,ds=\frac{\partial h(\Phi(F),x)}{\partial x}D\Phi+\operatorname{Re}m_4,$$

where for $i=3,4$, $|\operatorname{Re}m_i|\leq d_i|D\Phi|^2$. Hence, since

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

it follows that

(15) $\gamma(F+H)-\gamma(F)=D\gamma(F,H)+R\gamma(F,H)$, where

(16) $D\gamma(F)=\frac{\partial f(\Phi(F),x)}{\partial x}D\Phi(F)+\int_{\infty}^{\Phi(F)}\frac{\partial h(s,x)}{\partial x}ds$ and

$$R\gamma(F)=\frac{\partial f(\Phi(F),x)}{\partial x}R\Phi(F)+\frac{\partial h(\Phi(F),x)}{\partial x}D\Phi(F)+\operatorname{Re}m_3+\operatorname{Re}m_4.$$

It then follows by (3) and (5) that there exists $c_5<\infty$ such that

(17) $|D\gamma(F)|\leq c_5\|H\|$ and $|R\gamma(F)|\leq c_5\|H\|^2.$

We are now in a position to obtain a first order Taylor expansion for $\Xi(F+H)$. Denote $\mu(F), \beta(F), \gamma(F), \Lambda(F), \Phi(F)$, and $\Xi(F)$ by $\mu, \beta, \gamma, \Lambda, \Phi$, and $\Xi$, respectively, and denote $Dw(F,H)$ and $Rw(F,H)$ by $Dw$ and $Rw$, respectively, for $w = \mu, \beta, \gamma, \Lambda, \Phi, \Xi$. It is easy to show that

(18) $\Xi(F+H) - \Xi(F) = D\Xi + R\Xi,$

where

$$D\Xi = D\mu\, \beta\, \Lambda + \mu\, D\beta\, \Lambda + \mu\, \beta\, D\Lambda - D\mu\, \gamma - \mu\, D\gamma, \text{ and}$$

$$R\Xi = \mu\, \beta\, R\Lambda + \mu\, D\beta(D\Lambda + R\Lambda) + D\mu\, \beta\, (D\Lambda + R\Lambda)$$

$$+ D\mu\, D\beta\, \Lambda(F+H) + R\mu\, \beta(F+H)\, \Lambda(F+H) - \mu\, R\gamma - R\mu\, \gamma$$

$$-(D\mu + R\mu)(D\gamma + R\gamma).$$

By (8), (11), (14), and (17), there exists $c_6 < \infty$ such that

(19) $|D\Xi| \leq |D\mu|\,|\beta\, \Lambda| + |\mu\Lambda|\,|D\beta| + |\mu\beta|\,|D\Lambda| + |D\mu|\,|\gamma| + |\mu|\,|D\gamma|$

$$\leq c_6 \left\| H \right\|,$$

and

(20) $|R\Xi| \leq c_6 \left\| H \right\|^2.$

We will now use this first order Taylor expansion of $\Xi$ to show the consistency of $\partial m(\widehat{x,e})/\partial x$ and to derive its asymptotic distribution. Let $H = \widehat{F} - F$. Then, $\Xi(F+H) - \Xi(F) = \Xi(\widehat{F}) - \Xi(F) = \partial m(\widehat{x,e})/\partial x - \partial m(x,e)/\partial x$. By (19) and (20),

$$\left| \frac{\partial \widehat{m}(x,e)}{\partial x} - \frac{\partial m(x,e)}{\partial x} \right| \leq c_6 \left\| \widehat{F} - F \right\| + c_6 \left\| \widehat{F} - F \right\|^2$$

and by Assumptions C.1-C.3 and C.4', $\left\| \widehat{F} - F \right\| \to 0$ in probability. Hence, $\partial m(\widehat{x,e})/\partial x \to \partial m(x,e)/\partial x$ in probability. By (18)-(20), $\Xi$ is Hadamard differentiable at $F$. It then follows by Theorem 3.9.4 in van der Vaart and Wellner (1996) and the Lemma in Appendix B that

$$\sqrt{N}\sigma_N^{L/2+1} \left( \Xi(\widehat{F}) - \Xi(F) \right) - D\Xi \left( F, \sqrt{N}\sigma_N^{(L/2)+1} \left( \widehat{F} - F \right) \right)$$

converges in outer probability to 0. Let

$$D_1\Xi = \frac{1}{f(\Phi(F),x)} F_{Y|X=\widetilde{x}}(\widetilde{e}) \frac{\partial h(x)}{\partial x} - \frac{1}{f(\Phi(F),x)} \int^{\Phi(F)} \frac{\partial h(s,x)\, ds}{\partial x}$$

and

$$D_2\Xi = D\Xi - D_1\Xi.$$

Since

$$D_1\Xi = \frac{F_{Y|X=x}(m(x,e))}{f(m(x,e),x)} \left( \frac{\partial \widehat{f}(x)}{\partial x} - \frac{\partial f(x)}{\partial x} \right) - \frac{1}{f(m(x,e),x)} \int^{m(x,e)} \left( \frac{\partial \widehat{f}(s,x)}{\partial x} - \frac{\partial f(s,x)}{\partial x} \right) \; ds,$$

it follows, by the Lemma in Appendix B, that

$$\sqrt{N}\sigma_N^{(L/2)+1} D_1\Xi \to N(0, V_{\partial x})$$

where

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

Using the same Lemma and the definition of $D_2\Xi$, it is easy to show that

$$D_2\Xi \left( F, \sqrt{N}\sigma_N^{(L/2)+1} \left( \widehat{F} - F \right) \right) \to 0 \quad \text{in probability.}$$

