ESTIMATION OF NONPARAMETRIC MODELS WITH SIMULTANEITY

Rosa L. Matzkin*
Department of Economics
University of California, Los Angeles

May 2010

Abstract

We introduce methods for estimating nonparametric, nonadditive models with simultaneity. The methods are developed by directly connecting the elements of the structural system to be estimated with features of the density of the observable variables, such as the values of derivatives and of average derivatives of this density. The estimators are easily computed functionals of nonparametric estimators of these features. We consider in detail a model where to each structural equation there corresponds an exclusive regressor and a model with one equation of interest and one instrument. For the first model, our estimator for the matrix of derivatives of the structural function has a form analogous to the one of a standard Least Squares estimator, $(X'X)^{-1}(X'Y)$, except that the elements of the matrices X and Y are constructed from average derivative estimators of the conditional density of the observed endogenous variables given the observed exogenous variables. For the second model, with one equation of interest and one instrument, we provide several identification and estimation results. The estimators that we develop are based on using the estimated density of the observable variables to find particular values of the instrument where one can read off the derivative of the function of interest. We show that our estimators are consistent and asymptotically normal. We also indicate several ways in which our new identification results can be used to develop new estimation methods.

——————————————

*The support of NSF is gratefully acknowledged. I have benefitted from discussions with Richard Blundell and Whitney Newey, comments by participants at the 2008 All UC Econometrics Conference at Berkeley, 2009 Winter Meeting of the Econometric Society, 2009 BIRS Conference on Semiparametric and Nonparametric Methods in Econometrics, Conference on Consumer Demand at Brown University, SETA Conference at Kyoto University, Invited Session on Advances in Econometrics at the Latin American Meeting of the Econometric Society, 2009 Minnesota Lecture, econometric seminars at Harvard/MIT, UCL, UCLA, Stanford, and Maryland, and the excellent research assistance of Yong Hyeon Yang.

1. Introduction

Estimation of structural models has been one of the main objectives of econometrics since its early times. The analysis of counterfactuals, the evaluation of welfare, and the prediction of the evolution of markets, among others, require knowledge of primitive functions and distributions in the economy, such as technologies and distributions of preferences, which often can only be estimated using structural models.

Estimation of parametric structural models dates back to the early works of Haavelmo (1943, 1944), Hurwicz (1950), Koopmans (1949), Koopmans and Reiersol (1950), Koopmans, Rubin, and Leipnik (1950), (1950), Fisher (1959, 1961, 1966), Wegge (1965), Rothenberg (1971), and Bowden (1973). (See Hausman (1983) and Hsiao (1983) for early review articles.)

In more recent years, estimation of semiparametric and nonparametric models has received increasing interest and significant development. Several estimators have been developed based on conditional moment restrictions. These include Newey and Powell (1989, 2003), Darolles, Florens, and Renault (2002), Ai and Chen (2003), Hall and Horowitz (2003), and specifically for models with nonadditive random terms, Chernozhukov and Hansen (2005) and Chernozhukov, Imbens, and Newey (2007). Identification in these models has been mostly analyzed in terms of restrictions leading to completeness type conditions. Estimation has often required dealing with ill-posed inverse problems.

In this paper, we make assumptions and construct nonparametric estimators in ways that are different from those nonparametric methods for models with simultaneity. In particular, our estimators are closely tied to conditions on the structural model, which allow us to directly read off the density of the observable variables the particular elements of the structural model that we are interested in estimating. In this vein, estimators for conditional expectations can be easily constructed by integrating nonparametric estimators for conditional probability densities, such as the kernel estimators of Nadaraya (1964) and Watson (1964). Conditional quantiles estimators can be easily constructed by inverting nonparametric estimators for conditional distribution functions, such as in Bhattacharya (1963) and Stone (1977).[1] For structural functions with nonadditive unobservable random terms, several methods exist to estimate the nonparametric function directly from estimators for the distribution of the observable variables. These include Matzkin (1999, 2003), Altonji and Matzkin (2001, 2005), Chesher (2003), and Imbens and Newey (2003, 2009). These methods cannot handle simultaneous equation models that do not satisfy control function separability (Blundell and Matzkin (2010)). The goal of this paper is to fill this important gap.

[1] See Koenker (2005) for other quantile methods.

Our simultaneous equations models are nonparametric and with nonadditive unobservable random terms. Unlike linear models with additive errors, each reduced form function in the nonadditive model depends separately on the value of each of the unobservable variables in the system. We show that our estimators are consistent and asymptotically normal.

Alternative estimators for nonparametric simultaneous equations can be formulated using a nonparametric version of Manski (1983) Minimum Distance from Independence, as in Brown and Matzkin (1998). Those estimators are defined as the minimizers of a distance between the joint and the multiplication of the marginal distributions of the exogenous variables, and typically do not have a closed form. To our knowledge, no asymptotic distribution is known for estimators of nonparametric functions defined in this way.

To describe the estimation approach, we focus on two particular models. Our first model is a system where to each equation there corresponds an exclusive regressor. Consider for example a model where the vector of observable endogenous variables consists of the Nash equilibrium actions of a set of players. Each player chooses his or her action as a function of his or her individual observable and unobservable costs, taking the other players' actions as given. In this model, each individual player's observable cost would be the exclusive observable variable corresponding to the reaction function of that player. Our method allows to estimate nonparametrically the reaction functions of each of the players, from the distribution of observable equilibrium actions and players' costs. The calculation of our estimator for the reaction functions requires only a simple matrix inversion and a multiplication, analogous to the solution of Linear Least Squares estimators. The difference is that the elements in our matrices are calculated using nonparametric average derivative methods. In this sense, our estimators can be seen as the extension to models with simultaneity of the average derivative methods of Stoker (1986) and Powell, Stock and Stoker (1989). As in those papers, we extract the structural parameters using averages of nonparametrically estimated derivatives of the densities of the observable variables.[2]

The second model for which we develop estimators is one such as a demand function, where the object of interest is the derivative of the demand with respect to price. Price is determined by another function, the supply function, which depends on quantity produced, an unobservable shock, and at least one observable cost. We provide conditions under which the derivative of the demand function with respect to price can be easily read off the joint density of the equilibrium price, the equilibrium quantity and the observable cost, and we develop consistent and asymptotically normal estimators for this derivative. We also present new results that can be used to develop new identification results and related estimators.

[2]Existent extensions of the average derivative methods of Stoker (1986) and Powell, Stock, and Stoker (1989) for models with endogeneity, such as Altonji and Ichimura (2000), Altonji and Matzkin (2001, 2005), Blundell and Powell (2003a), and Imbens and Newey (2003, 2009), require conditions that are generally not satisfied by models with simultaneity.

We focus in this paper on the most simple models we can deal with, which exhibit simultaneity. However, our proposed techniques can be used in models where simultaneity is only one of many other possible features of the model. For example, our results can be used in models with simultaneity in latent dependent variables, models with unobserved heterogeneity, and models where the unobservable variables are only conditionally independent of the explanatory variables. (See Matzkin (2010a) for identification of some such models and Matzkin (2010b) for applications of the estimation methods in this paper to such models.)

The structure of the paper is as follows. In the next section we present a basic model and discuss some of its features. In Section 3 we present an estimator for a simultaneous model with exclusive regressors. Section 4 deals with the model of one equation of interest and one excluded instrument. It presents identification and estimation results. Section 5 presents results for this latter model that can be used to develop additional identification results and estimators. The Appendix contains some of the proofs

2. Nonadditive simultaneous equations

Our basic model can be described as

$$(2.1) \qquad s(Y, X, \varepsilon) = 0$$

where Y denotes a G-dimensional vector of observable endogenous variables, X denotes a K-dimensional vector of observable exogenous variables, ε denotes a G-dimensional vector of unobserved variables, and $s : R^{G+K+G} \to R^G$ is an unknown function.

We assume that the function s is such that for any value (x, ε) of (X, ε), there exists a unique value y of Y such that $s(y, x, \varepsilon) = 0$ and for any value (y, x) there exists a unique value ε such that $s(y, x, \varepsilon) = 0$. We will denote the function that assigns the values of y that satisfies (2.1) for (x, ε) by $h(x, \varepsilon)$ and we will denote the function that assigns the value of ε satisfying (2.1) for (y, x) by $r(y, x)$. The function h corresponds to the reduced form model of (2.1) while the function r corresponds to the structural form model of (2.1).

We will assume that the functions h and r are each twice continuously differentiable and that, for each fixed value x of X, both functions are onto R^G. The vector of unobservable variables ε will be assumed to be distributed independently of X and to possess an everywhere positive and continuously differentiable density ε. These assumptions were also made in Matzkin (2008) to analyze identification of nonparametric simultaneous equation models.

To describe the complications that arise in model (2.1) due to the nonlinearity of the structural system of equations, consider the textbook example of a system of demand and supply,

$$
(2.2) \quad \begin{aligned} Q &= D\left(P, I, \varepsilon_D\right) \\ P &= S\left(Q, W, \varepsilon_S\right) \end{aligned}
$$

where Q is observed quantity, P is observed price, I is observed income of the consumers, W is observed production costs, ε_D and ε_S are the unobservable random terms in, respectively, the demand and supply functions, and D and S are the unknown demand and supply functions.

Suppose that the functions D and S were linear in the endogenous variables and additive in, respectively, ε_D and ε_S,

$$
\begin{aligned} Q &= \beta_0 + \beta_1 P + \beta_2 I + \varepsilon_D \\ P &= \alpha_0 + \alpha_1 Q + \alpha_2 W + \varepsilon_S \end{aligned}
$$

Under conditions on the parameters guaranteeing the existence of unique solutions for (P, Q), this system generates reduced form functions for P and Q of the form

$$
\begin{aligned} P &= \gamma_0 + \gamma_1 I + \gamma_2 W + v_P \\ Q &= \delta_0 + \delta_1 I + \delta_2 W + v_Q \end{aligned}
$$

each with one additive unobservable variable. The linearity and additivity of the structural equations guarantees that the effect of the two unobservable variables, ε_D and ε_S, in each reduced form equation collapses into one unobservable additive variable for each equation. This allows one to estimate the reduced form equations by standard linear least squares methods. Estimation of the structural parameters can then be obtained, for example, from that of the reduced form parameters.

When the structural model is nonlinear in the endogenous variables or nonadditive in the unobservable variables, the effect of the unobservable variables ε_D and ε_S will not in general collapse into one unobserved variable for each reduced form equation. The reduced form function will then depend separately on both unobservable variables, ε_D and ε_S. In other words, we will only be able to establish that for some nonparametric functions h_D and h_S,

$$
\begin{aligned} Q &= h_D\left(I, W, \varepsilon_D, \varepsilon_S\right) \\ P &= h_S\left(I, W, \varepsilon_D, \varepsilon_S\right) \end{aligned}
$$

Nonparametric identification of the reduced form functions, h_D and h_S, the structural functions D and S, and the distribution of $(\varepsilon_D, \varepsilon_S)$, or of particular features of them, was studied in Matzkin (2008), following results by Brown (1983), Roehrig (1988), and Benkard and Berry

(2006). In the next section, we apply those results to show identification of the derivatives of the function r and we develop estimators that are calculated by an expression analogous to that of the standard Least Squares form, $(X'X)^{-1} X'Y$, except that the elements of the matrices are formed from estimated average derivative estimators. In Section 4, we consider the estimation of the function D in the model

$$\begin{aligned} Q &= D(P, \varepsilon_D) \\ P &= S(Q, W, \varepsilon_S) \end{aligned} \tag{2.3}$$

where I is not an argument of D. We provide conditions under which the derivative of D with respect to price can be read off directly from the density of (P, Q, W), and it can be estimated by an easily computable, consistent and asymptotically normal estimator.

3. A model with exclusive regressors

We consider the model

$$\begin{aligned} Y_1 &= m^1(Y_2, Y_3, ..., Y_G, Z, X_1, \varepsilon_1) \\ Y_2 &= m^2(Y_1, Y_3, ..., Y_G, Z, X_2, \varepsilon_2) \\ &\cdots \\ Y_G &= m^G(Y_1, Y_2, ..., Y_{G-1}, Z, X_G, \varepsilon_G) \end{aligned} \tag{3.1}$$

where $(Y_1, ..., Y_G)$ is a vector of observable endogenous variables, $(Z, X_1, ..., X_G)$ is a vector of observable exogenous variables, and $(\varepsilon_1, ..., \varepsilon_G)$ is a vector of unobservable variables. Inclusion in each of the functions m^g of the observable vector Z does not create any major complications. For simplicity of exposition, we will not include such a vector Z. If Z were included, all our assumptions and conclusions would hold conditionally on Z. We will assume that the functions $m^1, ..., m^G$ are invertible in, respectively, $\varepsilon_1, ..., \varepsilon_G$. This can be guaranteed if for each g, m^g is strictly increasing in ε_g. We will denote the corresponding inverse functions by $r^1, ..., r^G$. Hence, our system of indirect structural equations, denoting the mapping from the vectors of observable variables to the vector of unobservable variables, is expressed as

$$\begin{aligned} \varepsilon_1 &= r^1(Y_1, ..., Y_G, X_1) \\ \varepsilon_2 &= r^2(Y_1, ..., Y_G, X_2) \\ &\cdots \\ \varepsilon_G &= r^G(Y_1, ..., Y_G, X_G) \end{aligned}$$

In addition to the existence of this system of indirect structural equations, we will also assume that there exists a reduced form system. That is, we assume that for any values of $(X_1, ..., X_G, \varepsilon_1, ..., \varepsilon_G)$, the system in (3.1) has a unique solution. We will let the values of functions $h^1, ..., h^G$ denote the solution to these equations,

$$\begin{aligned} Y_1 &= h^1 (X_1, ..., X_G, \varepsilon_1, ..., \varepsilon_G) \\ Y_2 &= h^2 (X_1, ..., X_G, \varepsilon_1, ..., \varepsilon_G) \\ &\cdots \\ Y_G &= h^G (X_1, ..., X_G, \varepsilon_1, ..., \varepsilon_G) \end{aligned}$$

Our first assumption specifies the sense in which the explanatory observable regressors, $X_1, ..., X_G$, are exclusive. It states that, for any g, the $g - th$ structural function does not depend on X_k for all $k \neq g$.

Assumption 3.1: *For any g and any values y_{-g} of $(Y_1, ..., Y_{g-1}, Y_{g+1}, ..., Y_G)$, ε_g, x_g of X_g, and x_{-g} of $(X_1, ..., X_{g-1}, X_{g+1}, ..., X_G)$, the value of the function m^g at $(y_{-g}, x_g, \varepsilon_g)$ is constant over x_{-g}.*

Our second assumption can be interpreted as specifying units of measurement for each ε_g. Since ε_g is unobservable, its value can be determined only up to a monotone transformation. We tie the values of ε_g to that of x_g by requiring that the derivative of m^g with respect to ε_g is equal in absolute value to the derivative of m^g with respect to x_g. The assumption also imposes a strict monotonicity of m^g on ε_g, by bounding from below the derivative of m^g with respect to ε_g. Alternative conditions can be derived using results in Matzkin (2008).

Assumption 3.2: *For all g*

$$- \frac{\partial m^g (y_{-g}, x_g, \varepsilon_g)}{\partial x_g} = \frac{\partial m^g (y_{-g}, x_g, \varepsilon_g)}{\partial \varepsilon_g} > 0$$

The following set of assumptions imposes conditions guaranteeing that the mapping between the structural elements r and f_ε and the conditional densities of the observable variables, $f_{Y|X=x}$ is given by the transformation of variables equation

$$f_{Y|X=x}(y) = f_\varepsilon (r (y, x)) \left| \frac{\partial r (y, x)}{\partial y} \right| .$$

These assumptions also guarantee that both sides of these equation can be differentiated with respect to y and x. Some weakening of these assumptions is possible. One such direction is highlighted by Assumptions 3.4' and 3.4".

Assumption 3.3: $(\varepsilon_1, ..., \varepsilon_G)$ *is distributed independently of* $(X_1, ..., X_G)$ *with an everywhere positive and continuously differentiable density,* f_ε.

Assumption 3.4: $(X_1, ..., X_G)$ *possesses a differentiable density whose support is such that for any* $t \in R^G$ *and any* $y \in R^G$, *there exists* x *in this support such that* $t = r(y, x)$.

Assumption 3.5: *Conditional on* $(X_1, ..., X_G)$, *the functions* $r = \left(r^1, ..., r^G\right)$ *and* $h = \left(h^1, ..., h^G\right)$ *are twice continuously differentiable,* $1-1$, *onto* R^G, *and their Jacobian determinants are strictly positive.*

The next assumption imposes a restriction at $G+1$, not necessarily known, points on the support of ε. It states that at one point, $\varepsilon^{*(0)}$, the derivative of f_ε with respect to all its coordinates is zero, and at G points $\varepsilon^{*(1)}, ..., \varepsilon^{*(G)}$, the derivative of f_ε with respect to all its coordinates except one equals zero. For each g, the coordinate for which the derivative of f_ε is different from zero at $\varepsilon^{*(g)}$ is the $g-$th coordinate. For some of our results, this assumption can be replaced by alternative invertibility conditions on the matrix of values of $\partial \log f_\varepsilon \left(\varepsilon^{*(g)}\right) / \partial \varepsilon_k$ $(g = 1, ..., G; k = 1, ..., G)$.

Assumption 3.6: *The density,* f_ε, *of* $(\varepsilon_1, ..., \varepsilon_G)$ *is such that*
(i) for some, not necessarily known, value $\varepsilon^{*(0)} = (\varepsilon_1^*, ..., \varepsilon_G^*)$,

$$\partial \log f_\varepsilon (\varepsilon_1^*, ..., \varepsilon_G^*) / \partial \varepsilon_g = 0 \qquad for\ g = 1, ..., G,$$

(ii) for each g, *there exists a, not necessarily known, value* $\varepsilon^{*(g)}$ *such that*

$$\partial \log f_\varepsilon \left(\varepsilon^{*(g)}\right) / \partial \varepsilon_g \neq 0 \text{ and } for\ all\ k \neq g, \quad \partial \log f_\varepsilon \left(\varepsilon^{*(g)}\right) / \partial \varepsilon_k = 0,$$

Assumption 3.6 together with 3.4 implies that, given any value y, we will be able to find values $x^{*(0)}, x^{*(1)}, ..., x^{*(G)}$of X such that the value of $r(y, x)$ equals the values of $\varepsilon^{*(0)}, \varepsilon^{*(1)}, ..., \varepsilon^{*(G)}$. Identification requires only these values of X to recover the derivative of r with respect to y. Hence, when we are interested in identifying the derivative of r with respect to y, at only one particular value y, Assumption 3.4 is too strong. It suffices to guarantee that the values $x^{*(0)}, x^{*(1)}, ..., x^{*(G)}$of X are in the support of X. Moreover, if the interest is only on the identification of the derivatives of only one coordinate function, say r^1, of r we need to guarantee that only $x^{*(0)}$ and $x^{*(1)}$ are in the support of X. We state these weaker assumptions in Assumptions 3.4' and 3.4". For both assumptions, y^* denotes a given specific value of Y.