Hence,

$$D\Xi \left( F, \sqrt{N}\sigma_N^{(L/2)+1} \left( \widehat{F} - F \right) \right)$$

$$= D\Xi_1 \left( F, \sqrt{N}\sigma_N^{(L/2)+1} \left( \widehat{F} - F \right) \right) + D\Xi_2 \left( F, \sqrt{N}\sigma_N^{(L/2)+1} \left( \widehat{F} - F \right) \right)$$

$$\to N(0, V_{\partial x}).$$

It then follows that

$$\sqrt{N}\sigma_N^{(L/2)+1} \left( \frac{\widehat{\partial m(x,e)}}{\partial x} - \frac{\partial m(x,e)}{\partial x} \right) = \sqrt{N}\sigma_N^{L/2+1} \left( \Xi(\widehat{F}) - \Xi(F) \right) \to N(0, V_{\partial x})$$

in distribution, where

$$V_{\partial x} = \frac{F_{Y|X=x}(m(x,e)) \left(1-F_{Y|X=x}(m(x,e))\right)}{f_{Y|X=x}(m(x,e))^2 f(x)} \left[ \int \left( \int \frac{\partial K(s,z)}{\partial z} ds \right) \left( \int \frac{\partial K(s,z)}{\partial z} ds \right)' dz \right]$$

The result of the Theorem then follows by noticing that

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

**Proof of Theorem 6:** As in the proof of Theorem 5, we let $\Lambda$, $\Phi$, and $\mu$ denote the functionals defined by $\Lambda(G) = G_{Y|X=\widetilde{x}}(\widetilde{e})$, $\Phi(G) = G^{-1}_{Y|X=x}\left(G_{Y|X=\widetilde{x}}(\widetilde{e})\right)$, and $\mu(G) = 1/f(\Phi(G), x)$, where $G : R^{1+L} \to R$, and for all $y, x, z$, $G_{Y|X=x}(y) = \int_{-\infty}^{y} g(s,x)ds/g(x)$, $g(s,z) = \partial^{L+1}G(s,z)/\partial s \partial z_1 \cdots \partial z_L$ and $g(x) = \int_{-\infty}^{\infty} g(s,x)ds$. For all such $G$, we define the functionals

$$\widetilde{\beta}(G) = \left( \frac{\partial f(\widetilde{x})}{\partial x} \right)' \left( \frac{\overline{x}}{\overline{\varepsilon}} \right), \;\; \widetilde{\gamma}(G) = \int^{\widetilde{e}} \left( \frac{\partial f(s,\widetilde{x})}{\partial x} \right)' \left( \frac{\overline{x}}{\overline{\varepsilon}} \right) ds,$$

$\nu(G) = \frac{1}{f(\widetilde{x})}$, $\eta_1(G) = f(x)$, and $\eta_2(G) = f(\widetilde{e}, \widetilde{x})$.

Define also the functionals

$$\Psi_1(G) = \mu(G)\,\eta_1(G)\,\eta_2(G)\,\nu(G)\,\left(\frac{\alpha}{\overline{\varepsilon}}\right) + \mu(G)\,\eta_1(G)\,\nu(G)\,\widetilde{\gamma}(G)$$
$$-\mu(G)\,\eta_1(G)\,\nu(G)\;\Lambda(G)\,\widetilde{\beta}(G).$$

and

$$\Psi_2(G) = \mu(G)\,\eta_1(G)\,\eta_2(G)\,\nu(G)$$

Then, when $m$ satisfies specification (5.II), $\widehat{\frac{\partial m(x,e)}{\partial \varepsilon}} = \Psi_1(\widehat{F})$ and $\frac{\partial m(x,e)}{\partial \varepsilon} = \Psi_1(F)$, while when $m$ satisfies specification (5.I), $\widehat{\frac{\partial m(x,e)}{\partial \varepsilon}} = \Psi_2(\widehat{F})$ and $\frac{\partial m(x,e)}{\partial \varepsilon} = \Psi_2(F)$. To derive the asymptotic properties of $\Psi_1(\widehat{F})$ and $\Psi_2(\widehat{F})$, we first obtain first order Taylor expansions for $\Psi_1(F + H)$ and $\Psi_2(F + H)$.

Let $\|G\|$ denote the sup norm of $\partial g(s,x)/\partial x$. Let $H$ be such that $H$ vanishes outside a compact set, $H$ is differentiable up to the order $2L + 1$, and $\|H\|$ is sufficiently small so that, for some $a, b, d < \infty$

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

$|f(x) + h(x)| \geq b\,|f(x)|\,,\quad f(s,x) + h(s,x) \geq b\,|f(s,x)|\,,$ and

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

Equations (3)-(11) in the proof of Theorem 5 provide first order Taylor expansions for the functionals $\Phi$, $\mu$, and $\Lambda$. Using similar arguments as in the proof of that theorem, it is easy to establish that

(2) $\widetilde{\beta}(F + H) - \widetilde{\beta}(F) = D\widetilde{\beta}(F, H),$

where

(3) $D\widetilde{\beta}(F,H) = \left(\frac{\partial h(\widetilde{x})}{\partial x}\right)' \left(\frac{\overline{x}}{\overline{\varepsilon}}\right),$

and for some $k_1 < \infty$,

(4) $\left|D\widetilde{\beta}(F,H)\right| = k_1\,\|H\|\,;$

(5) $\widetilde{\gamma}(F + H) - \widetilde{\gamma}(F) = D\widetilde{\gamma}(F, H),$

where

(6) $D\widetilde{\gamma}(F,H) = \int^{\widetilde{e}} \left(\frac{\partial h(\widetilde{x})}{\partial x}\right)' \left(\frac{\overline{x}}{\overline{\varepsilon}}\right) ds,$

and for some $k_2 < \infty$,

(7) $|D\widetilde{\gamma}(F, H)| = k_2 \|H\|$;

(8) $\nu(F + H) - \nu(F) = D\nu(F, H) + R\nu(F, H)$,

where

(9) $D\nu(F, H) = \frac{-h(\widetilde{x})}{f(\widetilde{x})^2}$, $R\nu(F, H) = \frac{-h(\widetilde{x})^2}{f(\widetilde{x})^2\left(f(\widetilde{x})+h(\widetilde{x})\right)}$,

and for some $k_3 < \infty$,

(10) $|Dv(F, H)| = k_3 \|H\|$, and $|Rv(F, H)| = k_3 \|H\|$;

and

(11) $\eta_i(F + H) - \eta_i(F) = D\eta_i(F, H) \quad i = 1, 2$

where

(12) $D\eta_1 = h(x)$, $D\eta_2 = h(\widetilde{e}, \widetilde{x})$,

and for some $k_4 < \infty$,

(13) $|D\eta_i(F, H)| \leq k_4 \|H\| \qquad i = 1, 2.$

Denote $\mu(F), \widetilde{\beta}(F), \widetilde{\gamma}(F), \Lambda(F), \Phi(F), \nu(F), \eta_1(F)$, and $\eta_2(F)$ by $\mu, \widetilde{\beta}, \widetilde{\gamma}, \Lambda, \Phi, \nu, \eta_1$, and $\eta_2$, respectively, and denote $Dw(F, H)$ by $Dw$, for $w = \mu, \widetilde{\beta}, \widetilde{\gamma}, \Lambda, \Phi, \nu, \eta_1, \eta_2$. Define

$D\Psi_1(F, H) =$

$$D\mu \ \eta_1 \eta_2 \nu \left(\frac{\alpha}{\overline{\overline{\varepsilon}}}\right) + \mu D\eta_1 \eta_2 \nu \left(\frac{\alpha}{\overline{\overline{\varepsilon}}}\right) + \mu \eta_1 D \eta_2 \nu \left(\frac{\alpha}{\overline{\overline{\varepsilon}}}\right) + \mu \eta_1 \eta_2 D\nu \left(\frac{\alpha}{\overline{\overline{\varepsilon}}}\right)$$

$$+D\mu \eta_1 \nu \widetilde{\gamma} + \mu D\eta_1 \nu \widetilde{\gamma} + \mu \eta_1 D\nu \widetilde{\gamma} + \mu \eta_1 \nu D\widetilde{\gamma}$$

$$-D\mu \eta_1 \nu \Lambda \widetilde{\beta} - \mu D\eta_1 \nu \Lambda \widetilde{\beta} - \mu \eta_1 D\nu \Lambda \widetilde{\beta}$$

$$-\mu \eta_1 \nu D\Lambda \widetilde{\beta} - \mu \eta_1 \nu \Lambda D\widetilde{\beta}.$$

(14) $R\Psi_1(F, H) = \Psi_1(F + H) - \Psi_1(F) - D\Psi_1(F, H)$,

$$D\Psi_2(F, H) = D\mu \ \eta_1 \eta_2 \nu + \mu D\eta_1 \eta_2 \nu + \mu \eta_1 D \eta_2 \nu + \mu \eta_1 \eta_2 D\nu,$$

and

(15) $R\Psi_2(F,H) = \Psi_2(F+H) - \Psi_2(F) - D\Psi_2(F,H)$.

Using equations (2)-(13) above and equations (3)-(11) in the proof of Theorem 5, it is easy to show that there exists $k_5, k_6 < \infty$ such that

(16) $|D\Psi_1(F,H)| \leq k_5 \|H\|$, $|R\Psi_1(F,H)| \leq k_5 \|H\|^2$, and

(17) $|D\Psi_2(F,H)| \leq k_6 \|H\|$, and $|R\Psi_2(F,H)| \leq k_6 \|H\|^2$.

We will first use the first order Taylor expansion, (14) and (16), to derive the asymptotic properties of $\widehat{\partial m(x,e)}/\partial\varepsilon$ when $m$ satisfies specification (II). Let $H = \widehat{F} - F$. Then, $\Psi_1(F+H) - \Psi_1(F) = \widehat{\partial m(x,e)}/\partial\varepsilon - \partial m(x,e)/\partial\varepsilon$. By (14), (16) and Assumptions (C.1)-(C.3) and (C.4'), it then follows, by the arguments analogous to the ones used in the proofs of the previous theorems, that $\widehat{\partial m(x,e)}/\partial\varepsilon \to \partial m(x,e)/\partial\varepsilon$ in probability. Also by (14) and (16), $\Psi_1$ is Hadamard differentiable at $F$. Let

$$D_1\Psi_1(F, F+H) = \mu\,\eta_1\,\nu\, D\widetilde{\gamma} - \mu\,\eta_1\,\nu\,\Lambda\, D\widetilde{\beta} \quad \text{and} \quad D_2\Psi_1 = D\Psi_1 - D_1\Psi_1.$$