Assumption 3.4': *X possesses a twice continuously differentiable density with $G+1$ points $x^{*(0)}, x^{*(1)}, ..., x^{*(G)}$ in the interior of the support of X, such that $r\left(y^*, x^{*(0)}\right) = \varepsilon^{*(0)}$ and $r\left(y^*, x^{*(g)}\right) = \varepsilon^{*(g)}$ for $g = 1, ..., G$, where $\varepsilon^{*(0)}$ and $\varepsilon^{*(1)}, ..., \varepsilon^{*(G)}$ are as in Assumption 3.6.*

Assumption 3.4": *X possesses a twice continuously differentiable density with 2 points, $x^{*(0)}$ and $x^{*(1)}$, in the interior of the support of X, such that $r\left(y^*, x^{*(0)}\right) = \varepsilon^{*(0)}$ and $r\left(y^*, x^{*(1)}\right) = \varepsilon^{*(1)}$ where $\varepsilon^{*(0)}$ and $\varepsilon^{*(1)}$ are as in Assumption 3.6.*

3.1. Identification

Denote the matrices of derivatives of r with respect to y and x by r_y and r_x. The following result follows from Section 4.2 in Matzkin (2008), once we prove that Assumption 3.2 implies that for all (y, x), $r_x(y, x) = I$.[3]

Theorem 3.1: *Suppose that the model satisfies Assumptions 3.1-3.6. Then,* r_y and r_x *are identified. If, in addition, it is specified that at some values $\overline{y}$ of Y and $\overline{x}$ of X, and for some $\gamma \in R^G$, $r(\overline{y}, \overline{x}) = \gamma$, then the function r and the density f_ε are identified. Let y^* denote a specific value of Y. If Assumption 3.4 is substituted with Assumption 3.4', then $r_y(y^*, x)$ is still identified. If Assumption 3.4 is substituted with Assumption 3.4", then $r_y^1(y^*, x)$ is still identified.*

Proof of Theorem 3.1: Let $\widetilde{r}$ denote a function satisfying the same assumptions that r is assumed to satisfy. Let Δ_y denote the $G \times 1$ vector whose $j-th$ element is $\partial \log |r_y| / \partial y_j - \partial \log |\widetilde{r}_y| / \partial y_j$ and let Δ_x denote the $G \times 1$ vector whose $j-th$ element is $\partial \log |r_y| / \partial x_j - \partial \log |\widetilde{r}_y| / \partial x_j$, both evaluated at (y, x). We first note that our assumptions imply that for all (y, x)

$$(T.3.1) \qquad \widetilde{r}_x = r_x = I \ , \quad \Delta_x = 0, \quad \text{and} \quad \frac{d\Delta_y}{dx} = 0$$

To see this, note that by the definition of r^g,

[3] See also Matzkin (2005), Matzkin (2007b, Example 4), Berry and Haile (2009), and Chiappori and Komunjer (2009) for identification results in similar models.

$$y_g = m^g\left(y_{-g}, x_g, r^g\left(y, x_g\right)\right).$$

Differentiating with respect to x_g gives

$$0 = \frac{\partial m^g}{\partial x_g} + \frac{\partial m^g}{\partial \varepsilon_g}\frac{\partial r^g}{\partial x_g}$$

Assumption 3.2 then implies that $\partial r^g/\partial x_g = 1$. Similarly, $\partial \widetilde{r}^g/\partial x_g = 1$. By Assumption 3.1, $r^g_{x_j} = \widetilde{r}^g_{x_j} = 0$ for $j \neq g$. Hence,

$$\widetilde{r}_x = r_x = I$$

Since $r^g_{x_g} = \widetilde{r}^g_{x_g} = 1$ and $r^g_{x_j} = \widetilde{r}^g_{x_j} = 0$ holds for all (y, x), it follows that for all g, k, j,

$$r^g_{y_k, x_j} = \widetilde{r}^g_{y_k, x_j} = 0.$$

Hence, for all g, j

$$\frac{\partial \log\left[|r_y|\right]}{\partial x_g} = \frac{\partial \log\left[|r_y|\right]}{\partial x_j} = \frac{\partial \log\left[|\widetilde{r}_y|\right]}{\partial x_g} = \frac{\partial \log\left[|\widetilde{r}_y|\right]}{\partial x_j} = 0$$

By the same arguments, for all g and j,

$$\frac{\partial^2 \log\left[|r_y|\right]}{\partial x_j \partial y_g} = \frac{\partial^2 \log\left[|r_y|\right]}{\partial x_g \partial y_g} = \frac{\partial^2 \log\left[|\widetilde{r}_y|\right]}{\partial x_j \partial y_g} = \frac{\partial^2 \log\left[|\widetilde{r}_y|\right]}{\partial x_g \partial y_g} = 0$$

Hence, for all g, the values of $\partial \log\left[|r_y|\right]/\partial y_g$ and $\partial \log\left[|\widetilde{r}_y|\right]/\partial y_g$ are not functions of $(x_1, ..., x_G)$. We have then shown that $\Delta_x = 0$, and $\partial \Delta_y/\partial x = 0$.

$(T.3.1)$ implies that the properties of r and $\widetilde{r}$ are as those of the additive functions in Section 4.2 in Matzkin (2008). Repeating the arguments in that section, it follows that if $\widetilde{r}$ is observationally equivalent to r, then for any arbitrary argument (y, x) of $\widetilde{r}$ and r.

$$\left(\widetilde{r}'_y - r'_y\right)\ s_\varepsilon = 0$$

where s_ε denotes the $G \times 1$ vector whose $j-th$ element is $\partial \log f_\varepsilon\left(r\left(y, x\right)\right)/\partial \varepsilon_j$.

Since for all g, j, k, $\partial^2 r^g/\partial x_j \partial y_k = \partial^2 \widetilde{r}^g/\partial x_j \partial y_k = 0$, each of the elements of $\widetilde{r}_y$ and of r_y are constant over x. This together with our Assumptions 3.4 and 3.6 imply, as in Matzkin (2008, Section 4.2) that

$$\widetilde{r}'_y = r'_y$$

Hence, r_y and $r_x = I$ are identified. Assume that $r\left(\overline{y}, \overline{x}\right) = \gamma$. Then, since r_y and r_x are identified, r is identified also. And since r is identified, f_ε is identified also. (To see this, note that by the transformation of variables equation

$$f_\varepsilon\left(e\right)=f_{Y|X=x(e)}\left(\overline{y}\right)\quad\left|r_y\left(\overline{y},x(e)\right)\right|^{-1}$$

where $x(e)$ be the value of x at which $r\left(\overline{y},x(e)\right)=e$.)

We remark that, as shown in Matzkin (2008, Section 4.2) for any specific value y^*, we only require the existence of $x^{*(0)}$ to guarantee that for all values x,x'

$$\left(\widetilde{r}_y'\left(y^*,x\right)-r_y'\left(y^*,x\right)\right)\;s_\varepsilon\left(r\left(y^*,x'\right)\right)=0$$

This together with the existence of $x^{*(g)}$ as in Assumption 3.4 implies that the derivatives of the $g-$th coordinate of $\widetilde{r}\left(y^*,x\right)$ must equal those of $r\left(y^*,x\right)$. Hence, when Assumption 3.4' is satisfied, the derivatives of $r\left(y^*,x\right)$ are identified, and when Assumption 3.4" is satisfied, the derivatives of $r^1\left(y^*,x\right)$ are identified. This completes the proof.

3.2. Estimation

Our estimation methods make use of the transformation of variables relationship

$$f_{Y|X=x}\left(y\right)=f_\varepsilon\left(r\left(y,x\right)\right)\left|\frac{\partial r\left(y,x\right)}{\partial y}\right|. \tag{3.2}$$

Let $\widetilde{\mu}(y,x)$ denote a positive function satisfying $\int\widetilde{\mu}(y,x)\;dydx=1$. Let $\widehat{f}_{Y|X=x}\left(y\right)$ denote a nonparametric estimator for $f_{Y|X=x}\left(y\right)$. And let Θ denote the set of functions $\left(\widetilde{r}_y,\widetilde{r}_x\right)$ for which there exists a pair of a structural function and a density $\left(\widetilde{r},f_{\widetilde{\varepsilon}}\right)$ satisfying Assumptions 3.1-3.6. In the proof of Theorem 3.1, we showed that if $\left(r,f_\varepsilon\right)$ and $\left(\widetilde{r},f_{\widetilde{\varepsilon}}\right)$ is a pair such that $\left(r,f_\varepsilon\right)$ and $\left(\widetilde{r},f_{\widetilde{\varepsilon}}\right)$ satisfy Assumptions 3.1-3.6 and for all (y,x)

$$f_\varepsilon\left(r\left(y,x\right)\right)\quad\left|\frac{\partial r\left(y,x\right)}{\partial y}\right|=f_{\widetilde{\varepsilon}}\left(\widetilde{r}\left(y,x\right)\right)\quad\left|\frac{\partial\widetilde{r}\left(y,x\right)}{\partial y}\right|$$

then

$$r_y=\widetilde{r}_y\;\;\text{and}\;\;r_x=\widetilde{r}_x$$

This implies that if $f_{Y|X=x}$ were known, $\left(r_y,r_x\right)$ would be the unique minimizer over Θ of

$$\int\;\left[f_{Y|X=x}\left(y\right)-f_{\widetilde{\varepsilon}}\left(\widetilde{r}\left(y,x\right);\widetilde{r}_y,\widetilde{r}_x\right)\;\left|\widetilde{r}_y\right|\right]^2\;\;\widetilde{\mu}(y,x)\;\;dydx. \tag{3.3}$$

In such minimization, $\widetilde{r}$ and $f_{\widetilde{\varepsilon}}\left(\cdot;\widetilde{r}_y,\widetilde{r}_x\right)$ are any function and density that (i) are consistent with the model and with $\left(\widetilde{r}_y,\widetilde{r}_x\right)$, and (ii) are such that for all (y,x) with $\widetilde{\mu}\left(y,x\right)>0$, $f_{Y|X=x}\left(y\right)=f_{\widetilde{\varepsilon}}\left(\widetilde{r}(y,x);\widetilde{r}_y,\widetilde{r}_x\right)\;\left|\partial\widetilde{r}\left(y,x\right)/\partial y\right|$. A natural estimator for $\left(r_y,r_x\right)$ could then be defined as $(\widehat{r}_y, \widehat{r}_x)$ that minimizes over Θ the distance

$$(3.4) \int \left[\widehat{f}_{Y|X=x}(y) - f_{\widetilde{\varepsilon}}\left(\widetilde{r}(y,x); \widetilde{r}_y, \widetilde{r}_x\right) \, |\widetilde{r}_y| \right]^2 \, \widetilde{\mu}(y,x) \, dydx$$

where $\widehat{f}_{Y|X=x}$ is a nonparametric estimator of $f_{Y|X=x}$. In principle, one could impose sufficient regularity conditions to guarantee that (3.4) converges uniformly in probability to (3.3) and use this to show that $(\widehat{r}_y, \widehat{r}_x)$ converges in probability to (r_y, r_x). A normalization on the value of r at some point would also allow one to obtain consistent nonparametric estimators for r and f_ε over some domain.

Our estimator can be interpreted in a similar way, except that instead of minimizing (3.4), our estimator minimizes over (r_y, r_x) the distance between the derivatives with respect to the observable variables of the logarithms of each side of (3.2):

$$(3.5) \int \left\| \frac{\partial \log \widehat{f}_{Y|X=x}(y)}{\partial (y,x)} - \frac{\partial \log f_\varepsilon\left(r(y,x); r_y, r_x\right)}{\partial (y,x)} - \frac{\partial \log \, |r_y(y,x)|}{\partial (y,x)} \right\|^2 \, \widetilde{\mu}(y,x) \, dydx$$

We will develop below an estimator for $(r_y(y,x), r_x(y,x))$ at one arbitrary value for y. Hence, the weight function $\widetilde{\mu}$ will be defined only over x. We will denote it as $\mu(x)$. As we have shown in the proof of Theorem 3.1, under our assumptions, for all y and x, $r_x(y,x) = I$. Hence, the values of $r_y(y,x)$ are constant over x, and identification of $r_y(y,x)$ requires a condition, described in Assumption 3.4', on $G+1$ points in the support of X. We will assume that our weight function, $\mu(x)$, is strictly positive at those $G+1$ points, but not necessarily strictly positive everywhere. We denote

$$q_y(y,x) = \frac{\partial \log f_{Y|X=x}(y)}{\partial y}; \quad q_x(y,x) = \frac{\partial \log f_{Y|X=x}(y)}{\partial x}, \text{ and}$$

$$l_\varepsilon(y,x) = \frac{\partial \log f_\varepsilon(r(y,x))}{\partial \varepsilon}.$$

Taking first logarithms and next derivatives of (3.2) with respect to y, x,we get that

$$q_y(y,x) = r_y(y,x)' \; l_\varepsilon(r(y,x)) + \frac{|r_y(y,x)|_y}{|r_y(y,x)|}$$

$$q_x(y,x) = r_x(y,x)' \; l_\varepsilon(r(y,x)) + \frac{|r_y(y,x)|_x}{|r_y(y,x)|}$$

Our assumptions imply that $|r_y(y,x)|_x = 0$ and $r_x(y,x) = I$. Hence, our equations become

$$q_y(y,x) = r_y(y,x)'\ l_\varepsilon(r(y,x)) + \frac{|r_y(y,x)|_y}{|r_y(y,x)|}$$

$$q_x(y,x) = l_\varepsilon(r(y,x))$$

Note that this means that if r is identified, one can obtain an estimator of $\partial \log f_\varepsilon / \partial \varepsilon$ directly from an estimator of q_x. Combining these we obtain an expression between the "known" $q_y(y,x)$ and $q_x(y,x)$ and the unknown $r_y(y,x)$ and $|r_y(y,x)|_y$:

$$q_y(y,x) = r_y(y,x)'\ q_x(y,x) + \frac{|r_y(y,x)|_y}{|r_y(y,x)|}$$

Since, as noted above, r_y is constant over x, we will denote $r_y(y,x)$ by $r_y(y)$. Equation (3.5) becomes

$$\text{(3.6)} \qquad \int \left\| q_y(y,x) - \ r_y(y)'\ q_x(y,x) - \frac{|r_y(y)|_y}{|r_y(y)|} \right\|^2 \ \mu(x)\ dx$$

where $\int \mu(x)dx = 1$, and our identification result implies that (3.6) is uniquely minimized at $r_y(y)$. For each g, denote

$$\int q_{y_g}(y) = \int q_{y_g}(y,x)\ \mu(x)\ dx = \int \frac{\partial \log f_{Y|X=x}(y)}{\partial y_g}\ \mu(x)\ dx$$

$$\int q_{x_g}(y) = \int q_{x_g}(y,x)\ \mu(x)\ dx = \int \frac{\partial \log f_{Y|X=x}(y)}{\partial x_g}\ \mu(x)\ dx$$

Hence, from

$$q_y(y,x) = r_y(y)'\ q_x(y,x) + \frac{|r_y(y)|_y}{|r_y(y)|}$$

we get

$$q_y(y,x) - \int q_y(y) = r_y(y)'\ \left(q_x(y,x) - \int q_x(y) \right)$$

Multiplying each row by $\left(q_{x_g}(y,x) - \int q_{x_g}(y)\right)$ and integrating with respect to μ, we get the system of equations, which is the solution to the minimization of (3.6) :

$$\Pi(y)\ \left(\frac{\partial r(y)}{\partial y}\right) = \Gamma(y), \qquad \text{where}$$

$$\Pi(y) = \begin{pmatrix} T_{x_1,x_1}(y) & T_{x_2,x_1}(y) & \cdot & \cdot & T_{x_G,x_1}(y) \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ T_{x_1,x_G}(y) & T_{x_2,x_G}(y) & \cdot & \cdot & T_{x_G,x_G}(y) \end{pmatrix}$$

$$\Gamma(y) = \begin{pmatrix} T_{y_1,x_1}(y) & T_{y_2,x_1}(y) & \cdot & \cdot & T_{y_G,x_1}(y) \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ T_{y_1,x_G}(y) & T_{y_2,x_G}(y) & \cdot & \cdot & T_{y_G,x_G}(y) \end{pmatrix}$$

$$T_{y_g,x_s}(y) = \int \left(q_{y_g}(y,x) - \int q_{y_g}(y) \right) \left(q_{x_s}(y,x) - \int q_{x_s}(y) \right) \mu(x)dx, \quad \text{and}$$

$$T_{x_j,x_s}(y) = \int \left(q_{x_j}(y,x) - \int q_{x_j}(y) \right) \left(q_{x_s}(y,x) - \int q_{x_s}(y) \right) \mu(x)dx.$$

Each of the elements, $T_{y_g,x_s}(y)$ and $T_{x_j,x_s}(y)$, in the matrices Π and Γ, can be estimated from the distribution of the observable variables using average derivative methods, by extending the methods in Stoker (1985) and Powell, Stock, and Stocker (1989). Denote such estimators by $\widehat{T}_{y_g,x_s}(y)$ and $\widehat{T}_{x_j,x_s}(y)$, and let $\widehat{\Pi(y)}$ and $\widehat{\Gamma(y)}$ denote the matrices whose elements are, respectively, $\widehat{T}_{y_g,x_s}(y)$ and $\widehat{T}_{x_j,x_s}(y)$. Then, the estimator for the matrix of derivatives $(\partial r(y)/\partial y)$ is defined as

$$\widehat{\left(\frac{\partial r(y)}{\partial y}\right)} = \left[\widehat{\Pi(y)}\right]^{-1} \widehat{\Gamma(y)}$$

3.3. Asymptotic properties

We develop the asymptotic properties of our estimators for the case when the estimator $\widehat{f}_{Y|X=x}(y)$ for the conditional density of Y given X is estimated by kernel methods. Let $\{Y^i, X^i\}_{i=1}^N$ denote N iid observations generated from $f_{Y,X}$. The estimator is

$$\widehat{f}_{Y|X=x}(y) = \frac{\sum_{i=1}^N K\left(\frac{Y^i-y}{\sigma_N}, \frac{X^i-x}{\sigma_N}\right)}{\sigma_N^G \quad \sum_{i=1}^N K\left(\frac{X^i-x}{\sigma_N}\right)}$$

where K is a kernel function and σ_N is a bandwidth. The element in the j-th row, i-th column of our estimator for $\widehat{T}_{XX}$ is

$$\int \left(\widehat{q}_{x_i}(y,x) - \int \widehat{q}_{x_i}(y) \right) \left(\widehat{q}_{x_j}(y,x) - \int \widehat{q}_{x_j}(y) \right) \mu(x)dx$$

where for $k = 1, ..., G$

$$\widehat{q}_{x_k}(y,x) = \frac{\partial \log \widehat{f}_{Y|X=x}(y)}{\partial x_k} \quad \text{and} \quad \int \widehat{q}_{x_k}(y) = \int \frac{\partial \log \widehat{f}_{Y|X=x}(y)}{\partial x_k} \; \mu(x)\; dx$$

Similarly, the element in the j-th row, i-th column of our estimator for $\widehat{T}_{YX}$ is

$$\int \left(\widehat{q}_{y_i}(y,x) - \int \widehat{q}_{y_i}(y) \right) \left(\widehat{q}_{x_j}(y,x) - \int \widehat{q}_{x_j}(y) \right) \mu(x)dx$$

where for $g = 1, ..., G$

$$\widehat{q}_{y_g}(y,x) = \frac{\partial \log \widehat{f}_{Y|X=x}(y)}{\partial y_g} \quad \text{and} \quad \int \widehat{q}_{y_g}(y) = \int \frac{\partial \log \widehat{f}_{Y|X=x}(y)}{\partial y_g} \; \mu(x) \; dx$$

We will specify the weight function $\mu(x)$ to be bounded, twice continuously differentiable on R^G, strictly positive in the interior of a compact and convex subset, $\overline{M}^x$, of R^G, and zero outside $\overline{M}^x$. We will let $\overline{M}^y$ denote a compact set such that the value y at which we estimate r_y is an interior point of $\overline{M}^y$. Our results use the following assumptions.

Assumption 3.7: *The density $f_{Y,X}$ generated by f_ε and r, is bounded, everywhere positive, and continuously differentiable of order $d \geq s+2$, where s denotes the order of the kernel function. Moreover, there exists $\delta > 0$ such that for all $x \in \overline{M}^x$, $f_X(x) > \delta$ and $f_{Y,X}(y,x) > \delta$.*

Assumption 3.8: *The $G+1$ values satisfying Assumption 3.4' are interior points of $\overline{M}^x$, where $\overline{M}^x$ is the compact and convex subset of R^G such that for every x in the interior of $\overline{M}^x$, $\mu(x) > 0$.*

Assumption 3.9: *The kernel function K is of order s, where $s+2 \leq d$. It attains the value zero outside a compact set, integrates to 1, is differentiable of order Δ, and its derivatives of order Δ are Lipschitz, where $\Delta \geq 2$.*

Assumption 3.10: *The sequence of bandwidths, σ_N, is such that $\sigma_N \to 0$, $N\sigma_N^{G+2} \to \infty$, $\sqrt{N}\sigma_N^{(G/2)+1+s} \to 0$, $\left[N\sigma_N^{2G+2}/\ln(N)\right] \to \infty$, and $\sqrt{N}\sigma_N^{(G/2)+1}\left[\sqrt{\ln(N)/N\sigma_N^{2G+2}} + \sigma_N^s\right]^2 \to 0$.*

To describe the asymptotic behavior of our estimator, we will denote by $rr_y(y)$ the vector in R^{G^2} formed by stacking the columns of $r_y(y)$, so that $rr_y(y) = vec(r_y(y)) = (r^1_{y_1}(y), ..., r^G_{y_1}(y); \; r^1_{y_2}(y), ..., r^G_{y_2}(y); \; ...; \; r^1_{y_G}(y), ..., r^G_{y_G}(y))'$. Let $\widehat{r}_{yy}(y)$ denote the estimator for $rr_y(y)$. Accordingly, we will denote the matrix $TT_{xx}(y)$ by $I_G \otimes T_{xx}(y)$ and its estimator $\widehat{TT}_{xx}(y) = I_G \otimes \widehat{T}_{xx}(y)$. The vector $TT_{yx}(y)$ will be the vector formed by stacking the columns of $T_{yx}(y)$: $TT_{yx}(y) = (T_{y_1,x_1}(y), ..., T_{y_1 x_G}(y); \; T_{y_2,x_1}(y), ..., T_{y_2 x_G}(y); \; ...;$

$T_{y_G,x_1}(y),...,T_{y_Gx_G}(y))'$, with its estimator defined by substituting each coordinate by its estimator. For each s, denote

$$\Delta\partial_{x_s}\log f_{Y|X=x}(y)=\frac{\partial\log f_{Y|X=x}(y)}{\partial x_s}-\int\frac{\partial\log f_{Y|X=x}(y)}{\partial x_s}\mu(x)dx$$

and for each j,k, denote

$$\widetilde{KK}_{y_j,y_k}=\left\{\int\left[\int\left(\frac{\partial K(\widetilde{y},\widetilde{x})}{\partial y_j}\right)d\widetilde{x}\right]\left[\int\left(\frac{\partial K(\widetilde{y},\widetilde{x})}{\partial y_k}\right)d\widetilde{x}\right]d\widetilde{y}\right\}$$

In the proof of Theorem 3.2, which we present in the Appendix, we show that under our assumptions

$$\sqrt{N\sigma_N^{G+2}}\left(\widehat{TT}_{yx}(y)-TT_{yx}(y)\right)\underset{d}{\rightarrow}N\left(0,V(y)\right)$$

where the element in the diagonal of $V(y)$ corresponding to $T_{y_jx_s}$ is

$$\left\{\int\left(\Delta\partial_{x_s}\log f_{Y|X=x}(y)\right)^2\left(\frac{\mu(x)^2}{f_{Y,X}(y,x)}\right)\ dx\right\}\left\{\int\left[\int\left(\frac{\partial K(\widetilde{y},\widetilde{x})}{\partial y_j}\right)d\widetilde{x}\right]^2d\widetilde{y}\right\}$$

and the element in $V(y)$ corresponding to the covariance between $T_{y_jx_s}$ and $T_{y_kx_e}$ is

$$\left\{\int\left(\Delta\partial_{x_s}\log f_{Y|X=x}(y)\right)\left(\Delta\partial_{x_e}\log f_{Y|X=x}(y)\right)\left(\frac{\mu(x)^2}{f_{Y,X}(y,x)}\right)\ dx\right\}\widetilde{KK}_{y_j,y_k}.$$

Theorem 3.2: *Suppose that the model satisfies Assumptions 3.1-3.10. Then,*

$$\sqrt{N\ \sigma_N^{G+2}}\left(\widehat{rr}(y)-rr(y)\right)\underset{d}{\rightarrow}N\left(0,\left(TT_{xx}(y)\right)^{-1}V(y)\left(TT_{xx}(y)\right)^{-1}\right)$$

4. A system with two equations and one instrument

In this section, we consider the model

$$\text{(4.1)}\qquad \begin{aligned} Y_1 &= m^1\left(Y_2,\varepsilon_1\right)\\ Y_2 &= m^2\left(Y_1,X,\varepsilon_2\right)\end{aligned}$$

where the object of interest is the derivative $\partial m^1\left(y_2,\varepsilon_1\right)/\partial y_2$ and where we can estimate the joint density $f_{Y_1,Y_2,X}$ of the observable variables. We concentrate on conditions that are closely related to those used in Section 3. In Section 5, we present additional results that can be used for the identification and estimation of $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ under other assumptions in the model. We make the following assumptions, which are analogous to the assumptions made in Section 3. We note in particular that the support of X does not need to be the real line.

Assumption 4.1: *For all* (y_1, x, ε_2)

$$-\frac{\partial m^2(y_1, x, \varepsilon_2)}{\partial x} = \frac{\partial m^2(y_1, x, \varepsilon_2)}{\partial \varepsilon_2} > 0$$

Assumption 4.2: $(\varepsilon_1, \varepsilon_2)$ *is distributed independently of* X *with an everywhere positive and continuously differentiable density,* f_ε.

Assumption 4.3: X *possesses a twice continuously differentiable density.*

Assumption 4.4: *The functions* $r = (r^1, r^2)$ *and* $h = (h^1, h^2)$ *exist and are twice continuously differentiable,* $1-1$, *and onto* R^2.

We will show identification first under Assumption 4.5 and next under Assumption 4.5'.

Assumption 4.5: *The density,* $f_{\varepsilon_1,\varepsilon_2}$, *of* $(\varepsilon_1, \varepsilon_2)$ *is such that for all* ε_1 *there exists at least one value* $\varepsilon_2^*(\varepsilon_1)$ *of* ε_2 *such that*

$$\frac{\partial^2 \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_2^2} = 0\ .$$

At any such value $\partial^2 \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))/\partial\varepsilon_2\partial\varepsilon_1 \neq 0$

Assumption 4.5': *The density,* $f_{\varepsilon_1,\varepsilon_2}$, *of* $(\varepsilon_1, \varepsilon_2)$ *is such that for all* ε_1, *there exist distinct values* $\varepsilon_2^*(\varepsilon_1)$ *and* $\varepsilon_2^{**}(\varepsilon_1)$ *of* ε_2 *such that*

$$\frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_2} = \frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^{**}(\varepsilon_1))}{\partial \varepsilon_2} = 0$$

For any such two values, $\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))/\partial\varepsilon_1 \neq \partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^{**}(\varepsilon_1))/\partial\varepsilon_1$.

Assumption 4.6: *For any* y_1, y_2 *and* $\varepsilon_1 = r^1(y_1, y_2)$, *there exists a value* x^* *in the interior of the support of* X *such that for* $\varepsilon_2^*(\varepsilon_1)$ *as in Assumption 4.5,* $\varepsilon_2^* = r^2(y_1, y_2, x^*)$.

Assumption 4.6': *For any y_1, y_2 and $\varepsilon_1 = r^1(y_1, y_2)$, there exists distinct values x^* and x^{**} in the interior of the support of X such that for $\varepsilon_2^*(\varepsilon_1)$ and $\varepsilon_2^{**}(\varepsilon_1)$ as in Assumption 4.5', $\varepsilon_2^* = r^2(y_1, y_2, x^*)$ and $\varepsilon_2^{**} = r^2(y_1, y_2, x^{**})$.*

Assumptions 4.5-4.6 and 4.5'-4.6' play a role similar to Assumptions 3.6 and 3.4" in Section 3. Assumption 4.5 is satisfied, for example, when there exists a function $a(\varepsilon_1)$ with $\partial a(\varepsilon_1)/\partial\varepsilon_1 \neq 0$ and the value and derivatives of the conditional density of ε_2 given ε_1 when $\varepsilon_2 = 0$ coincide with the values and derivatives of a conditional density of the form $f_{\varepsilon_2|\varepsilon_1}(\varepsilon_2) = c\ \exp\left(-a(\varepsilon_1)\ \varepsilon_2^3\right)$, for some c which could depend on ε_1. The conditional density $f_{\varepsilon_2|\varepsilon_1}(\varepsilon_2)$ is not restricted to possess this form for values of ε_2 other than $\varepsilon_2 = 0$. An example that satisfies Assumption 4.5' is where the conditional density of ε_2 given ε_1 has the form $f_{\varepsilon_2|\varepsilon_1}(\varepsilon_2) = c\ \exp\left(-a(\varepsilon_1)\ \varepsilon_2^3 - b\ \varepsilon_2^2\right)$ locally at $\varepsilon_2 = 0$ and $\varepsilon_2 = (-2b)/(3a(\varepsilon_1))$. Again, the conditional density $f_{\varepsilon_2|\varepsilon_1}(\varepsilon_2)$ is not restricted to possess this form for other values of ε_2. These conditional densities also satisfy a local invertibility condition at $\varepsilon_2 = 0$ in the first example, and at $\varepsilon_2 = 0$ and $\varepsilon_2 = (-2b)/(3a(\varepsilon_1))$ in the second, which is required in Assumptions 4.8 and 4.8' below, to guarantee desired asymptotic properties for our estimators of $\partial m^1(y_2, \varepsilon_1)/\partial y_2$. Assumptions 4.6 and 4.6' guarantee that the values of ε_2 satisfying, respectively, Assumptions 4.5 and 4.5' possess corresponding values of x in the support of X.

4.1. Identification

The following theorem establishes an identification result for $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ when Assumptions 4.1-4.6 are satisfied. It uses a relationship between $\log f_\varepsilon$ and $\log f_{Y|X}$ in terms of second order derivatives.[4] Theorem 4.2 establishes identification of $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ when Assumptions 4.1-4.4 and 4.5'-4.6' are satisfied, using a relationship between $\log f_\varepsilon$ and $\log f_{Y|X}$ in terms of first order derivatives, as in Section 3. In each case, our proof proceeds by first showing how to identify the value(s) of x satisfying Assumption 4.6 (4.6'), and then using Assumption 4.5 (4.5') to find the appropriate functional of $f_{Y,X}$ from which one can calculate $\partial m^1(y_2, \varepsilon_1)/\partial y_2$.

Theorem 4.1: *Suppose that Assumptions 4.1-4.6 are satisfied. Then,*

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} \text{ is identified}$$

[4] The closest work that uses the relationship between second order derivatives of observed and unobserved densities is, we believe, the generic identification result for multiple choice models in Chiappori and Komunjer (2009).

Proof: The proof is constructive. Differentiating the identity $y_1 = m^1(y_2, r^1(y_1, y_2))$ with respect to y_1 and with respect to y_2, we get that

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = \frac{-r^1_{y_2}}{r^1_{y_1}}$$

where $r^1_{y_1} = \partial r^1(y_1, y_2)/\partial y_1$ and $r^1_{y_2} = \partial r^1(y_1, y_2)/\partial y_2$. Hence, we only need to show identification of the ratio, $r^1_{y_2}/r^1_{y_1}$. As in the proof of Theorem 3.1, Assumption 4.1 implies that $r^2_x = \partial r^2(y_1, y_2, x)/\partial x = 1$. We use the transformation of variables equation

$$f_{Y|X=x}(y_1, y_2) = f_{\varepsilon_1,\varepsilon_2}\left(r^1(y_1, y_2), r^2(y_1, y_2, x)\right) \quad \left|\frac{\partial r(y_1, y_2, x)}{\partial y}\right|$$

to show that under our assumptions, for any (y_1, y_2), the ratio $r^1_{y_2}/r^1_{y_1}$ can be read off the density of the observable variables, $f_{Y|X}(y_1, y_2)$. Taking logarithms, differentiating both sides with respect to x, and noting that $r^2_x = 1$ implies that $\partial\left|\partial r(y_1, y_2, x)/\partial y\right|/\partial x = 0$, we get that for all y_1, y_2, x

$$\frac{\partial \log f_{Y|X=x}(y_1, y_2)}{\partial x} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}\left(r^1(y_1, y_2), r^2(y_1, y_2, x)\right)}{\partial \varepsilon_2}$$

Differentiating with respect to y_1, y_2 and x, we get

$$\frac{\partial^2 \log f_{Y|X=x}(y)}{\partial x \partial y_1} = \frac{\partial^2 \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^2)}{\partial \varepsilon_2 \partial \varepsilon_1} r^1_{y_1} + \frac{\partial^2 \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^2)}{\partial \varepsilon_2 \partial \varepsilon_2} r^2_{y_1}$$

$$\frac{\partial^2 \log f_{Y|X=x}(y)}{\partial x \partial y_2} = \frac{\partial^2 \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^2)}{\partial \varepsilon_2 \partial \varepsilon_1} r^1_{y_2} + \frac{\partial^2 \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^2)}{\partial \varepsilon_2 \partial \varepsilon_2} r^2_{y_2}$$

$$\frac{\partial^2 \log f_{Y|X=x}(y)}{\partial x \partial x} = \frac{\partial^2 \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^2)}{\partial \varepsilon_2 \partial \varepsilon_2}$$

By Assumption 4.5 there exists ε_2^* such that $\partial^2 f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1, \varepsilon_2^*)/\partial \varepsilon_2^2 = 0$, and by Assumptions 4.3 and 4.6, there exists a value x^* such that $\varepsilon_2^* = r^2(y_1, y_2, x^*)$. Conditioning on such value of x, we get that

$$\begin{aligned} \frac{\partial^2 \log f_{Y|X=x^*}(y_1, y_2)}{\partial x \partial y_1} &= \frac{\partial^2 \log f_{\varepsilon_1,\varepsilon_2}}{\partial \varepsilon_2 \partial \varepsilon_1} r^1_{y_1} \quad \text{and} \\ \frac{\partial^2 \log f_{Y|X=x^*}(y_1, y_2)}{\partial x \partial y_2} &= \frac{\partial^2 \log f_{\varepsilon_1,\varepsilon_2}}{\partial \varepsilon_2 \partial \varepsilon_1} r^1_{y_2} \end{aligned}$$

Since by our assumptions $\partial^2 \log f_{\varepsilon_1,\varepsilon_2}\left(r^1(y_1, y_2), r^2(y_1, y_2, x^*)\right)/\partial \varepsilon_2 \partial \varepsilon_1 = \partial^2 \log f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1, \varepsilon_2^*)/\partial \varepsilon_2 \partial \varepsilon_1$

$\neq 0$, it follows from these equations that

$$\frac{r^1_{y_2}}{r^1_{y_1}} = \left[\frac{\partial^2 \log f_{Y|X=x^*}(y_1,y_2)}{\partial x \partial y_1}\right]^{-1}\left[\frac{\partial^2 \log f_{Y|X=x^*}(y_1,y_2)}{\partial x \partial y_2}\right]$$

Hence, $\partial m^1(y_2,\varepsilon_1)/\partial y_2 = -\left[\partial^2 \log f_{Y|X=x^*}(y_1,y_2)/\partial x \partial y_1\right]^{-1}\left[\partial^2 \log f_{Y|X=x^*}(y_1,y_2)/\partial x \partial y_2\right].$ This completes the proof.

Our next theorem establishes identification when assumptions 4.1-4.4 and 4.5'-4.6' are satisfied.

Theorem 4.2: *Suppose that Assumptions 4.1-4.4 and 4.5'-4.6' are satisfied. Then,* $\partial m^1(y_2,\varepsilon_1)/\partial y_2$ *is identified*

Proof: Let (y_1,y_2) be given and $\varepsilon_1 = r^1(y_1,y_2)$. By the same arguments in the proof of Theorem 4.1, we need to show that we can recover the ratio $r^1_{y_2}/r^1_{y_1}$. Taking logarithms and differentiating the transformation of variables equation used in the proof of Theorem 4.1, we get that for all y_1, y_2, x

$$(T.4.1) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_1} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_1} r^1_{y_1} + \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_2} r^2_{y_1} + \frac{\partial \log |r_y|}{\partial y_1}$$

$$(T.4.2) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_2} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_1} r^1_{y_2} + \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_2} r^2_{y_2} + \frac{\partial \log |r_y|}{\partial y_2}$$

$$(T.4.3) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial x} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_2}$$

By Assumption 4.6', there exist x^* and x^{**} such that for ε_2^* and ε_2^{**} as in Assumption 4.5', $\varepsilon_2^* = r^2(y_1,y_2,x^*)$ and $\varepsilon_2^{**} = r^2(y_1,y_2,x^{**})$. Let r^{2*} and r^{2**} denote the value of r^2 at, respectively, (y_1,y_2,x^*) and (y_1,y_2,x^{**}). By $(T.4.1) - (T.4.3)$,

$$\frac{\partial \log f_{Y|X=x^*}(y)}{\partial x} = \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial x} = 0.$$

When $x = x^*$

$$\frac{\partial \log f_{Y|X=x^*}(y)}{\partial y_1} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^{2*})}{\partial \varepsilon_1} r^1_{y_1} + \frac{\partial \log |r_y|}{\partial y_1}$$

$$\frac{\partial \log f_{Y|X=x^*}(y)}{\partial y_2} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^{2*})}{\partial \varepsilon_1} r^1_{y_2} + \frac{\partial \log |r_y|}{\partial y_2}$$

while when $x = x^{**}$,

$$\frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2**})}{\partial \varepsilon_1} r^1_{y_1} + \frac{\partial \log |r_y|}{\partial y_1}$$

$$\frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_2} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2**})}{\partial \varepsilon_1} r^1_{y_2} + \frac{\partial \log |r_y|}{\partial y_2}$$

Taking differences, we get

$$\frac{\partial \log f_{Y|X=x^{*}}(y)}{\partial y_1} - \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1} = \left(\frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2*})}{\partial \varepsilon_1} - \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2**})}{\partial \varepsilon_1}\right) r^1_{y_1}$$

$$\frac{\partial \log f_{Y|X=x^{*}}(y)}{\partial y_2} - \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_2} = \left(\frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2*})}{\partial \varepsilon_1} - \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2**})}{\partial \varepsilon_1}\right) r^1_{y_2}$$

By Assumption 4.5', $\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2*})/\partial \varepsilon_1 \neq \partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2**})/\partial \varepsilon_1$. Hence,

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = \frac{-r^1_{y_2}}{r^1_{y_1}} = \frac{\frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_2} - \frac{\partial \log f_{Y|X=x^{*}}(y)}{\partial y_2}}{\frac{\partial \log f_{Y|X=x^{*}}(y)}{\partial y_1} - \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1}} \text{ is identified.}$$

4.2. Estimation and asymptotic properties

Our estimation methods for $\partial m^1(y_2, \varepsilon_1)/\partial y_2$, under either Assumption 4.5 or 4.5', are closely related to our proofs of identification. When Assumptions 4.1-4.4 and 4.5-4.6 are made, the estimator for $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ is obtained by first estimating nonparametrically the derivatives $\partial^2 \log f_{Y|X=x}(y)/\partial x \partial y_1$, $\partial^2 \log f_{Y|X=x}(y)/\partial x \partial y_2$, and $\partial^2 \log f_{Y|X=x}(y)/\partial x \partial x$ at the particular value of (y_1, y_2) for which we want to estimate $\partial m^1(y_2, \varepsilon_1)/\partial y_2$. The next step consists of finding a value x^* of x satisfying

$$\frac{\partial^2 \widehat{\log f_{Y|X=x^*}}(y)}{\partial x \partial x} = 0$$

The estimator for $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ is then defined by

$$\frac{\partial \widehat{m^1(y_2, \varepsilon_1)}}{\partial y_2} = \frac{-\frac{\partial^2 \widehat{\log f_{Y|X=x^*}}(y)}{\partial x \partial y_2}}{\frac{\partial^2 \widehat{\log f_{Y|X=x^*}}(y)}{\partial x \partial y_1}} \tag{4.2}$$

When $\partial^2 \log f_{Y|X=x}(y)/\partial x \partial y_1$, $\partial^2 \log f_{Y|X=x}(y)/\partial x \partial y_2$, and $\partial^2 \log f_{Y|X=x}(y)/\partial x \partial x$ are estimated using kernel methods, the asymptotic distribution of the estimator for $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ defined in this way can be shown to be consistent and asymptotically normal.