Since, when $H = \widehat{F} - F$,

$$\begin{aligned} D_1\Psi_1(F, \widehat{F} - F) &= \frac{\int^{\widetilde{e}} \left(\frac{\partial h(\widetilde{x})}{\partial x}\right)' \binom{\overline{x}}{\overline{\overline{\varepsilon}}} ds}{f_{Y|X=x}(m(x,e))\; f(\widetilde{x})} - \frac{F_{Y|X=\widetilde{x}}(\widetilde{e}) \left(\frac{\partial h(\widetilde{x})}{\partial x}\right)' \binom{\overline{x}}{\overline{\overline{\varepsilon}}}}{f_{Y|X=x}(m(x,e))\; f(\widetilde{x})} \\ &= \frac{\int^{\widetilde{e}} \left(\frac{\partial \widehat{f}(\widetilde{x})}{\partial x} - \frac{\partial f(\widetilde{x})}{\partial x}\right)' \binom{\overline{x}}{\overline{\overline{\varepsilon}}} ds}{f_{Y|X=x}(m(x,e))\; f(\widetilde{x})} - \frac{F_{Y|X=\widetilde{x}}(\widetilde{e}) \left(\frac{\partial \widehat{f}(\widetilde{x})}{\partial x} - \frac{\partial f(\widetilde{x})}{\partial x}\right)' \binom{\overline{x}}{\overline{\overline{\varepsilon}}}}{f_{Y|X=x}(m(x,e))\; f(\widetilde{x})} \end{aligned}$$

it follows, by applying the Lemma in Appendix B, that

$$D_1\Psi_1(F, \sqrt{N}\sigma_N^{L/2+1}(\widehat{F} - F)) \to N(0, V_{II,\partial\varepsilon})$$

in distribution, where

$$V_{II,\partial\varepsilon} = \frac{F_{Y|X=\widetilde{x}}(\widetilde{e})\;(1 - F_{Y|X=\widetilde{x}}(\widetilde{e}))}{f_{Y|X=x}(m(x,e))^2 f(\widetilde{x})} \left\{ \binom{\overline{x}}{\overline{\overline{\varepsilon}}}' \left[ \int \left( \int \frac{\partial K(s,z)}{\partial z} ds \right) \left( \int \frac{\partial K(s,z)}{\partial z} ds \right)' dz \right] \binom{\overline{x}}{\overline{\overline{\varepsilon}}} \right\}.$$

while

$$D_2\Psi_1(F, \sqrt{N}\sigma_N^{L/2+1}(\widehat{F} - F)) \to 0 \;\text{ in probability}$$

Since, by Theorem 3.9.4 in van der Vaar and Wellner (1996),

$$\sqrt{N}\sigma_N^{L/2+1}\left(\Psi_1(\widehat{F}) - \Psi_1(F)\right) - \sqrt{N}\sigma_N^{L/2+1}\left(D_1\Psi_1(F, \widehat{F} - F)\right)$$

converges in outer probability to 0, the above implies that

$$\sqrt{N}\sigma_N^{(L/2)+1}\left(\frac{\widehat{\partial m(x,e)}}{\partial\varepsilon} - \frac{\partial m(x,e)}{\partial\varepsilon}\right) = \sqrt{N}\sigma_N^{L/2+1}\left(\Psi_1(\widehat{F}) - \Psi_1(F)\right) \to N(0, V_{II,\partial\varepsilon})$$

in distribution, where

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

The statement in the Theorem follows by noticing that $F_{Y|X=\widetilde{x}}(\widetilde{e}) = F_\varepsilon(e)$.

We next use the first order Taylor expansion, (15) and (17), to derive the asymptotic properties of $\partial \widehat{m(x,e)}/\partial\varepsilon$ when $m$ satisfies specification (5.I). Let $H = \widehat{F} - F$. Then, $\Psi_2(F+H) - \Psi_2(F) = \partial \widehat{m(x,e)}/\partial\varepsilon - \partial m(x,e)/\partial\varepsilon$. By (15), (17) and Assumptions (C.1)-(C.3) and (C.4'), it follows, as previously, that $\partial \widehat{m(x,e)}/\partial\varepsilon \to \partial m(x,e)/\partial\varepsilon$ in probability. Also by (15) and (17), $\Psi_2$ is Hadamard differentiable at $F$. Let

$$D_1\Psi_2 = -\frac{h(m(x,e),x)}{f(m(x,e),x)^2}\frac{f(\widetilde{e},\widetilde{x})\,f(x)}{f(\widetilde{x})} + \frac{h(\widetilde{e},\widetilde{x})\,f(x)}{f(m(x,e),x)\,f(\widetilde{x})}.$$

By following arguments similar to the ones used earlier, and using the Lemma in Appendix B, we can show that the asymptotic distribution of $\partial\widehat{m}(x,e)/\partial\varepsilon - \partial m(x,e)/\partial\varepsilon$ will be driven by $D_1\Psi_2$, because in the terms in $D_1\Psi_2$, all the $L+1$ coordinates in the argument of the function $h$ are fixed, while in the other terms of $D\Psi_2$, only $L$ coordinates are fixed. Since

$$D_1\Psi_2(F,\widehat{F}-F) = -\frac{\left(\widehat{f}(m(x,e),x) - f(m(x,e),x)\right)}{f(m(x,e),x)^2}\frac{f(\widetilde{e},\widetilde{x})\,f(x)}{f(\widetilde{x})} + \frac{\left(\widehat{f}(\widetilde{e},\widetilde{x}) - f(\widetilde{e},\widetilde{x})\right)\,f(x)}{f(m(x,e),x)\,f(\widetilde{x})}$$

the results of the Lemma in Appendix B imply that

$$D_1\Psi_2\left(F, \sqrt{N}\sigma_N^{(L+1)/2}\left(\widehat{F}-F\right)\right) \to N(0, V_{I,\partial\varepsilon})$$

where, since $e \neq m(x,e)$,

$$\begin{aligned} V_{I,\partial\varepsilon} &= \left(\int K(s,z)^2 ds dz\right)\left\{ \frac{f(\widetilde{e},\widetilde{x})}{f_{Y|X=x}(m(x,e))^2\, f(\widetilde{x})^2} + \frac{f_{Y|X=\widetilde{x}}(\widetilde{e})^2 f(m(x,e),x)}{f_{Y|X=x}(m(x,e))^4 f(x)^2} \right\} \\ &= \left(\int K(s,z)^2 ds dz\right)\left\{ \frac{f_{Y|X=\widetilde{x}}(\widetilde{e})}{f_{Y|X=x}(m(x,e))^2\, f(\widetilde{x})} + \frac{f_{Y|X=\widetilde{x}}(\widetilde{e})^2}{f_{Y|X=x}(m(x,e))^3 f(x)} \right\}. \end{aligned}$$