When instead of Assumptions 4.5-4.6, we make Assumption 4.5'-4.6', our estimator for $\partial m^1(y_2,\varepsilon_1)/\partial y_2$, is obtained by first estimating nonparametrically $\partial \log f_{Y|X=x}(y)/\partial x$, $\partial \log f_{Y|X=x}(y)/\partial y_1$, and $\partial \log f_{Y|X=x}(y)/\partial y_2$ at the particular value of (y_1,y_2) for which we want to estimate $\partial m^1(y_2,\varepsilon_1)/\partial y_2$. The next step consists of finding values x^* and x^{**} of x satisfying

$$\frac{\partial \log \widehat{f_{Y|X=x^*}(y)}}{\partial x} = \frac{\partial \log \widehat{f_{Y|X=x^{**}}(y)}}{\partial x} = 0$$

Our estimator for $\partial m^1(y_2,\varepsilon_1)/\partial y_2$ is then defined by

$$(4.3) \qquad \frac{\partial \widehat{m^1(y_2,\varepsilon_1)}}{\partial y_2} = \frac{\frac{\partial \log \widehat{f_{Y|X=x^{**}}(y)}}{\partial y_2} - \frac{\partial \log \widehat{f_{Y|X=x^*}(y)}}{\partial y_2}}{\frac{\partial \log \widehat{f_{Y|X=x^*}(y)}}{\partial y_1} - \frac{\partial \log \widehat{f_{Y|X=x^{**}}(y)}}{\partial y_1}}$$

Again, when $\partial \log f_{Y|X=t}(y)/\partial y_1$, $\partial \log f_{Y|X=t}(y)/\partial y_2$, and $\partial \log f_{Y|X=t}(y)/\partial x$ are estimated using kernels, the estimator for $\partial m^1(y_2,\varepsilon_1)/\partial y_2$ defined in this way will be consistent and asymptotically normal, under standard conditions. We present below the asymptotic properties for both estimators, without exploiting possible averaging, as in Section 3.

Asymptotic properties

To derive the asymptotic properties of the estimator defined in (4.2), we make the following assumptions

Assumption 4.7: *The density f_ε and the density $f_{Y,X}$ generated by f_ε and r, are bounded, everywhere positive, and continuously differentiable of order d, where $d \geq 5+s$ and s denotes the order of the kernel function $K(\cdot)$, specified below in Assumption 4.10.*

Assumption 4.8: *For any x' such that $\partial^2 \log f_\varepsilon\left(r^1(y_1,y_2), r^2(y_1,y_2,x')\right)/\partial \varepsilon_2^2 = 0$, there exist a neighborhood $B'_{y,x}$ of (y_1,y_2,x') and B'_x of x' such that the density $f_X(x)$ and the density $f_{Y,X}(y,x) = f_\varepsilon\left(r^1(y_1,y_2), r^2(y_1,y_2,x)\right) \left|r_y(y_1,y_2,x)\right| f_X(x)$ are uniformly bounded away from zero on, respectively, B'_x and $B'_{y,x}$ and $\partial^3 \log f_\varepsilon\left(r^1(y_1,y_2), r^2(y_1,y_2,x)\right)/\partial \varepsilon_2^3$ is bounded away from zero on those neighborhoods.*

Assumption 4.9: *For any x' such that $\partial^2 \log f_\varepsilon\left(r^1(y_1,y_2), r^2(y_1,y_2,x')\right)/\partial \varepsilon_2^2 = 0$, $\partial^2 \log f_\varepsilon\left(r^1(y_1,y_2), r^2(y_1,y_2,x')\right)/\partial \varepsilon_1 \partial \varepsilon_2$ is uniformly bounded away from 0 on the neighborhood $B'_{y,x}$ defined in Assumption 4.8.*

Assumption 4.10: *The kernel function K attains the value zero outside a compact set,* *integrates to 1, is of order* s where $s+5 \le d$, *is differentiable of order* Δ, *and its derivatives of order* Δ *are Lipschitz, where* $\Delta \ge 5$.

Assumption 4.11: *The sequence of bandwidths,* σ_N*, is such that* $\sigma_N \to 0$, $\sqrt{N\sigma_N^{7+2s}} \to 0$, $\sqrt{N\sigma_N^{7}} \to \infty$, $\left[N\sigma_N^{11}/\ln(N)\right] \to \infty$, *and* $\sqrt{N\sigma_N^{7}}\left[\sqrt{\ln(N)/N\sigma_N^{11}} + \sigma_N^{s}\right]^2 \to 0$.

Assumptions 4.7, 4.10 and 4.11 are standard for derivations of asymptotic results of kernel estimators. Assumptions 4.8 are 4.9 are made to guarantee appropriate asymptotic behavior of the estimator for the value x^* at which $\widehat{\partial m^1 \left(y_2, \varepsilon_1\right) / \partial y_2}$ is calculated. Define $\overline{K}_{y_1,x}(\widetilde{y}_1,\widetilde{y}_2,x) = \frac{\partial^2 K(\widetilde{y}_1,\widetilde{y}_2,x)}{\partial y_1 \partial x}$, $\overline{K}_{y_2,x}(\widetilde{y}_1,\widetilde{y}_2,x) = \frac{\partial^2 K(\widetilde{y}_1,\widetilde{y}_2,x)}{\partial y_2 \partial x}$, and $\overline{K}_{x,x}(\widetilde{y}_1,\widetilde{y}_2,x) = \frac{\partial^2 K(\widetilde{y}_1,\widetilde{y}_2,x)}{\partial x^2}$. Let $\overline{K}(\widetilde{y}_1,\widetilde{y}_2,x)$ denote the 3×1 vector $\left(\overline{K}_{y_1,x}(\widetilde{y}_1,\widetilde{y}_2,x), \overline{K}_{y_2,x}(\widetilde{y}_1,\widetilde{y}_2,x), \overline{K}_{x,x}(\widetilde{y}_1,\widetilde{y}_2,x)\right)'$. Define the vector $\omega(y,x^*) = (\omega_1,\omega_2,\omega_3)'$ where

$$\omega_1 = \frac{\frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial y_2 \partial x}}{\left[\frac{\partial^2 \log f_{Y|X=x^*}(x^*)}{\partial y_1 \partial x}\right]^2 f_{Y,X}(y,x^*)}; \omega_2 = \frac{-\ 1}{\frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial y_1 \partial x} f_{Y,X}(y,x^*)};$$

$$\omega_3 = \frac{\partial \left(\frac{\partial^2 \log f_{Y|X=x^*}(y)/\partial y_2 \partial x}{\partial^2 \log f_{Y|X=x^*}(y)/\partial y_1 \partial x}\right)/\partial x}{f_{Y,X}(y,x^*)\left(\frac{\partial^3 \log f_{Y|X=x^*}(y)}{\partial x^3}\right)}$$

Let

$$\widetilde{V} = \omega(y,x^*)' \left[\int \overline{K}(\widetilde{y}_1,\widetilde{y}_2,x)\ \overline{K}(\widetilde{y}_1,\widetilde{y}_2,x)'\ d(\widetilde{y}_1,\widetilde{y}_2,x)\right] \omega(y,x^*)\ f_{Y,X}(y,x^*)$$

In the Appendix we prove

Theorem 4.3: *Suppose that the model satisfies Assumptions 4.1-4.11. Let the estimator for* $\partial m^1 \left(y_2, \varepsilon_1\right) / \partial y_2$ *be as defined in (4.2). Then,*

$$\sqrt{N\ \sigma_N^7}\left(\widehat{\partial m^1 \left(y_2, \varepsilon_1\right) / \partial y_2} - \partial m^1 \left(y_2, \varepsilon_1\right) / \partial y_2\right) \underset{d}{\to} N\left(0, \widetilde{V}\right)$$

To derive the asymptotic properties of the estimator defined in (4.3), we make the following assumptions

Assumption 4.7': *The density* f_ε *and the density* $f_{Y,X}$ *generated by* f_ε *and* r*, are bounded, everywhere positive, and continuously differentiable of order* d*, where* $d \ge 4+s$ *and* s *is the order of the kernel function* $K(\cdot)$ *in Assumption 4.10'.*

Assumption 4.8': *For any x' such that $\partial \log f_{\varepsilon}\left(r^{1}\left(y_{1}, y_{2}\right), r^{2}\left(y_{1}, y_{2}, x'\right)\right) / \partial \varepsilon_{2}=0$, there exist a neighborhood $B_{y,x}^{'}$ of (y_1, y_2, x') and B_x' of x' such that the density $f_X(x)$ and the density $f_{Y,X}(y,x) = f_{\varepsilon}\left(r^{1}\left(y_{1}, y_{2}\right), r^{2}\left(y_{1}, y_{2}, x\right)\right) \ \left|r_y\left(y_1,y_2,x\right)\right| \ f_X(x)$ are uniformly bounded away from zero on, respectively, B_x' and $B_{y,x}^{'}$ and $\partial^2 \log f_{\varepsilon}\left(r^{1}\left(y_{1}, y_{2}\right), r^{2}\left(y_{1}, y_{2}, x\right)\right) / \partial \varepsilon_{2}^2$ is bounded away from zero on those neighborhoods.*

Assumption 4.9': *For any two values x', x'' such that $\partial \log f_{\varepsilon}\left(r^{1}\left(y_{1}, y_{2}\right), r^{2}\left(y_{1}, y_{2}, x'\right)\right) / \partial \varepsilon_{2}=$ 0, and $\partial \log f_{\varepsilon}\left(r^{1}\left(y_{1}, y_{2}\right), r^{2}\left(y_{1}, y_{2}, x''\right)\right) / \partial \varepsilon_{2}=0$, $\left(\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2\prime})/\partial\varepsilon_1 - \partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2\prime\prime})/\partial\varepsilon_1\right)$ is uniformly bounded away from 0 on the neighborhoods $B_{y,x}^{'}, B_{y,x}^{''}$, B_x' and B_x'' defined on Assumption 4.8'.*

Assumption 4.10': *The kernel function K attains the value zero outside a compact set, integrates to 1, is of order s, where $s+4 \leq d$, is differentiable of order Δ, and its derivatives of order Δ are Lipschitz, where $\Delta \geq 4$.*

Assumption 4.11': *The sequence of bandwidths, σ_N, is such that $\sqrt{N\sigma_N^5} \to \infty$, $\sqrt{N\sigma_N^5}\sigma^s \to 0$, $\left[\sqrt{\ln(N)/N\sigma_N^9} + \sigma_N^s\right] \to 0$, and $\sqrt{N\sigma_N^5}\left[\sqrt{\ln(N)/N\sigma_N^9} + \sigma_N^s\right]^2 \to 0$.*

Define $\overline{K}_{y_1}(\widetilde{y}_1,\widetilde{y}_2,x) = \frac{\partial K(\widetilde{y}_1,\widetilde{y}_2,x)}{\partial y_1}$, $\overline{K}_{y_2}(\widetilde{y}_1,\widetilde{y}_2,x) = \frac{\partial K(\widetilde{y}_1,\widetilde{y}_2,x)}{\partial y_2}$, and $\overline{K}_{x}(\widetilde{y}_1,\widetilde{y}_2,x) = \frac{\partial K(\widetilde{y}_1,\widetilde{y}_2,x)}{\partial x}$. Let $\widetilde{K}(\widetilde{y}_1,\widetilde{y}_2,x)$ denote the 3×1 vector $\left(\overline{K}_{y_1}(\widetilde{y}_1,\widetilde{y}_2,x), \overline{K}_{y_2}(\widetilde{y}_1,\widetilde{y}_2,x), \overline{K}_{x}(\widetilde{y}_1,\widetilde{y}_2,x)\right)'$. Define the vectors $\omega^1 = (\omega_1^1, \omega_2^1, \omega_3^1)$ and $\omega^2 = (\omega_1^2, \omega_2^2, \omega_3^2)$ by

$$\omega_1^1 = \frac{-\left[\frac{\partial f_{Y,X}\left(y,x_2^*\right)}{\partial y_2} f_{Y,X}\left(y,x_1^*\right) - \frac{\partial f_{Y,X}\left(y,x_1^*\right)}{\partial y_2} f_{Y,X}\left(y,x_2^*\right)\right] f_{Y,X}\left(y,x_2^*\right)}{\left[\frac{\partial f_{Y,X}\left(y,x_1^*\right)}{\partial y_1} f_{Y,X}\left(y,x_2^*\right) - \frac{\partial f_{Y,X}\left(y,x_2^*\right)}{\partial y_1} f_{Y,X}\left(y,x_1^*\right)\right]^2};$$

$$\omega_2^1 = \frac{-f_{Y,X}\left(y,x_2^*\right)}{\left[\frac{\partial f_{Y,X}\left(y,x_1^*\right)}{\partial y_1} f_{Y,X}\left(y,x_2^*\right) - \frac{\partial f_{Y,X}\left(y,x_2^*\right)}{\partial y_1} f_{Y,X}\left(y,x_1^*\right)\right]};$$

$$\omega_3^1 = \frac{-\frac{\partial}{\partial x_1}\left(\frac{\partial \log f_{Y|X=x_2^*}(y)/\partial y_2 - \partial \log f_{Y|X=x_1}(y)/\partial y_2}{\partial \log f_{Y|X=x_1}(y)/\partial y_1 - \partial \log f_{Y|X=x_2^*}(y)/\partial y_1}\right)|_{x_1=x_1^*}}{\left(\frac{\partial^2 \log f_{Y|X=x_1^*}(y)}{\partial x^2}\right) f_{Y,X}\left(y,x_1^*\right)};$$

$$\omega_1^2 = \frac{\left[\frac{\partial f_{Y,X}\left(y,x_2^*\right)}{\partial y_2} f_{Y,X}\left(y,x_1^*\right) - \frac{\partial f_{Y,X}\left(y,x_1^*\right)}{\partial y_2} f_{Y,X}\left(y,x_2^*\right)\right] f_{Y,X}\left(y,x_1^*\right)}{\left[\frac{\partial f_{Y,X}\left(y,x_1^*\right)}{\partial y_1} f_{Y,X}\left(y,x_2^*\right) - \frac{\partial f_{Y,X}\left(y,x_2^*\right)}{\partial y_1} f_{Y,X}\left(y,x_1^*\right)\right]^2};$$

$$\omega_2^2 = \frac{f_{Y,X}\left(y, x_1^*\right)}{\left[\frac{\partial f_{Y,X}\left(y,x_1^*\right)}{\partial y_1} f_{Y,X}\left(y, x_2^*\right) - \frac{\partial f_{Y,X}\left(y,x_2^*\right)}{\partial y_1} f_{Y,X}\left(y, x_1^*\right)\right]}; \text{ and}$$

$$\omega_3^2 = \frac{-\frac{\partial}{\partial x_2}\left(\frac{\partial \log f_{Y|X=x_2}(y)/\partial y_2 - \partial \log f_{Y|X=x_1^*}(y)/\partial y_2}{\partial \log f_{Y|X=x_1^*}(y)/\partial y_1 - \partial \log f_{Y|X=x_2}(y)/\partial y_1}\right)|_{x_2=x_2^*}}{\left(\frac{\partial^2 \log f_{Y|X=x_2^*}(y)}{\partial x^2}\right) f_{Y,X}\left(y, x_2^*\right)}.$$

Define

$$\begin{aligned}\overline{V} &= \omega^{1\prime}\left[\int \widetilde{K}(\widetilde{y}_1, \widetilde{y}_2, x)\widetilde{K}(\widetilde{y}_1, \widetilde{y}_2, x)' d(\widetilde{y}_1, \widetilde{y}_2, x)\right] \omega^1 \; f_{Y,X}\left(y, x_1^*\right) \\ &\quad + \omega^{2\prime}\left[\int \widetilde{K}(\widetilde{y}_1, \widetilde{y}_2, x)\widetilde{K}(\widetilde{y}_1, \widetilde{y}_2, x)' d(\widetilde{y}_1, \widetilde{y}_2, x)\right] \omega^2 \; f_{Y,X}\left(y, x_2^*\right)\end{aligned}$$

In the Appendix we prove

Theorem 4.4: *Suppose that Assumptions 4.1-4.4, and 4.5'-4.11' are satisfied. Define* $\partial m^1\widehat{(y_1, \varepsilon_1)}/\partial y_2$ *as in (4.3). Then,*

$$\sqrt{N\sigma_N^5}\left(\partial m^1\widehat{(y_2, \varepsilon_1)}/\partial y_2 - \partial m^1(y_2, \varepsilon_1)/\partial y_2\right) \to N\left(0, \overline{V}\right)$$

5. Further identification and estimation results

The conditions in Section 4 under which we showed the identification of $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ are by no means unique. Several other sets of conditions could be used, and estimators based on them could be developed. We provide here results that can guide one to those other sets of conditions. We consider again the model

$$\begin{aligned} y_1 &= m^1\left(y_2, \varepsilon_1\right) \\ y_2 &= m^2\left(y_1, x, \varepsilon_2\right) \end{aligned} \tag{5.1}$$

where m^1, m^2, and $f_{\varepsilon_1,\varepsilon_2}$ are unknown. We will assume the existence of the functions $h = (h^1, h^2)$ and $r = (r^1, r^2)$ and Assumption 4.2-4.4. Our objective is to estimate

$$\frac{\partial m^1\left(y_2, \varepsilon_1\right)}{\partial y_2},$$

the partial effect of exogenously changing the value of the observable explanatory variable, y_2, leaving the value of the unobservable ε_1 unchanged.

5.1. Imposing additional structure

One possibility to obtain additional estimators is to impose more structure on the functions m^1 and m^2. Suppose that the model is

$$\begin{aligned} y_1 &= m(y_2) + \varepsilon_1 \\ y_2 &= \beta\, y_1 + \gamma\, x + \varepsilon_2 \end{aligned}$$

The derivative of m at y_2 is identified when $(\varepsilon_1, \varepsilon_2)$ has an everywhere positive, differentiable density $f_{\varepsilon_1,\varepsilon_2}$ such that for two, not necessarily known a-priori, values $(\overline{\varepsilon}_1, \overline{\varepsilon}_2)$ and $(\varepsilon_1'', \varepsilon_2'')$,

$$\frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\overline{\varepsilon}_1, \overline{\varepsilon}_2)}{\partial \varepsilon_1} \neq \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1'', \varepsilon_2'')}{\partial \varepsilon_1}$$

and

$$\frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\overline{\varepsilon}_1, \overline{\varepsilon}_2)}{\partial \varepsilon_2} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1'', \varepsilon_2'')}{\partial \varepsilon_2} = 0$$

(See Matzkin (2007a).) Fix the value of y_2. Taking logarithms and derivatives of the transformation of variables equation as in Section 3.2, but integrating this time with respect to a measure of both x and y_1, we obtain the relationship

$$\begin{pmatrix} T_{y_1,y_1} & T_{y_1,x} \\ T_{y_1,x} & T_{x,x} \end{pmatrix} \begin{pmatrix} -\frac{\partial m(y_2)}{\partial y_2} \\ \frac{\beta\left(\frac{\partial m(y_2)}{\partial y_2}\right)-1}{\gamma} \end{pmatrix} = \begin{pmatrix} T_{y_2,y_1} \\ T_{y_2,x} \end{pmatrix}.$$

Substituting the elements in the first and third matrices by nonparametric estimators, we obtain an estimator of the derivative of m with respect to y_2.

5.2. Estimating derivatives of the reduced form

In many situations it is easier to consider estimation of the derivatives of the reduced form functions, h, instead of estimation of the derivatives of the structural functions r. The following theorem establishes the relationship between the derivatives of h and the derivative of m^1. It states that to identify the derivative of m^1 with respect to y_2, it suffices to identify, at only one value of x, the ratio of the derivatives with respect to x of the reduced form functions h^1 and h^2.

Theorem 5.1: *Suppose that model (5.1) satisfies Assumption 4.2-4.4. Then, for any* (y_2, ε_1) *and all* x,

$$(5.1.1) \qquad \frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2} = \frac{\frac{\partial h^1(x,\varepsilon_1,\varepsilon_2)}{\partial x}}{\frac{\partial h^2(x,\varepsilon_1,\varepsilon_2)}{\partial x}}\Big|_{\varepsilon_2=r^2(m^1(y_2,\varepsilon_1),y_2,x)}$$

Proof of Theorem 5.1: Differentiating with respect to x the identity

$$\varepsilon = r\left(h(x,\varepsilon),x\right)$$

we get

$$\left(\frac{\partial h}{\partial x}\right) = -\left(\frac{\partial r}{\partial y}\right)^{-1}\left(\frac{\partial r}{\partial x}\right)$$

which in our model becomes

$$h_x^1 = \frac{r_{y_2}^1 \; r_x^2}{|r_y|} \qquad \text{and} \qquad h_x^2 = \frac{-r_{y_1}^1 \; r_x^2}{|r_y|}$$

Hence, for all x,

$$\frac{\partial m_1(y_2,\varepsilon_1)}{\partial y_2} = \frac{\frac{\partial h^1(x,\varepsilon_1,\varepsilon_2)}{\partial x}}{\frac{\partial h^2(x,\varepsilon_1,\varepsilon_2)}{\partial x}}\Big|_{\varepsilon_2=r^2(m^1(y_2,\varepsilon_1),y_2,x)}$$

This completes the proof.

To estimate the derivative $\partial m^1(y_2,\varepsilon_1)/\partial y_2$ through estimation of the derivatives of h, we can use the following transformation of variables

$$f_\varepsilon(\varepsilon) = f_{Y|X=x}(h(x,\varepsilon)) \quad \left|\frac{\partial h(x,\varepsilon)}{\partial \varepsilon}\right|$$

Taking logarithms and differentiating with respect to x, we get

$$\frac{-\,\partial \log f_{Y_1,Y_2|X=x}(y_1,y_2)}{\partial x} = \frac{\partial \log f_{Y_1,Y_2|X=x}(y_1,y_2)}{\partial y_1}h_x^1 + \frac{\partial \log f_{Y_1,Y_2|X=x}(y_1,y_2)}{\partial y_2}h_x^2 + \frac{\partial \log\left|\frac{\partial h(x,\varepsilon)}{\partial \varepsilon}\right|}{\partial x}$$

To get an expression for the last term, we can differentiate $h\left(x, r\left(y,x\right)\right) = y$ with respect to y, to get that

$$\left(\frac{\partial h(x,\varepsilon)}{\partial \varepsilon}\right)\left(\frac{\partial r(y,x)}{\partial y}\right) = I$$

which implies that

$$\begin{aligned}
\frac{\partial \log \left|\frac{\partial h(x,\varepsilon)}{\partial \varepsilon}\right|}{\partial x} &= -\frac{d \log \left|\frac{\partial r(h(x,\varepsilon),x)}{\partial y}\right|}{dx} \\
&= -\left(\frac{\partial h(x,\varepsilon)}{\partial x}\right)' \frac{\partial \log \left|\frac{\partial r(y,x)}{\partial y}\big|_{y=h(x,\varepsilon)}\right|}{\partial y} - \frac{\partial \log \left|\frac{\partial r(y,x)}{\partial y}\big|_{y=h(x,\varepsilon)}\right|}{\partial x}
\end{aligned}$$

It can be verified that this last term equals $\partial\left(\left(r^1_{y_2}\ r^2_x\right)/\left|r_y\right|\right)/\partial y_1 + \partial\left(\left(-r^1_{y_1}\ r^2_x\right)/\left|r_y\right|\right)/\partial y_2$, which we denote loosely as $\partial\left(h^1_x\right)/\partial y_1 + \partial\left(h^2_x\right)/\partial y_2$. Hence, we can state that

$$\text{(5.2)} \qquad \begin{aligned}
\frac{-\partial \log f_{Y_1,Y_2|X=x}(y_1,y_2)}{\partial x} &= \frac{\partial \log f_{Y_1,Y_2|X=x}(y_1,y_2)}{\partial y_1} h^1_x + \frac{\partial \log f_{Y_1,Y_2|X=x}(y_1,y_2)}{\partial y_2} h^2_x \\
&\quad + \frac{\partial}{\partial y_1}\left(h^1_x\right) + \frac{\partial}{\partial y_2}\left(h^2_x\right)
\end{aligned}$$

where

$$h^1_x = \frac{r^1_{y_2}\ r^2_x}{\left|r_y\right|} \qquad \text{and} \qquad h^2_x = \frac{-r^1_{y_1}\ r^2_x}{\left|r_y\right|}$$

Equation (5.2) expresses a linear relationship between the "observable" derivatives of log $f_{Y_1,Y_2|X=x}(y_1,y_2)$ and the "unobservable" derivatives of the reduced form functions, h^1 and h^2. Imposing restrictions on the functions r and on the density $f_{\varepsilon_1,\varepsilon_2}$ one can use this equation to recover either h^1_x and h^2_x, or the ratio, h^1_x/h^2_x.

The two estimators developed in Section 4 can be interpreted as being derived from (5.2). Consider for example the first estimator in Section 4. The assumptions we made in that section on the function r imply that $\partial\left(h^1_x\right)/\partial x = \partial\left(h^2_x\right)/\partial x = \partial^2\left(h^1_x\right)/\partial x \partial y_1 = \partial^2\left(h^2_x\right)/\partial x \partial y_2 = 0$. Differentiating both sides of (5.2) with respect to x, we then get

$$\text{(5.3)} \qquad \frac{-\partial^2 \log f_{Y_1,Y_2|X=x}(y_1,y_2)}{\partial x^2} = \frac{\partial^2 \log f_{Y_1,Y_2|X=x}(y_1,y_2)}{\partial x \partial y_1} h^1_x + \frac{\partial^2 \log f_{Y_1,Y_2|X=x}(y_1,y_2)}{\partial x \partial y_2} h^2_x$$

Hence, at x^* such that

$$\partial \log f_{Y_1,Y_2|X=x}(y_1,y_2)/\partial x^2 = 0$$

we have that

$$- \frac{h_x^1}{h_x^2} = \frac{- \partial^2 \log f_{Y_1,Y_2|X=x^*} (y_1, y_2) / \partial x \partial y_2}{\partial^2 \log f_{Y_1,Y_2|X=x^*} (y_1, y_2) / \partial x \partial y_1}$$

The assumptions we made in Section 4 on $f_{\varepsilon_1,\varepsilon_2}$ guarantee that such value of x^* exists.

The second estimator uses two values, x^* and x^{**}, of x at which $\partial \log f_{Y_1,Y_2|X=x} (y_1, y_2) / \partial x = 0$. The assumptions we made in Section 4 imply that h_x^1, h_x^2, and their derivatives are constant across these two values. Hence, solving for h_x^1/h_x^2 using (5.2) at those two values, we get

$$- \frac{h_x^1}{h_x^2} = \frac{\frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_2} - \frac{\partial \log f_{Y|X=x^{*}}(y)}{\partial y_2}}{\frac{\partial \log f_{Y|X=x^{*}}(y)}{\partial y_1} - \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1}}.$$

By Theorem 5.1,

$$\frac{\partial m_1 (y_2, \varepsilon_1)}{\partial y_2} = \frac{- \partial^2 \log f_{Y_1,Y_2|X=x^*} (y_1, y_2) / \partial x \partial y_2}{\partial^2 \log f_{Y_1,Y_2|X=x^*} (y_1, y_2) / \partial x \partial y_1}$$

in the first model and, in the second,

$$\frac{\partial m_1 (y_2, \varepsilon_1)}{\partial y_2} = \frac{\frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_2} - \frac{\partial \log f_{Y|X=x^{*}}(y)}{\partial y_2}}{\frac{\partial \log f_{Y|X=x^{*}}(y)}{\partial y_1} - \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1}}.$$

where, in both, $\varepsilon_1 = r^1 (y_1, y_2)$.

5.3. Observational equivalence

Characterizations of observational equivalence are useful to determine restrictions guaranteeing that the true pair (r, f_ε), or some feature of it, is the unique solution to a set of equations that depend on the density of the observable variables. These restrictions provide the basis upon which consistent nonparametric and parametric estimation can be developed. When the values or structures of an estimator are restricted to belong to such a set, one can guarantee that the critical identification conditions for consistency of the estimator will be satisfied. Our next theorem provides a set of such characterizations for Model (5.1). The characterizations are expressed first in terms of the reduced form functions, $h (x, \varepsilon_1, \varepsilon_2)$, then in terms of the normalized Jacobian determinant, $|r_y| / r_x^2$, and finally in terms of the elements of the second equation, $\left(r^2, f_{\varepsilon_2|\varepsilon_1}\right)$. The identification results in Section 3 and 4 are particular cases of these general characterizations. For other identified cases that can similarly be derived from the characterizations below, one can modify the estimation methods that we developed in those sections.

Theorem 5.2: *Suppose that in Model (5.1), Assumptions 4.2-4.4 are satisfied. Then, the following statements are equivalent*

(i) $(\widetilde{r}, f_{\widetilde{\varepsilon}})$ *is observationally equivalent to* (r, f_{ε}),

(ii) for all y, x

$$0 = \frac{\partial \left(f_{Y|X=x}\left(y_1, y_2\right) \ \left(\widetilde{h}_x^1\left(x, \widetilde{r}^1, \widetilde{r}^2\right) - h_x^1\left(x, r^1, r^2\right) \right) \right)}{\partial y_1}$$

$$+ \frac{\partial \left(f_{Y|X=x}\left(y_1, y_2\right) \ \left(\widetilde{h}_x^2\left(x, \widetilde{r}^1, \widetilde{r}^2\right) - h_x^2\left(x, r^1, r^2\right) \right) \right)}{\partial y_2},$$

where h *and* $\widetilde{h}$ *are the reduced form functions of, respectively,* r *and* $\widetilde{r}$, *and where the arguments of* $r^1, r^2, \widetilde{r}^1, \widetilde{r}^2$ *are* (y_1, y_2, x).

(iii) for all y, x

$$\widetilde{r}_{y_2}^1 \frac{\partial}{\partial y_1}\left[f_{Y_1,Y_2|X=x}\left(y_1, y_2\right) \ \frac{\widetilde{r}_x^2}{|\widetilde{r}_y|} \right] - \widetilde{r}_{y_1}^1 \frac{\partial}{\partial y_2}\left[f_{Y_1,Y_2|X=x}\left(y_1, y_2\right) \ \frac{\widetilde{r}_x^2}{|\widetilde{r}_y|} \right]$$

$$= r_{y_2}^1 \frac{\partial}{\partial y_1}\left[f_{Y_1,Y_2|X=x}\left(y_1, y_2\right) \ \frac{r_x^2}{|r_y|} \right] - r_{y_1}^1 \frac{\partial}{\partial y_2}\left[f_{Y_1,Y_2|X=x}\left(y_1, y_2\right) \ \frac{r_x^2}{|r_y|} \right]$$

where $r^1 = r^1(y_1, y_2)$, $\widetilde{r}^1 = \widetilde{r}^1(y_1, y_2)$, $r^2 = r^2(y_1, y_2, x)$ *and* $\widetilde{r}^2 = \widetilde{r}^2(y_1, y_2, x)$,

(iv) for all (y, x),

$$\frac{\partial F_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)}{\partial y_1} \frac{\partial F_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)}{\partial y_2} - \frac{\partial F_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)}{\partial y_2} \frac{\partial F_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)}{\partial y_1}$$

$$= \frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_1} \frac{\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)}{\partial y_2} - \frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_2} \frac{\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)}{\partial y_1}$$

where $r^1 = r^1(y_1, y_2)$, $\widetilde{r}^1 = \widetilde{r}^1(y_1, y_2)$, $r^2 = r^2(y_1, y_2, x)$ *and* $\widetilde{r}^2 = \widetilde{r}^2(y_1, y_2, x)$, *and where* $F_{\widetilde{\varepsilon}_1}$ *and* F_{ε_1} *denote the marginal cumulative distributions of, respectively,* $\widetilde{\varepsilon}_1$ *and* ε_1, *and* $F_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1}$ *and* $F_{\varepsilon_2|\varepsilon_1}$ *denote the conditional cumulative distributions of, respectively,* $\widetilde{\varepsilon}_2$ *given* $\widetilde{\varepsilon}_1$, *and* ε_2 *given* ε_1.

The proof of this theorem is presented in the Appendix. Condition *(ii)* characterizes observational equivalence in terms of the relationship between $f_{Y|X=x}$ and the derivatives of the reduced form functions, (h^1, h^2) and $\left(\widetilde{h}^1, \widetilde{h}^2\right)$. After dividing by $f_{Y|X=x}(y_1, y_2) > 0$, the condition can be written, for all y, x, as

$$0 = \frac{\partial \log\left(f_{Y|X=x}\left(y_1, y_2\right)\right)}{\partial y_1} \ \left(\widetilde{h}_x^1\left(x, \widetilde{r}^1, \widetilde{r}^2\right) - h_x^1\left(x, r^1, r^2\right) \right)$$

$$+\frac{\partial \log\left(f_{Y|X=x}\left(y_1,y_2\right)\right)}{\partial y_2}\left(\widetilde{h}_x^2\left(x,\widetilde{r}^1,\widetilde{r}^2\right)-h_x^2\left(x,r^1,r^2\right)\right)$$

$$+\frac{\partial\left(\widetilde{h}_x^1\left(x,\widetilde{r}^1,\widetilde{r}^2\right)-h_x^1\left(x,r^1,r^2\right)\right)}{\partial y_1}+\frac{\partial\left(\widetilde{h}_x^2\left(x,\widetilde{r}^1,\widetilde{r}^2\right)-h_x^2\left(x,r^1,r^2\right)\right)}{\partial y_2}$$

When the only solution to this expression is $\widetilde{h}_x^1 = h_x^1$, $\widetilde{h}_x^2 = h_x^2$, and $\partial\left(\widetilde{h}_x^1 - h_x^1\right)/\partial y_1 + \partial\left(\widetilde{h}_x^1 - h_x^1\right)/\partial y_2 = 0$, (5.2) will possess a unique solution in the unknown functions.

Condition *(iii)* in Theorem 5.3 can be used to determine restrictions on $r_x^2/\left|r_y\right|$ guaranteeing that $r_{y_1}^1$ and $r_{y_2}^1$, or their ratio, are uniquely determined from $f_{Y|X=x}$. The identification results in Sections 3 and 4 assumed that $r_x^2/\left|r_y\right|$ is a constant function of x. Condition *(iii)* provides the means to develop new identification results when $r_x^2/\left|r_y\right|$ is not constant over x.

Condition *(iv)* provides conditions for identification of the elements of the first equation in Model (5.1), in terms of restrictions on the elements, $f_{\varepsilon_2|\varepsilon_1}$ and r^2, of the second equation in Model 5.1. The first equation is characterized by $F_{\varepsilon_1}\left(r^1\right)$, while the second, conditional on the value of ε_1, is characterized by $F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)$. To see how *(iv)* can be used to develop an estimator for the ratio $\left(r_{y_2}^1/r_{y_1}^1\right)$, we note that $f_{Y_1,Y_2|X=x}(y_1,y_2)$ can be expressed as

$$(5.4)\quad f_{Y_1,Y_2|X=x}(y_1,y_2)=\frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_1}\frac{\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)}{\partial y_2}-\frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_2}\frac{\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)}{\partial y_1}$$

Note also that $\partial F_{\varepsilon_1}\left(r^1\right)/\partial y_1$ and $\partial F_{\varepsilon_1}\left(r^1\right)/\partial y_2$ are both functions of only y_1 and y_2. Moreover,

$$\frac{\partial F_{\varepsilon_1}\left(r^1\right)/\partial y_1}{\partial F_{\varepsilon_1}\left(r^1\right)/\partial y_2}=\frac{r_{y_1}^1}{r_{y_2}^1},$$

which is the object of interest. Hence, identification of the derivative $\partial m^1\left(y_2,\varepsilon_1\right)/\partial y_2$ is established once the identification of $\partial F_{\varepsilon_1}\left(r^1\right)/\partial y_1$ and $\partial F_{\varepsilon_1}\left(r^1\right)/\partial y_2$ up to a common constant is achieved. As y_1, y_2 remain fixed and x varies, the value of these two functions is constant. Changes in x affect only the other two functions in (5.4), which are $\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)/\partial y_2$ and $\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)/\partial y_1$. Note that these two functions do not depend on any of the features of the function of interest, r^1. For particular fixed values of y_1 and y_2, the identification problem can then be expressed as the problem of finding coefficients a_1, a_2 satisfying the relationship

$$p\left(x\right)=a_1\phi_1(x)+a_2\phi_2(x)$$

where $p(x)$ is a known function. When $p(x)=f_{Y_1,Y_2|X=x}\left(y_1,y_2\right)$, the functions ϕ_1,ϕ_2 are $\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)/\partial y_2$ and $\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)/\partial y_1$. When $p(x)$ is the derivative of $f_{Y_1,Y_2|X=x}\left(y_1,y_2\right)$ with respect to x, the functions ϕ_1 and ϕ_2 are the derivatives of $\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)/\partial y_2$ and

$\partial F_{\varepsilon_2|\varepsilon_1=r^1}(r^2)/\partial y_1$ with respect to x. One can consider integration, differentiation, or any other transformation which which does not affect the values of the coefficients, a_1 and a_2, or their ratio. Clearly, one may impose many shape restrictions on $\phi_1(x)$ and $\phi_2(x)$ that guarantee identification. In particular, note that

$$\frac{\partial F_{\varepsilon_2|\varepsilon_1=r^1}(r^2)}{\partial y_2} = f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_{y_2} \quad and \quad \frac{\partial F_{\varepsilon_2|\varepsilon_1=r^1}(r^2)}{\partial y_1} = f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_{y_1}$$

Hence, when $p(x) = f_{Y_1,Y_2|X=x}(y_1,y_2)$ identification will depend on the way in which the values of $r^2_{y_2}$ and $r^2_{y_1}$ vary with x.

6. Conclusions

In this paper, we have introduced several new methods for estimation of nonparametric simultaneous equations models. We developed in detail two models. In our first model, each structural equation contained an exclusive regressor. We introduced for this model an estimator of the standard Least Squares form, $(X'X)^{-1}(X'Y)$, except that the elements of the matrices X and Y were constructed from average derivative estimators from the density of the observable variables. Our second model had one function of interest and one instrument. We introduced estimators for the derivative of the function of interest, which were expressed in terms of ratios of derivatives of the conditional density of the observed endogenous variables at particular estimated values of the instrument.

The estimators that we developed were special cases of new general approaches to estimation for models with simultaneity, which we presented in the paper. These approaches can be easily adapted to handle many other alternative models, satisfying different identifying assumptions. We have indicated directions in which alternative identified models can be found and how our estimation methods can be modified for such models.