By the Delta method, it then follows that

$$\begin{aligned} \sqrt{N}\sigma_N^{(L+1)/2}\left(\frac{\partial \widehat{m(x,e)}}{\partial e} - \frac{\partial m(x,e)}{\partial e}\right) &= \sqrt{N}\sigma_N^{(L+1)/2}\left(\Psi_2(\widehat{F}) - \Psi_2(F)\right) \\ &\to N(0, V_{I,\partial\varepsilon}) \end{aligned}$$

Finally, we note that the asymptotic properties of $\partial \widehat{m(x,e)}/\partial\varepsilon$ when $m$ satisfy specification (5.III) follow from Theorem 5.

# Appendix B

The following Lemma presents well known results about kernel estimators (see, for example, Newey (1994)). They are presented here for completeness.

LEMMA: Suppose that the following assumptions are satisfied:

(i) $\{y_i, x_i\}$ is iid, $y_i$ has values in $R^L$ and $x_i$ has values in $R^Q$.

(ii) The joint density $f(y,x)$ has a compact support $\Theta \subset R^{L+Q}$, $f$ is continuously differentiable up to the order $g = t + s$ for some even $s > 0$.

(iii) The kernel function $K(\cdot,\cdot)$ is continuously differentiable, $K$ vanishes outside a compact set, $\int K(y,x)dydx = 1$, and $K$ is a kernel of order $s$.

(iv) The function $r(y,x)$ is continuous and bounded a.e.

Then,

(I) if $t = 0$, $N\sigma_N^Q \to \infty$ and $\sigma_N^s\sqrt{N\sigma_N^Q} \to 0,$ then

$\sqrt{N\sigma_N^Q}\int r(y,x)\left(\widehat{f}(y,x) - f(y,x)\right)dy \;\to N(0,V_1)$

where $V_1 = \left\{\int [r(y,x)]^2 f(y,x)dy\right\}\left\{\int \left(\int K(y,x)dy\right)^2 dx\right\}$,

and for any two distinct points $x^{(1)}$ and $x^{(2)}$,

$\sqrt{N\sigma_N^Q}\int r(y,x^{(1)})\left(\widehat{f}(y,x^{(1)}) - f(y,x^{(1)})\right)dy$ and

$\sqrt{N\sigma_N^Q}\int r(y,x^{(1)})\left(\widehat{f}(y,x^{(1)}) - f(y,x^{(1)})\right)dy$ are asymptotically independent.

(II) if $t = 1$, $N\sigma_N^{Q+2} \to \infty$ and $\sigma_N^s\sqrt{N\sigma_N^{Q+2}} \to 0$ then

$\sqrt{N\sigma_N^{Q+2}}\int r(y,x)\left(\frac{\partial\widehat{f}(y,x)}{\partial x} - \frac{\partial f(y,x)}{\partial x}\right)dy \;\to N(0,V_2)$

where $V_2 = \left\{\int [r(y,x)]^2 f(y,x)dy\right\}\widetilde{K_2}$

and $\widetilde{K_2} = \left\{\int \left(\int\int \frac{\partial K(y,x)}{\partial x}dy\right)\left(\int\int \frac{\partial K(y,x)}{\partial x}dy\right)' dx\right\}$.

Moreover, for any two distinct points $x^{(1)}$ and $x^{(2)}$,

$\sqrt{N\sigma_N^{Q+2}}\int r(y,x^{(1)})\left(\frac{\partial\widehat{f}(y,x^{(1)})}{\partial x} - \frac{\partial f(y,x^{(1)})}{\partial x}\right)dy$ and

$\sqrt{N\sigma_N^{Q+2}} \int r(y,x^{(2)}) \left( \frac{\partial \widehat{f}(y,x^{(2)})}{\partial x} - \frac{\partial f(y,x^{(2)})}{\partial x} \right) dy$ are asymptotically independent.

PROOF: We first show (I). By the definition of $\widehat{f}$,

$$\int r(y,x)\,(\widehat{f}(y,x) - f(y,x))\,dy$$

$$= \int r(y,x) \left( \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{\sigma^{L+Q}} K(\tfrac{y_i-y}{\sigma}, \tfrac{x_i-x}{\sigma},) - f(y,x) \right] \right) dy$$

$$= \frac{1}{N} \sum_{i=1}^N \left[ \int r(y,x) \frac{1}{\sigma^{L+Q}} K(\tfrac{y_i-y}{\sigma}, \tfrac{x_i-x}{\sigma}) dy - \int r(y,x) f(y,x) dy \right]$$

Let

$$w_i = \frac{1}{\sigma^{L+Q}} \int r(y,x) K(\tfrac{y_i-y}{\sigma}, \tfrac{x_i-x}{\sigma}) dy.$$

Then,

$$\int r(y,x)\,(\widehat{f}(y,x) - f(y,x))\,dy$$

$$= \frac{1}{N} \sum_{i=1}^N \left[ w_i - \int r(y,x)\; f(y,x)\,dy \right]$$

$$= \frac{1}{N} \sum_{i=1}^N \left[ w_i - E(w_i) \right] + \left[ E(w_i) - \int r(y,x)\; f(y,x)\,dy \right]$$