7. Appendix

Proof of Theorem 3.2: We apply the Delta method in Newey (1994). Let $\digamma$ denote the set of densities satisfying Assumption 3.7, and let $\|g\|$ denote the sum of sup norms of g and its derivatives over $\overline{M}^y \times \overline{M}^x$. Define the functionals $\alpha_{y_j}(g)$ and $\beta_{x_s}(g)$ by $\alpha_{y_j}(g) = \partial \log g_{Y|X=x}(y)/\partial y_j$ and $\beta_{x_s}(g) = \partial \log g_{Y|X=x}(y)/\partial x_s$.Then,

$$\alpha_{y_j}(g) = \frac{\frac{\partial g_{Y,X}(y,x)}{\partial y_j}}{g_{Y,X}(y,x)} \text{ and } \beta_{x_s}(g) = \left(\frac{\frac{\partial g_{Y,X}(y,x)}{\partial x_s}}{g_{Y,X}(y,x)} - \frac{\frac{\partial g_X(x)}{\partial x_s}}{g_X(x)}\right).$$

To simplify notation, we will denote $f_{Y,X}(y,x)$ by f, $f_X(x)$ by $\widetilde{f}$, $\partial f_{Y,X}(y,x)/\partial y_j$ by f_{y_j}, $\partial f_{Y,X}(y,x)/\partial x_s$ by f_{x_s}, and $\partial f_X(x)/\partial x_s$ by $\widetilde{f}_{x_s}$, with similar shorthands for functions g and h. For any h such that $\|h\|$ is small enough,

$$\alpha_{y_j}(f+h) - \alpha_{y_j}(f)$$

$$= \left[\frac{f_{y_j}+h_{y_j}}{f+h} - \frac{f_{y_j}}{f}\right] = \left[\frac{\left[h_{y_j}f - f_{y_j}h\right]}{f^2} - \frac{h\left[h_{y_j}f - f_{y_j}h\right]}{f^2(f+h)}\right] \text{ and}$$

$$\beta_{x_s}(f+h) - \beta_{x_s}(f) = \left[\frac{f_{x_s}+h_{x_s}}{f+h} - \frac{f_{x_s}}{f}\right] - \left[\frac{\widetilde{f}_{x_s}+\widetilde{h}_{x_s}}{\widetilde{f}+\widetilde{h}} - \frac{\widetilde{f}_{x_s}}{\widetilde{f}}\right]$$

$$= \left[\frac{\left[h_{x_s}f - f_{x_s}h\right]}{f^2} - \frac{h\left[h_{x_s}f - f_{x_s}h\right]}{f^2(f+h)}\right] - \left[\frac{\left[\widetilde{h}_{x_s}\widetilde{f} - \widetilde{f}_{x_s}\widetilde{h}\right]}{\widetilde{f}^2} - \frac{\widetilde{h}\left[\widetilde{h}_{x_s}\widetilde{f} - \widetilde{f}_{x_s}\widetilde{h}\right]}{\widetilde{f}^2(\widetilde{f}+\widetilde{h})}\right].$$

$$\text{Define} \quad D\alpha_{y_j}(f;h) = \frac{\left[h_{y_j}f - f_{y_j}h\right]}{f^2} \quad ; \quad R\alpha_{y_j}(f;h) = -\frac{h\left[h_{y_j}f - f_{y_j}h\right]}{f^2(f+h)}$$

$$D\beta_{x_s}(f;h) = \left[\frac{\left[h_{x_s}f - f_{x_s}h\right]}{f^2} - \frac{\left[\widetilde{h}_{x_s}\widetilde{f} - \widetilde{f}_{x_s}\widetilde{h}\right]}{\widetilde{f}^2}\right] \quad ; \text{ and}$$

$$R\beta_{x_s}(f;h) = -\left[\frac{h\left[h_{x_s}f - f_{x_s}h\right]}{f^2(f+h)} - \frac{\widetilde{h}\left[\widetilde{h}_{x_s}\widetilde{f} - \widetilde{f}_{x_s}\widetilde{h}\right]}{\widetilde{f}^2(\widetilde{f}+\widetilde{h})}\right].$$

$$\begin{aligned} \text{Then,} \quad \alpha_{y_j}(f+h) - \alpha_{y_j}(f) &= D\alpha_{y_j}(f;h) + R\alpha_{y_j}(f;h), \\ \text{and} \quad \beta_{x_s}(f+h) - \beta_{x_s}(f) &= D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h). \end{aligned}$$

Denote $\mu(x)\,dx$ by μ, and define

$$\Phi_{y_j,x_s}(g) = \int \alpha_{y_j}(g)\ \beta_{x_s}(g)\ \mu\ - \left(\int \alpha_{y_j}(g)\ \mu\right)\left(\int \beta_{x_s}(g)\ \mu\right).$$

It is easy to verify that for all h such that $\|h\|$ is small enough

$$\Phi_{y_j,x_s}(f+h) - \Phi_{y_j,x_s}(f)$$

$$= \int \left(\alpha_{y_j}(f+h) - \alpha_{y_j}(f)\right)\left(\beta_{x_s}(f) - \int \beta_{x_s}(f)\mu(x)dx\right)\ \mu(x)\ dx$$

$$+ \int \left(\alpha_{y_j}(f) - \int \alpha_{y_j}(f)\mu(x)dx\right)\left(\beta_{x_s}(f+h) - \beta_{x_s}(f)\right)\mu(x)\ dx$$

$$+\int \left(\alpha_{y_j}(f+h)-\alpha_{y_j}(f)\right)\left(\beta_{x_s}(f+h)-\beta_{x_s}(f)\right)\mu(x)\ dx$$

$$-\left(\int \left(\alpha_{y_j}(f+h)-\alpha_{y_j}(f)\right)\mu(x)\ dx\right)\left(\int \left(\beta_{x_s}(f+h)-\beta_{x_s}(f)\right)\mu(x)\ dx\right).$$

Hence,

$$\begin{aligned}
&\Phi_{y_j,x_s}(f+h)-\Phi_{y_j,x_s}(f)\\
&=\int D\alpha_{y_j}(f;h)\left(\beta_{x_s}(f)-\int \beta_{x_s}(f)\mu(x)dx\right)\ \mu(x)\ dx\\
&+\int D\beta_{x_s}(f;h)\ \left(\alpha_{y_j}(f)-\int \alpha_{y_j}(f)\mu(x)dx\right)\ \mu(x)\ dx\\
&+\int R\alpha_{y_j}(f;h)\left(\beta_{x_s}(f)-\int \beta_{x_s}(f)\mu(x)dx\right)\ \mu(x)\ dx\\
&+\int R\beta_{x_s}(f;h)\ \left(\alpha_{y_j}(f)-\int \alpha_{y_j}(f)\mu(x)dx\right)\ \mu(x)\ dx\\
&+\int \left(D\alpha_{y_j}(f;h)+R\alpha_{y_j}(f;h)\right)\left(D\beta_{x_s}(f;h)+R\beta_{x_s}(f;h)\right)\mu(x)\ dx\\
&-\left(\int \left(D\alpha_{y_j}(f;h)+R\alpha_{y_j}(f;h)\right)\mu(x)\ dx\right)\left(\int \left(D\beta_{x_s}(f;h)+R\beta_{x_s}(f;h)\right)\mu(x)\ dx\right)
\end{aligned}$$

Denote the first two terms in this last sum by $D\Phi_{y_j,x_s}(f;h)$ and the last four terms by $R\Phi_{y_j,x_s}(f;h)$. Our assumptions imply that for some $a<\infty$,

$$\left|D\Phi_{y_j,x_s}(f;h)\right|\le a\,\|h\| \ \text{ and } \ \left|R\Phi_{y_j,x_s}(f;h)\right|\le a\,\|h\|^2.$$

Expanding the first term in the sum, we get

$$\begin{aligned}
(T.1)\quad &\int D\alpha_{y_j}(f;h)\left(\beta_{x_s}(f)-\int \beta_{x_s}(f)\mu(x)dx\right)\ \mu(x)\ dx\\
= \quad &\int h_{y_j}\left[\frac{\mu(x)\left(\beta_{x_s}(f)-\int \beta_{x_s}(f)\mu(x)dx\right)}{f}\right]\ dx\\
&-\int h\left[\frac{f_{y_j}\mu(x)\left(\beta_{x_s}(f)-\int \beta_{x_s}(f)\mu(x)dx\right)}{f^2}\right]dx.
\end{aligned}$$

Expanding the second term in the sum, we get

$$(T.2)\quad \int D\beta_{x_s}(f;h)\ \left(\alpha_{y_j}(f)-\int \alpha_{y_j}(f)\mu(x)dx\right)\ \mu(x)\ dx$$

$$= \int \frac{[h_{x_s} f - f_{x_s} h]}{f^2} \frac{f_{y_j}}{f} \mu(x)\ dx - \int \frac{\left[\widetilde{h}_{x_s} \widetilde{f} - \widetilde{f}_{x_s} \widetilde{h}\right]}{\widetilde{f}^2} \frac{f_{y_j}}{f}\ \mu(x)\ dx$$

$$- \left(\int \frac{[h_{x_s} f - f_{x_s} h]}{f^2} \mu(x)\ dx \right) \left(\int \left(\frac{f_{y_j}}{f} \right) \mu(x) dx \right)$$

$$+ \left(\int \frac{\left[\widetilde{h}_{x_s} \widetilde{f} - \widetilde{f}_{x_s} \widetilde{h}\right]}{\widetilde{f}^2} \mu(x) dx \right) \left(\int \left(\frac{f_{y_j}}{f} \right) \mu(x) dx \right)$$

Our assumptions imply that $[h\ \mu]/f$, $\left[h\ \mu\ f_{y_j}\right]/f^2$, $\left[\widetilde{h}\ \mu\right]/\widetilde{f}$ and $\left[\widetilde{h}\ \mu\ f_{y_j}\right]/\left[\widetilde{f} f\right]$ vanish on the boundary of the integration. Hence, integration by parts of the terms in $(T.2)$ containing h_{x_s} or $\widetilde{h}_{x_s}$ gives

$$\int D\beta_{x_s}(f; h)\ \left(\alpha_{y_j}(f) - \int \alpha_{y_j}(f) \mu(x) dx \right)\ \mu(x)\ dx$$

$$= - \int h\ \left[\frac{\partial}{\partial x_s} \left(\frac{\mu\ f_{y_j}}{f^2} \right) + \frac{f_{x_s} f_{y_j} \mu}{f^3} \right]\ dx + \int \widetilde{h}\ \left[\frac{\partial}{\partial x_s} \left(\frac{\mu\ f_{y_j}}{\widetilde{f}\ f} \right) + \frac{\widetilde{f}_{x_s} f_{y_j} \mu}{\widetilde{f}^2 f} \right]\ dx$$

$$+ \left(\int h \left[\frac{\partial}{\partial x_s} \left(\frac{\mu}{f} \right) + \frac{f_{x_s} \mu}{f^2} \right]\ dx \right)\ \left(\int \left(\frac{f_{y_j}}{f} \right) \mu(x) dx \right)$$

$$- \left(\int \widetilde{h} \left[\frac{\partial}{\partial x_s} \left(\frac{\mu}{\widetilde{f}} \right) + \frac{\widetilde{f}_{x_s} \mu}{\widetilde{f}^2} \right] dx \right) \left(\int \left(\frac{f_{y_j}}{f} \right) \mu(x) dx \right).$$

Letting $h = \widehat{f} - f$, it follows by our assumptions and standard kernel methods (see, e.g., Newey (1994)) that

$$\sqrt{N \sigma_N^{G+2}} \left[\left(\int D\beta_{x_s}(f; \widehat{f} - f)\ \left(\alpha_{y_j}(f) - \int \alpha_{y_j}(f) \mu(x) dx \right)\ \mu(x) \right) + R\Phi_{y_j, x_s} \left(f; \widehat{f} - f \right) \right] \xrightarrow{p} 0$$

and also for the second term in $(T.1)$,

$$\sqrt{N \sigma_N^{G+2}} \left(\int h \left[\frac{f_{y_j} \mu(x) \left(\beta_{x_s}(f) - \int \beta_{x_s}(f) \mu(x) dx \right)}{f^2} \right] dx \right) \xrightarrow{p} 0$$

Hence, by the definition of Φ, it follows that

$$\sqrt{N \sigma_N^{G+2}} \left(\Phi_{y_j, x_s} \left(\widehat{f} \right) - \Phi_{y_j, x_s} (f) \right)$$

$$= \sqrt{N \sigma_N^{G+2}} \left(\int D\alpha_{y_j}(f; \widehat{f} - f) \left(\beta_{x_s}(f) - \int \beta_{x_s}(f) \mu(x) dx \right)\ \mu(x)\ dx \right) + o_p(1)$$

$$= \sqrt{N\sigma_N^{G+2}} \left(\int \left(\frac{\partial \widehat{f}_{Y,X}(y,x)}{\partial y_j} - \frac{\partial f_{Y,X}(y,x)}{\partial y_j} \right) \left(\frac{\left(\Delta \partial_{x_s} \log f_{Y|X=x}(y)\right) \ \mu(x)}{f_{Y,X}(y,x)} \right) dx \right) + o_p(1)$$

$$\text{where} \quad \Delta \partial_{x_s} \log f_{Y|X=x}(y) = \frac{\partial \log f_{Y|X=x}(y)}{\partial x_s} - \int \frac{\partial \log f_{Y|X=x}(y)}{\partial x_s} \mu(x) dx$$

Using the notation of Section 3, and applying Newey (1994), this implies that for $V(y)$ as defined in Section 3

$$(T.3) \qquad \sqrt{N\sigma_N^{G+2}} \left(\widehat{TT}_{yx}(y) - TT_{yx}(y) \right) \underset{d}{\to} N\left(0, V(y)\right).$$

Define next the functional $\Upsilon_{x_j,x_s}(g)$ by

$$\Upsilon_{x_j,x_s}(g) = \int \beta_{x_j}(g)\ \beta_{x_s}(g)\ \mu \ - \left(\int \beta_{x_j}(g)\ \mu \right) \left(\int \beta_{x_s}(g)\ \mu \right).$$

Using arguments very similar to the above, we can conclude that under our assumptions,

$$\Upsilon_{x_j,x_s}(\widehat{f}) - \Upsilon_{x_j,x_s}(f) = D\Upsilon_{x_j,x_s}\left(f; \widehat{f} - f\right) + R\Upsilon_{x_j,x_s}\left(f; \widehat{f} - f\right)$$

$$\text{where} \quad D\Upsilon_{x_j,x_s}\left(f; \widehat{f} - f\right)$$

$$= \int \left[D\beta_{x_j}(f; \widehat{f} - f) \left(\Delta \partial_{x_s} \log f_{Y|X=x}(y)\right) + D\beta_{x_s}(f; \widehat{f} - f) \left(\Delta \partial_{x_j} \log f_{Y|X=x}(y)\right) \right] \ \mu(x)\ dx$$

and for some $b < \infty$,

$$\left| D\Upsilon_{x_j,x_s}\left(f; \widehat{f} - f\right) \right| \leq b \left\| \widehat{f} - f \right\| \quad \text{and} \quad \left| R\Upsilon_{x_j,x_s}\left(f; \widehat{f} - f\right) \right| \leq b \left\| \widehat{f} - f \right\|^2$$

This implies, under our assumptions that $\widehat{TT}_{xx}(y) \underset{p}{\to} TT_{xx}(y)$. The result of the theorem then follows from this, $(T.3)$, Slutsky's Theorem, and the definition of $\widehat{rr}_y(y)$.

Proof of Theorem 4.3: Let F denote the set of densities g that satisfy Assumption 4.7. Let $\|g\|$ denote the maximum of the supremum of the values and derivatives up to the fourth order of g over the compact set defined as the closure of the neighborhood defined in Assumption 4.8. We apply again Newey (1994). For this, we first analyze the functional that for any g, assigns the value of x at which $\partial^2 \log g_{Y|X=x}(y)/\partial x^2 = 0$. Define the functional $\Phi(g,x)$ by $\Phi(g,x) = \partial^2 \log g_{Y|X=x}(y)/\partial x^2$. We will show that there exists a functional $\kappa(g)$ on a neighborhood of f which is defined implicitly by $\Phi(g, \kappa(g)) = 0$ and satisfies a Taylor expansion of the form $\kappa(f+h) = \kappa(f) + D\kappa(f;h) + R\kappa(f;h)$ with $|R\kappa(f;h)|$ of the order $\|h\|^2$. We then use this to analyze the functional defining our estimator. We will denote

$g_{Y,X}(y,x)$ by $g(x)$ and $g_X(x)$ by $\widetilde{g}(x)$, with similar notation for other functions in $\mathcal{F}$. For any h such that $\|h\|$ is small enough, any x in a neighborhood of x^*, and any small enough δ such that $|\delta| > 0$,

$$\Phi(g+h,x) - \Phi(g,x)$$

$$= \frac{\frac{\partial^2 g(x)}{\partial x^2} + \frac{\partial^2 h(x)}{\partial x^2}}{g(x)+h(x)} - \frac{\left[\frac{\partial g(x)}{\partial x} + \frac{\partial h(x)}{\partial x}\right]^2}{[g(x)+h(x)]^2} - \frac{\frac{\partial^2 \widetilde{g}(x)}{\partial x^2} + \frac{\partial^2 \widetilde{h}(x)}{\partial x^2}}{\widetilde{g}(x)+\widetilde{h}(x)} + \frac{\left[\frac{\partial^2 \widetilde{g}(x)}{\partial x^2} + \frac{\partial^2 \widetilde{h}(x)}{\partial x^2}\right]^2}{\left[\widetilde{g}(x)+\widetilde{h}(x)\right]^2}$$

$$- \left(\frac{\frac{\partial^2 g(x)}{\partial x^2}}{g(x)} - \frac{\left[\frac{\partial g(x)}{\partial x}\right]^2}{[g(x)]^2} - \frac{\frac{\partial^2 \widetilde{g}(x)}{\partial x^2}}{\widetilde{g}(x)} + \frac{\left[\frac{\partial^2 \widetilde{g}(x)}{\partial x^2}\right]^2}{[\widetilde{g}(x)]^2} \right) \textit{ and}$$

$$\Phi(g,x+\delta) - \Phi(g,x)$$

$$= \frac{\frac{\partial^2 g(x+\delta)}{\partial x^2}}{g(x+\delta)} - \frac{\left[\frac{\partial g(x+\delta)}{\partial x}\right]^2}{[g(x+\delta)]^2} - \frac{\frac{\partial^2 \widetilde{g}(x+\delta)}{\partial x^2}}{\widetilde{g}(x+\delta)} + \frac{\left[\frac{\partial^2 \widetilde{g}(x+\delta)}{\partial x^2}\right]^2}{[\widetilde{g}(x+\delta)]^2}$$

$$- \left(\frac{\frac{\partial^2 g(x)}{\partial x^2}}{g(x)} - \frac{\left[\frac{\partial g(x)}{\partial x}\right]^2}{[g(x)]^2} - \frac{\frac{\partial^2 \widetilde{g}(x)}{\partial x^2}}{\widetilde{g}(x)} + \frac{\left[\frac{\partial^2 \widetilde{g}(x)}{\partial x^2}\right]^2}{[\widetilde{g}(x)]^2} \right)$$

Define

$$D_g\Phi(g,x;h) = \frac{\frac{\partial^2 h(x)}{\partial x^2} g(x)^2 - \frac{\partial^2 g(x)}{\partial x^2} h(x) g(x) - 2\frac{\partial h(x)}{\partial x}\frac{\partial g(x)}{\partial x} g(x) + 2\left(\frac{\partial g(x)}{\partial x}\right)^2 h(x)}{[g(x)]^3}$$

$$- \frac{\frac{\partial^2 \widetilde{h}(x)}{\partial x^2} \widetilde{g}(x)^2 - \frac{\partial^2 \widetilde{g}(x)}{\partial x^2} \widetilde{h}(x) \widetilde{g}(x) - 2\frac{\partial \widetilde{h}(x)}{\partial x}\frac{\partial \widetilde{g}(x)}{\partial x} \widetilde{g}(x) + 2\left(\frac{\partial \widetilde{g}(x)}{\partial x}\right)^2 \widetilde{h}(x)}{[\widetilde{g}(x)]^3}, \text{ and}$$

$$D_x\Phi(g,x;\delta) = \frac{\partial^3 \log g_{Y|X=x}(y)}{\partial x^3}\delta$$

$$R_f\Phi(g,x;h) = \Phi(g+h,x) - \Phi(g,x) - D_g\Phi(g,x;h), \text{ and}$$
$$R_x\Phi(g,x;\delta) = \Phi(g,x+\delta) - \Phi(g,x) - D_x\Phi(g,x;\delta).$$

Our assumptions imply that there exists $a < \infty$ such that for all (g,x) in a neighborhood of (f,x^*),

$$\|D_x\Phi(g,x;\delta)\| \leq a|\delta|; \quad \|R_x\Phi(g,x;\delta)\| \leq a|\delta|^2;$$
$$\|D_g\Phi(g,x;h)\| \leq a\|h\|; \quad \textit{and} \quad \|R_g\Phi(g,x;h)\| \leq a\|h\|^2$$

Moreover, it can be verified that on a neighborhood of (f,x^*), $D_x\Phi(g,x;\delta)$ and $D_g\Phi(g,x;h)$ are also Fréchet differentiable on (g,x) and their derivatives are continuous on (g,x). By our assumptions, for all (g,x) in a neighborhood of (f,x^*), $D_x\Phi(g,x;\delta)$ is invertible. It then follows by the Implicit Function Theorem on Banach spaces that there exists a unique functional κ such that for all g in a neighborhood of f

$$\Phi(f,\kappa(f))=0$$

The Fréchet derivative at g is given by

$$D\kappa(g;h)=\left(\frac{\partial^3\log g_{Y|X=x}(y)}{\partial x^3}\right)^{-1}\left[-\,D_g\Phi(g,x;h)\right]$$

Since Φ is a C^2 map on a neighborhood of (f,x^*) and its second order derivatives are uniformly bounded on such neighborhood, κ is a C^2 map with uniformly bounded second derivatives on a neighborhood of f. Hence, by Taylor's Theorem on Banach spaces, it follows that there exists $c<\infty$ such that for sufficiently small $\|h\|$, $|\kappa(f+h)-\kappa(f)-D\kappa(f;h)|\leq c\,\|h\|^2$.