We will show that under the above assumptions, the first term is asymptotically normal and the second converges to 0. For this, we note that

$$E(w_i) = E\left( \frac{1}{\sigma^{L+Q}} \int r(y,x) K(\tfrac{y_i-y}{\sigma}, \tfrac{x_i-x}{\sigma}) dy \right)$$

$$= \int \int \left( \frac{1}{\sigma^{L+Q}} \int r(y,x) K(\tfrac{y_i-y}{\sigma}, \tfrac{x_i-x}{\sigma}) dy \right) f(y_i,x_i) dy_i dx_i$$

$$= \int \int \left( \int r(y,x) K(\widetilde{y},\widetilde{x}) dy \right) f(y+\sigma\widetilde{y}, x+\sigma\widetilde{x}) d\widetilde{y} d\widetilde{x}$$

$$= \int r(y,x) \left( \int \int K(\widetilde{y},\widetilde{x}) f(y+\sigma\widetilde{y}, x+\sigma\widetilde{x}) d\widetilde{y} d\widetilde{x} \right) dy$$

Using a Taylor's expansion of $f(y+\sigma\widetilde{y}, x+\sigma\widetilde{x})$ around $f(y,x)$ and using the assumption that the kernel function $K$ integrates to 1 and is of order $s$, it follows that

(1) $E(w_i) = \int r(y,x) f(y,x) dy + O(\sigma^s)$.

Next,

$$E(w_i^2) = E\left( \frac{1}{\sigma^{2L+2Q}} \left[ \int r(y,x) K(\tfrac{y_i-y}{\sigma}, \tfrac{x_i-x}{\sigma}) dy \right]^2 \right)$$

$$= E\left( \frac{1}{\sigma^{2Q}} \left[ \int r(y_i - \sigma\widetilde{y}, x) K(\widetilde{y}, \tfrac{x_i-x}{\sigma}) d\widetilde{y} \right]^2 \right)$$

$= \int\int \frac{1}{\sigma^Q} \left[\int r(y_i - \sigma\widetilde{y}, x)K(\widetilde{y}, \widetilde{x})d\widetilde{y}\right]^2 f(y_i, x + \sigma\widetilde{x})dy_i d\widetilde{x}$

Then, by the continuity and boundedness of $f$ and $r$, it follows by the Bounded Convergence Theorem that

$$(2)\ \sigma^Q E(w_i^2) \to \left[\int r(y,x)^2 f(y,x)dy\right] \left\{\int \left(\int K(y,x)dy\right)^2 dx\right\}.$$

From (1), (2), and $\sigma \to 0$ it follows that

$$\sigma^Q Var(w_i) \to \left[\int r(y,x)^2 f(y,x)dy\right] \left\{\int \left(\int K(y,x)dy\right)^2 dx\right\}$$

Let $\delta > 0.$ Since

$$E\left|\frac{1}{N}(w_i - E(w_i))\right|^{2+\delta} \leq \frac{2^{2+\delta}E|w_i|^{2+\delta}}{N^{2+\delta}}$$

$$E\left|w_i\right|^{2+\delta} = E\left|\frac{1}{\sigma^{L+Q}}\int r(y,x)K(\frac{y_i-y}{\sigma}, \frac{x_i-x}{\sigma})dy\right|^{2+\delta}$$

$$= \int\int \left(\frac{1}{\sigma^{Q(2+\delta)}}\left|\int r(y_i - \sigma\widetilde{y}, x)K(\widetilde{y}, \frac{x_i-x}{\sigma})d\widetilde{y}\right|^{2+\delta}\right) f(y_i, x_i)dy_i dx_i$$

$$= \int\int \left(\frac{1}{\sigma^{Q(1+\delta)}}\left|\int r(y_i - \sigma\widetilde{y}, x_i - \sigma\widetilde{x})K(\widetilde{y}, \widetilde{x})d\widetilde{y}\right|^{2+\delta}\right) f(y_i, x + \sigma\widetilde{x})d\widetilde{y}d\widetilde{x}$$

$$= O\left(\frac{1}{\sigma^{Q(1+\delta)}}\right)$$

and

$$\left(\frac{Var(w_i)}{N}\right)^{\frac{2+\delta}{2}} = O\left(\frac{1}{(N\sigma^Q)^{1+\frac{\delta}{2}}}\right),$$

it follows that

$$\frac{\sum_{i=1}^N E\left|(\frac{w_i}{N} - E(\frac{w_i}{N})\right|^{2+\delta}}{\left[\left(Var\sum_{i=1}^N \frac{w_i}{N}\right)^{1/2}\right]^{2+\delta}} = \frac{\sum_{i=1}^N E\left|(\frac{w_i}{N} - E(\frac{w_i}{N})\right|^{2+\delta}}{\left(Var(\frac{w_i}{N})\right)^{1+\frac{\delta}{2}}} = O\left(\frac{1}{N^{1+(\delta/2)}\sigma^{Q\delta/2}}\right) \to 0.$$

By Liapounov's Theorem it then follows that

$$\sqrt{N\sigma_N^Q}\left(\frac{1}{N}\sum_{i=1}^N w_i - E(w_i)\right) \to N(0, V_1)$$

where

$$V_1 = \left[\int r(y,x)^2 f(y,x)dy\right] \left\{\int \left(\int K(y,x)dy\right)^2 dx\right\}.$$

Since by (1) and by assumption,

$$\sqrt{N\sigma_N^Q}\left(E(w_i) - \int r(y,x)f(y,x)dy\right) = O\left(\sigma_N^s\sqrt{N\sigma_N^Q}\right) \to 0$$

the first part of (I) is proved.