We now analyze the functional of f that defines our estimator. This functional uses κ as an input. Define the functional $\Psi(g,\kappa(g))$ by

$$\Psi(g,\kappa(g))=\frac{-\left[\frac{\partial^2 g_{Y,X}(y,\kappa(g))}{\partial y_2\partial x}\,g_{Y,X}(y,\kappa(g))-\frac{\partial g_{Y,X}(y,\kappa(g))}{\partial y_2}\frac{\partial g_{Y,X}(y,\kappa(g))}{\partial x}\right]}{\left[\frac{\partial^2 g_{Y,X}(y,\kappa(g))}{\partial y_1\partial x}\,g_{Y,X}(y,\kappa(g))-\frac{\partial g_{Y,X}(y,\kappa(g))}{\partial y_1}\frac{\partial g_{Y,X}(y,\kappa(g))}{\partial x}\right]}$$

Then, $\Psi(f,\kappa(f))=\partial m^1(y_2,\varepsilon_1)/\partial y_2$ and $\Psi\left(\widehat{f},\kappa(\widehat{f})\right)=\partial m^1\widehat{(y_2,\varepsilon_1)}/\partial y_2$. For h and δ such that $\|h\|$ and $|\delta|$ are small enough, define

$$D_g\Psi(g,x^*;h)$$

$$=\frac{-\left[\frac{\partial^2 h(x^*)}{\partial y_2\partial x}\,g(x^*)+\frac{\partial^2 g(x^*)}{\partial y_2\partial x}\,h(x^*)-\frac{\partial h(x^*)}{\partial y_2}\frac{\partial g(x^*)}{\partial x}-\frac{\partial g(x^*)}{\partial y_2}\frac{\partial h(x^*)}{\partial x}\right]}{\left[\frac{\partial^2 g(x^*)}{\partial y_1\partial x}g(x^*)-\frac{\partial g(x^*)}{\partial y_1}\frac{\partial g(x^*)}{\partial x}\right]}$$

$$+\frac{\left[\frac{\partial^2 g(x^*)}{\partial y_2\partial x}\,g(x^*)-\frac{\partial g(x^*)}{\partial y_2}\frac{\partial g(x^*)}{\partial x}\right]\left[\frac{\partial^2 h(x^*)}{\partial y_1\partial x}g(x^*)+\frac{\partial^2 g(x^*)}{\partial y_1\partial x}h(x^*)-\frac{\partial h(x^*)}{\partial y_1}\frac{\partial g(x^*)}{\partial x}-\frac{\partial g(x^*)}{\partial y_1}\frac{\partial h(x^*)}{\partial x}\right]}{\left[\frac{\partial^2 g(x^*)}{\partial y_1\partial x}g(x^*)-\frac{\partial g(x^*)}{\partial y_1}\frac{\partial g(x^*)}{\partial x}\right]^2}$$

$$D_x\Psi(g,x^*;\delta)=\frac{\partial\left(\frac{-\,\partial^2\log g_{Y|X=x^*}(y)/\partial y_2\partial x}{\partial^2\log g_{Y|X=x^*}(y)/\partial y_1\partial x}\right)}{\partial x}\delta$$

Then,

$$\begin{aligned} D\Psi\left(f,\kappa(f);h\right) &= D_f\Psi\left(f,x^*;h\right)+D_x\Psi\left(f,x^*;D\kappa(f;h)\right) \quad \text{and} \\ R\Psi\left(f,\kappa(f);h\right) &= \Psi\left(f+h,\kappa(f+h)\right)-\Psi\left(f,\kappa(f)\right)-D\Psi\left(f,\kappa(f);h\right) \end{aligned}$$

The properties we derived on $D\kappa$ and $R\kappa$ and our assumptions imply that for some $b<\infty$, $|D\Psi\left(f,\kappa(f);h\right)|\leq b\left\|h\right\|$ and $|R\Psi\left(f,\kappa(f);h\right)|\leq b\left\|h\right\|^2$. By standard kernel methods and our assumptions it follows that when $h=\widehat{f}-f$,

$$\begin{aligned} &\sqrt{N\sigma_N^7}\; D\Psi\left(f,\kappa(f);h\right) \\ &= \sqrt{N\sigma_N^7}\frac{-\;[\;f(x^*)]}{\left[\frac{\partial^2 f(x^*)}{\partial y_1\partial x}f(x^*)-\frac{\partial f(x^*)}{\partial y_1}\frac{\partial f(x^*)}{\partial x}\right]}\frac{\partial^2 h\left(x^*\right)}{\partial y_2\partial x} \\ &+\sqrt{N\sigma_N^7}\frac{\left[\frac{\partial^2 f(x^*)}{\partial y_2\partial x}\;f(x^*)-\frac{\partial f(x^*)}{\partial y_2}\frac{\partial f(x^*)}{\partial x}\right]\left[f(x^*)\right]}{\left[\frac{\partial^2 f(x^*)}{\partial y_1\partial x}f(x^*)-\frac{\partial f(x^*)}{\partial y_1}\frac{\partial f(x^*)}{\partial x}\right]^2}\frac{\partial^2 h\left(x^*\right)}{\partial y_1\partial x} \\ &+\sqrt{N\sigma_N^7}\frac{\partial\left(\frac{-\;\partial^2\log f_{Y|X=x^*}(y)/\partial y_2\partial x}{\partial^2\log f_{Y|X=x^*}(y)/\partial y_1\partial x}\right)}{\partial x}\left(\frac{\partial^3\log f_{Y|X=x^*}(y)}{\partial x^3}\right)^{-1}\left[\frac{-1}{f(x^*)}\right]\frac{\partial^2 h(x^*)}{\partial x^2}+o_p(1) \end{aligned}$$

Hence, when $h=\widehat{f}-f$,

$$\begin{aligned} &\sqrt{N\sigma_N^7}\; D\Psi\left(f,\kappa(f);h\right) \\ &= \sqrt{N\sigma_N^7}\frac{\frac{\partial^2\log f_{Y|X=x^*}(y)}{\partial y_2\partial x}}{\left[\frac{\partial^2\log f_{Y|X=x^*}(y)}{\partial y_1\partial x}\right]^2 f_{Y,X}(y,x^*)}\frac{\partial^2 h\left(x^*\right)}{\partial y_1\partial x} \\ &+\sqrt{N\sigma_N^7}\frac{-\;1}{\frac{\partial^2\log f_{Y|X=x^*}(y)}{\partial y_1\partial x}f_{Y,X}(y,x^*)}\frac{\partial^2 h\left(x^*\right)}{\partial y_2\partial x} \\ &+\sqrt{N\sigma_N^7}\frac{\frac{-\;\partial\left(\frac{-\;\partial^2\log f_{Y|X=x^*}(y)/\partial y_2\partial x}{\partial^2\log f_{Y|X=x^*}(y)/\partial y_1\partial x}\right)}{\partial x}}{f_{Y,X}(y,x^*)\left(\frac{\partial^3\log f_{Y|X=x^*}(y)}{\partial x^3}\right)}\frac{\partial^2 h(x^*)}{\partial x^2}+o_p(1) \end{aligned}$$

Standard results for kernel estimators imply then that

$$\sqrt{N\sigma_N^7}\; D\Psi\left(f,\kappa(f);\widehat{f}-f\right)\underset{d}{\rightarrow}N(0,\widetilde{V})$$

where $\widetilde{V}$ is as defined prior to the statement of Theorem 4.3. Our assumptions imply that $\sqrt{N\sigma_N^7}R\Psi\left(f,\kappa(f);\widehat{f}-f\right)=o_p(1)$. Hence,

$$\sqrt{N\sigma_N^7}\left[\partial m^1\left(\widehat{y_2,\varepsilon_1}\right)/\partial y_2 - \partial m^1\left(y_2,\varepsilon_1\right)/\partial y_2\right]$$

$$= \sqrt{N\sigma_N^7}\left[\Psi\left(\widehat{f},\kappa(\widehat{f})\right) - \Psi\left(f,\kappa(f)\right)\right]$$

$$= \sqrt{N\sigma_N^7}D\Psi\left(f,\kappa(f);\widehat{f}-f\right) + o_p(1) \quad \underset{d}{\rightarrow} \quad N\left(0,\widetilde{V}\right)$$

Proof of Theorem 4.4: The proof is similar to the proof of Theorem 4.3. Let $\mathcal{F}$ denote the set of densities g that satisfy Assumption 4.7'. Let $\|g\|$ denote the maximum of the supremum of the values and derivatives up to the third order of g over a compact set that is defined by the union of the closures of the neighborhoods defined in Assumption 4.8'. We first analyze the functionals that for any g assign values x_1 and x_2, at which $\partial \log g_{Y|X=x_1}(y)/\partial x = 0$ and $\partial \log g_{Y|X=x_2}(y)/\partial x = 0$. As in the proof of Theorem 4.3, we will denote $g_{Y,X}(y,x)$ by $g(x)$ and $g_X(x)$ by $\widetilde{g}(x)$, with similar notation for other functions in $\mathcal{F}$. Since $x_1 \neq x_2$, the asymptotic covariance of our kernel estimators for the values of x_1 and x_2 is zero. Define the functional $\Phi\left(g,x_1,x_2\right) = \left(\partial \log g_{Y|X=x_1}(y)/\partial x,\ \partial \log g_{Y|X=x_2}(y)/\partial x\right)'$. We first show that there exists a functional $\kappa(g) = (\kappa^1(g),\kappa^2\left(g\right))$ satisfying $\kappa(f) = (x_1^*,x_2^*)$ which is defined implicitly in a neighborhood of f by

$$\Phi\left(g,\kappa^1(g),\kappa^2(g)\right) = 0.$$

Denote by x_1 any value of x in a small enough neighborhood of x_1^* and denote by x_2 any value of x in a small enough neighborhood of x_2^*. Let g denote a density in a small enough neighborhood of f. For any h such that $\|h\|$ is small enough, and any $\left(\delta_1,\delta_2\right)$, such that $|\delta_1|$ and $|\delta_2|$ are small enough

$$\Phi\left(g+h,x_1,x_2\right) - \Phi\left(g,x_1,x_2\right) = \begin{pmatrix} \frac{\frac{\partial g(x_1)}{\partial x}+\frac{\partial h(x_1)}{\partial x}}{g(x_1)+h(x_1)} - \frac{\frac{\partial \widetilde{g}(x_1)}{\partial x}+\frac{\partial \widetilde{h}(x_1)}{\partial x}}{\widetilde{g}(x_1)+\widetilde{h}(x_1)} - \frac{\frac{\partial g(x_1)}{\partial x}}{g(x_1)} + \frac{\frac{\partial \widetilde{g}(x_1)}{\partial x}}{\widetilde{g}(x_1)} \\ \\ \frac{\frac{\partial g(x_2)}{\partial x}+\frac{\partial h(x_2)}{\partial x}}{g(x_2)+h(x_2)} - \frac{\frac{\partial \widetilde{g}(x_2)}{\partial x}+\frac{\partial \widetilde{h}(x_2)}{\partial x}}{\widetilde{g}(x_2)+\widetilde{h}(x_2)} - \frac{\frac{\partial g(x_2)}{\partial x}}{g(x_2)} + \frac{\frac{\partial \widetilde{g}(x_2)}{\partial x}}{\widetilde{g}(x_2)} \end{pmatrix}$$

$$\begin{aligned} \Phi\left(g,x_1+\delta_1,x_2\right) - \Phi\left(g,x_1,x_2\right) &= \left(\frac{\frac{\partial g(x_1+\delta_1)}{\partial x}}{g(x_1+\delta_1)} - \frac{\frac{\partial \widetilde{g}(x_1+\delta_1)}{\partial x}}{\widetilde{g}(x_1+\delta_1)} - \frac{\frac{\partial g(x_1)}{\partial x}}{g(x_1)} + \frac{\frac{\partial \widetilde{g}(x_1)}{\partial x}}{\widetilde{g}(x_1)} ,0\right)' \text{ and} \\ \Phi\left(g,x_1,x_2+\delta_2\right) - \Phi\left(g,x_1,x_2\right) &= \left(0, \frac{\frac{\partial g(x_2+\delta_2)}{\partial x}}{g(x_2+\delta_2)} - \frac{\frac{\partial \widetilde{g}(x_2+\delta_2)}{\partial x}}{\widetilde{g}(x_2+\delta_2)} - \frac{\frac{\partial g(x_2)}{\partial x}}{g(x_2)} + \frac{\frac{\partial \widetilde{g}(x_2)}{\partial x}}{\widetilde{g}(x_2)} \right)' \end{aligned}$$

Define

$$D_g\Phi\left(g,x_1,x_2;h\right)=\begin{pmatrix}\frac{\frac{\partial h(x_1)}{\partial x}}{g(x_1)}-\frac{\frac{\partial g(x_1)}{\partial x}h(x_1)}{(g(x_1))^2}-\frac{\frac{\partial \widetilde{h}(x_1)}{\partial x}}{\widetilde{g}(x_1)}+\frac{\frac{\partial \widetilde{g}(x_1)}{\partial x}\widetilde{h}(x_1)}{(\widetilde{g}(x_1))^2}\\ \\ \frac{\frac{\partial h(x_2)}{\partial x}}{g(x_2)}-\frac{\frac{\partial g(x_2)}{\partial x}h(x_2)}{(g(x_2))^2}-\frac{\frac{\partial \widetilde{h}(x_2)}{\partial x}}{\widetilde{g}(x_2)}+\frac{\frac{\partial \widetilde{g}(x_2)}{\partial x}\widetilde{h}(x_2)}{(\widetilde{g}(x_2))^2}\end{pmatrix}$$

$$D_{x_1}\Phi\left(g,x_1,x_2;\delta_1\right)=\frac{\partial^2\log g_{Y|X=x_1}\left(y\right)}{\partial x^2}\delta_1;\quad D_{x_2}\Phi\left(g,x_1,x_2;\delta_2\right)=\frac{\partial^2\log g_{Y|X=x_2}\left(y\right)}{\partial x^2}\delta_2$$

$$R_f\Phi\left(g,x_1,x_2\right)=\Phi\left(g+h,x_1,x_2\right)-\Phi\left(g,x_1,x_2\right)-D_g\Phi\left(g,x_1,x_2;h\right),$$

$$R_{x_1}\Phi\left(g,x_1,x_2;\delta_1\right)=\Phi\left(g,x_1+\delta_1,x_2\right)-\Phi\left(g,x_1,x_2\right)-D_{x_1}\Phi\left(g,x_1,x_2;\delta_1\right),\text{ and}$$

$$R_{x_2}\Phi\left(g,x_1,x_2;\delta_2\right)=\Phi\left(g,x_1,x_2+\delta_2\right)-\Phi\left(g,x_1,x_2\right)-D_{x_2}\Phi\left(g,x_1,x_2;\delta_2\right).$$

Our assumptions imply that for some $a<\infty$,

$$\left\|D_{x_j}\Phi\left(g,x_1,x_2;\delta_j\right)\right\|\leq a\left|\delta_j\right|\text{ and }\left\|R_{x_j}\Phi\left(g,x_1,x_2;\delta_j\right)\right\|\leq a\left|\delta_j\right|^2\qquad(\text{for }j=1,2)$$

$$\left\|D_g\Phi\left(g,x_1,x_2;h\right)\right\|\leq a\left\|h\right\|;\quad and\quad\left\|R_f\Phi\left(g,x_1,x_2;h\right)\right\|\leq a\left\|h\right\|^2$$

Hence, $D_x\Phi\left(g,x_1,x_2;\delta\right)$ is the Fréchet derivative of Φ with respect to x and $D_g\Phi\left(g,x_1,x_2;h\right)$ is the Fréchet derivative of Φ with respect to g. By their definitions and our assumptions, it follows that both Fréchet derivatives are themselves Fréchet differentiable and their derivatives are continuous and uniformly bounded on a neighborhood of $\left(f,x_1^*,x_2^*\right)$. Moreover, again by our assumptions, each $D_{x_j}\Phi(g,x_1,x_2;\delta_j)$ $(j=1,2)$ has a continuous inverse on a neighborhood of $\Phi\left(f,x_1^*,x_2^*\right)$. It then follows by the Implicit Function Theorem on Banach spaces that there exist unique functionals κ^1 and κ^2 such that $\kappa^1\left(f\right)=x_1^*$, $\kappa^2\left(f\right)=x_2^*$, for all g in a neighborhood of f

$$\Phi\left(g,\kappa^1\left(g\right),\kappa^2\left(g\right)\right)=0$$

κ^1 and κ^2 are differentiable on a neighborhood of f and their Fréchet derivatives are given by, for j=1,2

$$D\kappa^j\left(g;h\right)=\left(\frac{\partial^2\log g_{Y|X=x_j}(y)}{\partial x^2}\right)^{-1}\left[-\,D_g\Phi\left(g,x_1,x_2;h\right)\right]_j$$

Moreover, κ^1 and κ^2 satisfy a First order Taylor expansion around f with remainder term for $\kappa^j\left(f+h\right)-\kappa^j\left(f\right)$ bounded by $\left\|h\right\|^2$. Define the functional $\Psi(g,x_1,x_2)$ by

$$\Psi\left(g,x_1,x_2\right)=\left[\frac{\frac{\partial g_{Y,X}(y,x_2)}{\partial y_2}}{g_{Y,X}\left(y,x_2\right)}-\frac{\frac{\partial g_{Y,X}(y,x_1)}{\partial y_2}}{g_{Y,X}\left(y,x_1\right)}\right]\left[\frac{\frac{\partial g_{Y,X}(y,x_1)}{\partial y_1}}{g_{Y,X}\left(y,x_1\right)}-\frac{\frac{\partial g_{Y,X}(y,x_2)}{\partial y_1}}{g_{Y,X}\left(y,x_2\right)}\right]^{-1}$$