Next, to show the asymptotic independence, we note that by (i) and the definition of $\widehat{f}$ the covariance equals

$$\frac{\sigma^Q}{\sigma^{2(L+Q)}}\left\{E\left[\left(\int r^1K^1\right)\left(\int r^2K^2\right)\right]-E\left(\int r^1K^1\right)E\left(\int r^2K^2\right)\right\}$$

where

$$\left(\int r^kK^k\right)=\int r(y,x^{(k)})K\left(\frac{y_i-y}{\sigma},\frac{x_i-x^{(k)}}{\sigma}\right)dy \quad k=1,2$$

Since

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

$$=\sigma^{2L+Q}\int\left(\int \widetilde{r}^1\widetilde{K}^1\right)\left(\int \widetilde{r}^1\widetilde{K}^1\right)f(y_i,x^{(1)}+\sigma\widetilde{x})dy_id\widetilde{x}$$

where $\int \widetilde{r}^1\widetilde{K}^1=\int r(y_i-\sigma\widetilde{y},x^{(1)})K(\widetilde{y},\widetilde{x})d\widetilde{y}$

and $\int \widetilde{r}^2\widetilde{K}^2=\int r(y_i-\sigma\widetilde{y},x^{(2)})K(\widetilde{y},\widetilde{x}+\frac{x^{(1)}-x^{(2)}}{\sigma})d\widetilde{y}$

it follows by bounded convergence, (1), and $\sigma\to 0$ that the covariance converges to 0.

We next show (II). By the definition of $\widehat{f}_x$,

$$\int\int r(y,x)\left(\widehat{f}_x(y,x)-f_x(y,x)\right)dy$$

$$=\int\int r(y,x)\left(\frac{1}{N}\sum_{i=1}^N\left[\frac{(-1)}{\sigma^{L+Q+1}}\frac{\partial K(\frac{y_i-y}{\sigma},\frac{x_i-x}{\sigma})}{\partial x}-\frac{\partial f(y,x)}{\partial x}\right]\right)dy$$

$$=\frac{1}{N}\sum_{i=1}^N\left[\int\int r(y,x)\ \frac{(-1)}{\sigma^{L+Q+1}}\frac{\partial K(\frac{y_i-y}{\sigma},\frac{x_i-x}{\sigma})}{\partial x}dy-\int\int r(y,x)\ \frac{\partial f(y,x)}{\partial x}dy\right]$$

Let $w_i=\frac{(-1)}{\sigma^{L+Q+1}}\int\int r(y,x)\frac{\partial K(\frac{y_i-y}{\sigma},\frac{x_i-x}{\sigma})}{\partial x}dy$. Then,

$$\int r(y,x)\left(\widehat{f}_x(y,x,z)-f_x(y,x,z)\right)dy$$

$$=\frac{1}{N}\sum_{i=1}^N\left[w_i-E(w_i)\right]+\left[E(w_i)-\int r(y,x)\ \frac{\partial f(y,x)}{\partial x}dy\right]$$

We note that

$$E(w_i)=E\left(\frac{(-1)}{\sigma^{L+Q+1}}\int r(y,x)\frac{\partial K(\frac{y_i-y}{\sigma},\frac{x_i-x}{\sigma})}{\partial x}dy\right)$$

$$=\int\int\left(\frac{(-1)}{\sigma^{L+Q+1}}\int r(y,x)\frac{\partial K(\frac{y_i-y}{\sigma},\frac{x_i-x}{\sigma})}{\partial x}dy\right)f(y_i,x_i)dy_idx_i$$

$$=\int\int\left(\frac{(-1)}{\sigma}\int r(y,x)\frac{\partial K(\widetilde{y},\widetilde{x})}{\partial x}dy\right)f(y+\sigma\widetilde{y},x+\sigma\widetilde{x})d\widetilde{y}d\widetilde{x}$$

$= \int r(y,x) \left( \int \int K(\widetilde{y},\widetilde{x}) \frac{\partial f(y+\sigma\widetilde{y}, x+\sigma\widetilde{x})}{\partial x} d\widetilde{y} d\widetilde{x} \right) dy$

where the last inequality follows by integration by parts. Using a Taylor's expansion of $\frac{\partial f(y+\sigma\widetilde{y}, x+\sigma\widetilde{x})}{\partial x}$ around $\frac{\partial f(y,x)}{\partial x}$ and using the assumption that the kernel function $K$ integrates to 1 and is of order $s$, it follows that

(3) $E(w_i) = \int \int r(y,x) \frac{\partial f(y,x)}{\partial x} dy + O(\sigma^s)$.

Next,

$E(w_i w_i')$

$= E\left( \frac{1}{\sigma^{2(L+Q+1)}} \left[ \int r(y,x) \frac{\partial K(\frac{y_i-y}{\sigma}, \frac{x_i-x}{\sigma})}{\partial x} dy \right] \left[ \int r(y,x) \frac{\partial K(\frac{y_i-y}{\sigma}, \frac{x_i-x}{\sigma})}{\partial x'} dy \right] \right)$

$= E\left( \frac{1}{\sigma^{2Q+2}} r_i r_i' \right)$

$= \int \int \left( \frac{1}{\sigma^{Q+2}} \overline{r}_i \overline{r}_i' \right) f(y_i, x+\sigma\widetilde{x}) dy_i d\widetilde{x}$

where $r_i = \int r(y_i - \sigma\widetilde{y}, x) \frac{\partial K(\widetilde{y}, \frac{x_i-x}{\sigma})}{\partial x} d\widetilde{y}$

and $\overline{r}_i = \int r(y_i - \sigma\widetilde{y}, x) \frac{\partial K(\widetilde{y}, \frac{x_i-x}{\sigma})}{\partial x} d\widetilde{y}$.

By the continuity and boundedness of $f$ and $r$, it follows by the Bounded Convergence Theorem that

(4) $\sigma^{Q+2} E(w_i w_{i'}) \to \left[ \int r(y,x)^2 f(y,x) dy \right] \widetilde{K}$

where $\widetilde{K} = \left\{ \int \left( \int \frac{\partial K(y,x)}{\partial x} dy \right) \left( \int \frac{\partial K(y,x)}{\partial x} dy \right)' dx \right\}$.

From (3), (4), and $\sigma \to 0$ it then follows that

$\sigma^{Q+2} Var(w_i) \to \left[ \int r(y,x)^2 f(y,x) dy \right] \widetilde{K}$

To apply Liapounov's Central Limit Theorem, we note that, for $\delta > 0$,

$E \left| \frac{1}{N}(w_i - E(w_i)) \right|^{2+\delta} \leq \frac{2^{2+\delta} E|w_i|^{2+\delta}}{N^{2+\delta}}$

where

$E\left|w_i\right|^{2+\delta} = E \left| \frac{(-1)}{\sigma^{L+Q+1}} \int r(y,x) \frac{\partial K(\frac{y_i-y}{\sigma}, \frac{x_i-x}{\sigma})}{\partial x} dy \right|^{2+\delta}$

$= \int \int \left( \frac{(-1)^{2+\delta}}{\sigma^{(Q+1)(2+\delta)}} \left| \int r(y_i - \sigma\widetilde{y}, x) \frac{\partial K(\widetilde{y}, \frac{x_i-x}{\sigma})}{\partial x} d\widetilde{y} \right|^{2+\delta} \right) f(y_i, x_i) dy_i dx_i$

$$= \int\int \left( \frac{(-1)^{2+\delta}}{\sigma^{Q(1+\delta)+2+\delta}} \left| \int r(y_i - \sigma\widetilde{y}, x) \frac{\partial K(\widetilde{y}, \widetilde{x})}{\partial x} d\widetilde{y} \right|^{2+\delta} \right) f(y_i, x + \sigma\widetilde{x}) dy_i y d\widetilde{x}$$

$$= O\left( \frac{1}{\sigma^{Q(1+\delta)+2+\delta}} \right)$$

and

$$\left( \frac{Var(w_i)}{N} \right)^{\frac{2+\delta}{2}} = O\left( \frac{1}{(N\sigma^{Q+2})^{1+\frac{\delta}{2}}} \right).$$

Hence,

$$\frac{\sum_{i=1}^N E\left| (\frac{w_i}{N} - E(\frac{w_i}{N}) \right|^{2+\delta}}{\left[ \left( Var \sum_{i=1}^N \frac{w_i}{N} \right)^{1/2} \right]^{2+\delta}} = \frac{\sum_{i=1}^N E\left| (\frac{w_i}{N} - E(\frac{w_i}{N}) \right|^{2+\delta}}{\left( Var(\frac{w_i}{N}) \right)^{1+\frac{\delta}{2}}} = O\left( \frac{1}{N^{1+\delta/2}\sigma^{Q\delta/2}} \right) \to 0.$$

By Liapounov's Theorem it then follows that

$$\sqrt{N\sigma_N^{Q+2}} \left( \frac{1}{N} \sum_{i=1}^N \left[ w_i - E(w_i) \right] \right) \to N(0, V_2)$$

where $V_2 = \left[ \int r(y,x)^2 f(y,x) dy \right] \widetilde{K}$

and $\widetilde{K} = \left\{ \int \left( \int \frac{\partial K(y,x)}{\partial x} dy \right) \left( \int \frac{\partial K(y,x)}{\partial x} dy \right)' dx \right\}.$

Since, by (3), $\sqrt{N\sigma_N^{Q+2}} \left( E(w_i) - \int\int r(y,x) \frac{\partial f(y,x)}{\partial x}) \, dy \right) = O(\sigma_N^s \sqrt{N\sigma_N^{Q+2}}) \to 0$. The first part of result (II) is proved. To prove the second part of (II), we note that by (i) and the definition of $\widehat{\frac{\partial f}{\partial x}}$, the covariance equals

$$\frac{\sigma^{Q+2}}{\sigma^{2(L+Q+1)}} \left\{ E\left[ \left( \int r^1 \partial K^1 \right) \left( \int r^2 \partial K^2 \right) \right] - E\left( \int r^1 \partial K^1 \right) E\left( \int r^2 \partial K^2 \right) \right\}$$

where

$$\left( \int r^k \partial K^k \right) = \int r(y, x^{(k)}) \frac{\partial K\left( \frac{y_i - y}{\sigma}, \frac{x_i - x^{(k)}}{\sigma} \right)}{\partial x} dy \quad k = 1, 2$$

Since

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

$$= \sigma^{2L+Q+2} \int \left( \int \widetilde{r}^1 \partial \widetilde{K}^1 \right) \left( \int \widetilde{r}^1 \partial \widetilde{K}^1 \right) f(y_i, x^{(1)} + \sigma\widetilde{x}) dy_i d\widetilde{x}$$

where $\int \widetilde{r}^1 \widetilde{K}^1 = \int r(y_i - \sigma\widetilde{y}, x^{(1)}) \frac{\partial K(\widetilde{y}, \widetilde{x})}{\partial x} d\widetilde{y}$

and $\int \widetilde{r}^2 \widetilde{K}^2 = \int r(y_i - \sigma\widetilde{y}, x^{(2)}) \frac{\partial K(\widetilde{y}, \widetilde{x} + \frac{x^{(1)} - x^{(2)}}{\sigma})}{\partial x} d\widetilde{y}$

it follows by bounded convergence, (3), and $\sigma \to 0$ that the covariance converges to 0.

# 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 Aproach," 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 T.F. Bewley (ed.) Advances of Econometrics, Cambridge University Press, Vol. I, 99-144.

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

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).

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. 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, by C. Hsiao, K. Morimune, and J. Powell 9editors), 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.

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.

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

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

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

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.

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," Ann. Math. Statist. 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.