Then, $\Psi\left(\widehat{f},\kappa^{1}\left(\widehat{f}\right),\kappa^{2}\left(\widehat{f}\right)\right)=\partial m^{1}\widehat{\left(y_{2},\varepsilon_{1}\right)}/\partial y_{2}$ and $\Psi\left(f,\kappa^{1}\left(f\right),\kappa^{2}\left(f\right)\right)=\partial m^{1}\left(y_{2},\varepsilon_{1}\right)/\partial y_{2}$. Denote $f_{Y,X}\left(y,x_{j}\right)$ by $f\left(x_{j}\right)$ and $h_{Y,X}\left(y,x_{j}\right)$ by $h\left(x_{j}\right)$ *(j=1,2)*. Then, for $\left\Vert h\right\Vert$, $\left\vert\delta_{1}\right\vert$, and $\left\vert\delta_{2}\right\vert$ sufficiently small,

$$\Psi\left(f+h,x_{1}^{*},x_{2}^{*}\right)-\Psi\left(f,x_{1}^{*},x_{2}^{*}\right)$$

$$=\frac{\left[\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{2}}+\frac{\partial h\left(x_{2}^{*}\right)}{\partial y_{2}}\right]\left[f\left(x_{1}^{*}\right)+h\left(x_{1}^{*}\right)\right]-\left[\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{2}}+\frac{\partial h\left(x_{1}^{*}\right)}{\partial y_{2}}\right]\left[f\left(x_{2}^{*}\right)+h\left(x_{2}^{*}\right)\right]}{\left[\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{1}}+\frac{\partial h\left(x_{1}^{*}\right)}{\partial y_{1}}\right]\left[f\left(x_{2}^{*}\right)+h\left(x_{2}^{*}\right)\right]-\left[\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{1}}+\frac{\partial h\left(x_{2}^{*}\right)}{\partial y_{1}}\right]\left[f\left(x_{1}^{*}\right)+h\left(x_{1}^{*}\right)\right]}$$

$$-\frac{\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{2}}f\left(x_{1}^{*}\right)-\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{2}}f\left(x_{2}^{*}\right)}{\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{1}}f\left(x_{2}^{*}\right)-\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{1}}f\left(x_{1}^{*}\right)}$$

$$\Psi\left(f,x_{1}^{*}+\delta_{1},x_{2}^{*}\right)-\Psi\left(f,x_{1}^{*},x_{2}^{*}\right)$$

$$=\frac{\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{2}}f\left(x_{1}^{*}+\delta_{1}\right)-\frac{\partial f\left(x_{1}^{*}+\delta_{1}\right)}{\partial y_{2}}f\left(x_{2}^{*}\right)}{\frac{\partial f\left(x_{1}^{*}+\delta_{1}\right)}{\partial y_{1}}f\left(x_{2}^{*}\right)-\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{1}}f\left(x_{1}^{*}+\delta_{1}\right)}-\frac{\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{2}}f\left(x_{1}^{*}\right)-\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{2}}f\left(x_{2}^{*}\right)}{\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{1}}f\left(x_{2}^{*}\right)-\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{1}}f\left(x_{1}^{*}\right)},\text{ and}$$

$$\Psi\left(f,x_{1}^{*},x_{2}^{*}+\delta_{2}\right)-\Psi\left(f,x_{1}^{*},x_{2}^{*}\right)$$

$$=\frac{\frac{\partial f\left(x_{2}^{*}+\delta_{2}\right)}{\partial y_{2}}f\left(x_{1}^{*}\right)-\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{2}}f\left(x_{2}^{*}+\delta_{2}\right)}{\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{1}}f\left(x_{2}^{*}+\delta_{2}\right)-\frac{\partial f\left(x_{2}^{*}+\delta_{2}\right)}{\partial y_{1}}f\left(x_{1}^{*}\right)}-\frac{\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{2}}f\left(x_{1}^{*}\right)-\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{2}}f\left(x_{2}^{*}\right)}{\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{1}}f\left(x_{2}^{*}\right)-\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{1}}f\left(x_{1}^{*}\right)}$$

Define

$$D_{f}\Psi\left(f,x_{1}^{*},x_{2}^{*};h\right)$$

$$=\frac{\left[\frac{\partial h\left(x_{2}^{*}\right)}{\partial y_{2}}f\left(x_{1}^{*}\right)-\frac{\partial h\left(x_{1}^{*}\right)}{\partial y_{2}}f\left(x_{2}^{*}\right)+\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{2}}h\left(x_{1}^{*}\right)-\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{2}}h\left(x_{2}^{*}\right)\right]}{\left[\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{1}}f\left(x_{2}^{*}\right)-\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{1}}f\left(x_{1}^{*}\right)\right]}$$

$$-\frac{\left[\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{2}}f\left(x_{1}^{*}\right)-\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{2}}f\left(x_{2}^{*}\right)\right]\left[\frac{\partial h\left(x_{1}^{*}\right)}{\partial y_{1}}f\left(x_{2}^{*}\right)-\frac{\partial h\left(x_{2}^{*}\right)}{\partial y_{1}}f\left(x_{1}^{*}\right)+\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{1}}h\left(x_{2}^{*}\right)-\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{1}}h\left(x_{1}^{*}\right)\right]}{\left[\frac{\partial f\left(x_{1}^{*}\right)}{\partial y_{1}}f\left(x_{2}^{*}\right)-\frac{\partial f\left(x_{2}^{*}\right)}{\partial y_{1}}f\left(x_{1}^{*}\right)\right]^{2}}$$

$$R_{f}\Psi\left(f,x_{1}^{*},x_{2}^{*};h\right)=\Psi\left(f+h,x_{1}^{*},x_{2}^{*}\right)-\Psi\left(f,x_{1}^{*},x_{2}^{*}\right)-D_{f}\Psi\left(f,x_{1}^{*},x_{2}^{*};h\right)$$

$$D_{x_1}\Psi\left(f,x_1^*,x_2^*;\delta_1\right) = \frac{\partial}{\partial x_1}\left(\frac{\partial \log f_{Y|X=x_2^*}(y)/\partial y_2 - \partial \log f_{Y|X=x_1}(y)/\partial y_2}{\partial \log f_{Y|X=x_1}(y)/\partial y_1 - \partial \log f_{Y|X=x_2^*}(y)/\partial y_1}\right)|_{x_1=x_1^*}\ \delta_1$$

$$D_{x_2}\Psi\left(f,x_1^*,x_2^*;\delta_2\right) = \frac{\partial}{\partial x_2}\left(\frac{\partial \log f_{Y|X=x_2}(y)/\partial y_2 - \partial \log f_{Y|X=x_1^*}(y)/\partial y_2}{\partial \log f_{Y|X=x_1^*}(y)/\partial y_1 - \partial \log f_{Y|X=x_2}(y)/\partial y_1}\right)|_{x_2=x_2^*}\ \delta_2$$

$$D\Psi\left(f,x_1^*,x_2^*;h\right)=D_f\Psi\left(f,x_1^*,x_2^*;h\right)+D_{x_1}\Psi\left(f,x_1^*,x_2^*;D\kappa^1(f;h)\right)+D_{x_2}\Psi\left(f,x_1^*,x_2^*;D\kappa^2(f;h)\right)$$

and $R\Psi\left(f,x_1^*,x_2^*;h\right)=\Psi\left(f+h,\kappa^1\left(f+h\right),\kappa^2\left(f+h\right)\right)-\Psi\left(f,x_1^*,x_2^*\right)-D\Psi\left(f,x_1^*,x_2^*;h\right).$

Our assumptions imply that

$$\left|D\Psi\left(f,x_1^*,x_2^*;h\right)\right|\le a\left\|h\right\| \text{ and } \left|R\Psi\left(f,x_1^*,x_2^*;h\right)\right|\le a\ \left\|h\right\|^2$$

By standard properties of kernel estimators, it follows that when $h=\widehat{f}-f$,

$$\sqrt{N\sigma_N^5}D\Psi\left(f,x_1^*,x_2^*;h\right)$$

$$=\sqrt{N\sigma_N^5}\frac{\left[\frac{\partial h\left(x_2^*\right)}{\partial y_2}f\left(x_1^*\right)-\frac{\partial h\left(x_1^*\right)}{\partial y_2}f\left(x_2^*\right)\right]}{\left[\frac{\partial f\left(x_1^*\right)}{\partial y_1}f\left(x_2^*\right)-\frac{\partial f\left(x_2^*\right)}{\partial y_1}f\left(x_1^*\right)\right]}$$

$$-\sqrt{N\sigma_N^5}\frac{\left[\frac{\partial f\left(x_2^*\right)}{\partial y_2}f\left(x_1^*\right)-\frac{\partial f\left(x_1^*\right)}{\partial y_2}f\left(x_2^*\right)\right]\left[\frac{\partial h\left(x_1^*\right)}{\partial y_1}f\left(x_2^*\right)-\frac{\partial h\left(x_2^*\right)}{\partial y_1}f\left(x_1^*\right)\right]}{\left[\frac{\partial f\left(x_1^*\right)}{\partial y_1}f\left(x_2^*\right)-\frac{\partial f\left(x_2^*\right)}{\partial y_1}f\left(x_1^*\right)\right]^2}$$

$$+\sqrt{N\sigma_N^5}\frac{\frac{\partial}{\partial x_1}\left(\frac{\partial \log f_{Y|X=x_2^*}(y)/\partial y_2-\partial \log f_{Y|X=x_1}(y)/\partial y_2}{\partial \log f_{Y|X=x_1}(y)/\partial y_1-\partial \log f_{Y|X=x_2^*}(y)/\partial y_1}\right)|_{x_1=x_1^*}}{\left(\frac{\partial^2 \log f_{Y|X=x_1^*}(y)}{\partial x^2}\right)f(x_1^*)}\left(\frac{-\ \partial h(x_1^*)}{\partial x}\right)$$

$$+\sqrt{N\sigma_N^5}\frac{\frac{\partial}{\partial x_2}\left(\frac{\partial \log f_{Y|X=x_2}(y)/\partial y_2-\partial \log f_{Y|X=x_1^*}(y)/\partial y_2}{\partial \log f_{Y|X=x_1^*}(y)/\partial y_1-\partial \log f_{Y|X=x_2}(y)/\partial y_1}\right)|_{x_2=x_2^*}}{\left(\frac{\partial^2 \log f_{Y|X=x_2^*}(y)}{\partial x^2}\right)f(x_2^*)}\left(\frac{-\ \partial h(x_2^*)}{\partial x}\right)+o_p(1)$$

Hence, by standard results for kernel estimators, $\sqrt{N\sigma_N^5}D\Psi\left(f,x_1^*,x_2^*;h\right)\to N(0,\overline{V})$ where $\overline{V}$ is as defined prior to the statement of Theorem 4.4. Since our assumptions guarantee that $\sqrt{N\sigma_N^5}R\Psi\left(f,x_1^*,x_2^*;h\right)=o_p(1)$, we can conclude that

$$\sqrt{N\sigma_N^5}\left(\frac{\partial \widehat{m^1(y_2,\varepsilon_1)}}{\partial y_2}-\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2}\right)$$

$$=\sqrt{N\sigma_N^5}\left(\Psi\left(\widehat{f},\kappa^1\left(\widehat{f}\right),\kappa^2\left(\widehat{f}\right)\right)-\Psi\left(f,x_1^*,x_2^*\right)\right)\xrightarrow{d}N\left(0,\overline{V}\right)$$

This concludes the proof.

Proof of Theorem 5.2: To show the equivalence between (i) and (ii), we use an observational equivalence result in Matzkin (2008). Define

$$b\left(y,x;\partial r,\partial\widetilde{r},\partial^2 r,\partial^2\widetilde{r}\right)=-\Delta_x'+\Delta_y'\left(\frac{\partial\widetilde{r}\,(y,x)}{\partial y}\right)^{-1}\frac{\partial\widetilde{r}\,(y,x)}{\partial x}.$$

$$s\,(y,x;f_\varepsilon,r)=\frac{\partial\log f_\varepsilon\left(r\,(y,x)\right)}{\partial\varepsilon}$$

$$A\,(y,x;\partial r,\partial\widetilde{r})=\left[\frac{\partial r\,(y,x)}{\partial x}-\frac{\partial r\,(y,x)}{\partial y}\left(\frac{\partial\widetilde{r}\,(y,x)}{\partial y}\right)^{-1}\frac{\partial\widetilde{r}\,(y,x)}{\partial x}\right]$$

Matzkin (2008) shows that r and $\widetilde{r}$ are observationally equivalent if and only if for all (y,x),

$$s\,(y,x;f_\varepsilon,r)'\;A\,(y,x;\partial r,\partial\widetilde{r})\;=\;\left(\frac{\partial\log f_\varepsilon\,(r^1,r^2)}{\partial\varepsilon}\right)'\left(r_x-r_y\widetilde{r}_y^{\,-1}\widetilde{r}_x\right)$$

Letting $\left|r_y\right|_w$ denote the derivative of $|r_y|$ with respect to w, for $w=x,y_1,y_2$, we have that

$$b\left(y,x;\partial r,\partial\widetilde{r},\partial^2 r,\partial^2\widetilde{r}\right)$$

$$=\;-\frac{\left|r_y\right|_x}{\left|r_y\right|}+\frac{\left|\widetilde{r}_y\right|_x}{\left|\widetilde{r}_y\right|}+\left(\frac{\left|r_y\right|_{y_1}}{\left|r_y\right|}-\frac{\left|\widetilde{r}_y\right|_{y_1}}{\left|\widetilde{r}_y\right|},\frac{\left|r_y\right|_{y_2}}{\left|r_y\right|}-\frac{\left|\widetilde{r}_y\right|_{y_2}}{\left|\widetilde{r}_y\right|}\right)\widetilde{r}_y^{\,-1}\widetilde{r}_x$$

As shown in the proof of Theorem 5.1, in our model,

$$\begin{pmatrix} h_x^1(x,r^1,r^2) \\ \\ h_x^2(x,r^1,r^2)\end{pmatrix}=-r_y^{-1}r_x=\begin{pmatrix} r_{y_2}^1\left(\frac{r_x^2}{|r_y|}\right) \\ \\ -r_{y_1}^1\left(\frac{r_x^2}{|r_y|}\right)\end{pmatrix}$$

with similar expressions when h and r are substituted with $\widetilde{h}$ and $\widetilde{r}$. Taking the derivative of $h_x^1(x,r^1,r^2)$ with respect to y_1 and of $h_x^2(x,r^1,r^2)$ with respect to y_2, we get that

$$\begin{aligned}\frac{\partial h_x^1(x,r^1,r^2)}{\partial y_1}\;&=\;\frac{\partial}{\partial y_1}\left(r_{y_2}^1\left(\frac{r_x^2}{|r_y|}\right)\right)\\ &=\;r_{y_1,y_2}^1\left(\frac{r_x^2}{|r_y|}\right)+r_{y_2}^1\left(\frac{r_{y_1,x}^2}{|r_y|}\right)-\left(\frac{r_{y_2}^1\;r_x^2\;\left|r_y\right|_{y_1}}{\left|r_y\right|^2}\right),\text{ and}\end{aligned}$$

$$\frac{\partial h_x^2(x,r^1,r^2)}{\partial y_2} = \frac{\partial}{\partial y_2}\left(-r^1_{y_1}\left(\frac{r_x^2}{|r_y|}\right)\right)$$
$$= -r^1_{y_1,y_2}\left(\frac{r_x^2}{|r_y|}\right) - r^1_{y_1}\left(\frac{r^2_{y_2,x}}{|r_y|}\right) + \left(\frac{r^1_{y_1}\ r^2_x\ |r_y|_{y_2}}{|r_y|^2}\right)$$

Hence,

$$\frac{\partial h_x^1(x,r^1,r^2)}{\partial y_1} + \frac{\partial h_x^2(x,r^1,r^2)}{\partial y_2} = \frac{\partial}{\partial y_1}\left(r^1_{y_2}\left(\frac{r_x^2}{|r_y|}\right)\right) - \frac{\partial}{\partial y_2}\left(r^1_{y_1}\left(\frac{r_x^2}{|r_y|}\right)\right)$$
$$= \frac{-|r_y|_x}{|r_y|} + \left(\frac{|r_y|_{y_1}}{|r_y|}, \frac{|r_y|_{y_2}}{|r_y|}\right) r_y^{-1} r_x.$$

It follows that

$$b\left(y,x;\partial r,\partial\widetilde{r},\partial^2 r,\partial^2\widetilde{r}\right)$$
$$= -\frac{|r_y|_x}{|r_y|} + \frac{|\widetilde{r}_y|_x}{|\widetilde{r}_y|} + \left(\frac{|r_y|_{y_1}}{|r_y|} - \frac{|\widetilde{r}_y|_{y_1}}{|\widetilde{r}_y|}, \frac{|r_y|_{y_2}}{|r_y|} - \frac{|\widetilde{r}_y|_{y_2}}{|\widetilde{r}_y|}\right)\widetilde{r}_y^{-1}\widetilde{r}_x$$
$$= \frac{\partial h_x^1(x,r^1,r^2)}{\partial y_1} + \frac{\partial h_x^2(x,r^1,r^2)}{\partial y_2} - \left(\frac{|r_y|_{y_1}}{|r_y|}, \frac{|r_y|_{y_2}}{|r_y|}\right) r_y^{-1} r_x$$
$$-\frac{\partial \widetilde{h}_x^1(x,\widetilde{r}^1,\widetilde{r}^2)}{\partial y_1} - \frac{\partial \widetilde{h}_x^2(x,\widetilde{r}^1,\widetilde{r}^2)}{\partial y_2} + \left(\frac{|\widetilde{r}_y|_{y_1}}{|\widetilde{r}_y|}, \frac{|\widetilde{r}_y|_{y_2}}{|\widetilde{r}_y|}\right)\widetilde{r}_y^{-1}\widetilde{r}_x$$
$$+\left(\frac{|r_y|_{y_1}}{|r_y|} - \frac{|\widetilde{r}_y|_{y_1}}{|\widetilde{r}_y|}, \frac{|r_y|_{y_2}}{|r_y|} - \frac{|\widetilde{r}_y|_{y_2}}{|\widetilde{r}_y|}\right)\widetilde{r}_y^{-1}\widetilde{r}_x$$
$$= \frac{\partial\left[h_x^1(x,r^1,r^2) - \widetilde{h}_x^1(x,\widetilde{r}^1,\widetilde{r}^2)\right]}{\partial y_1} + \frac{\partial\left[h_x^2(x,r^1,r^2) - \widetilde{h}_x^2(x,\widetilde{r}^1,\widetilde{r}^2)\right]}{\partial y_2}$$
$$+\left(\frac{|r_y|_{y_1}}{|r_y|}, \frac{|r_y|_{y_2}}{|r_y|}\right)\left(\widetilde{r}_y^{-1}\widetilde{r}_x - r_y^{-1}r_x\right)$$

Hence, in our model, the expression

$$s\left(y,x;f_\varepsilon,r\right)'\ A\left(y,x;\partial r,\partial\widetilde{r}\right)\ = b\left(y,x;\partial r,\partial\widetilde{r},\partial^2 r,\partial^2\widetilde{r}\right)$$

is equivalent to the expression

$$\left(\frac{\partial \log f_\varepsilon\left(r^1,r^2\right)}{\partial\varepsilon}\right)' r_y\left(r_y^{-1}r_x - \widetilde{r}_y^{-1}\widetilde{r}_x\right)$$

$$= \frac{\partial \left[h_x^1(x, r^1, r^2) - \widetilde{h}_x^1(x, \widetilde{r}^1, \widetilde{r}^2) \right]}{\partial y_1} + \frac{\partial \left[h_x^2(x, r^1, r^2) - \widetilde{h}_x^2(x, \widetilde{r}^1, \widetilde{r}^2) \right]}{\partial y_2}$$

$$+ \left(\frac{|r_y|_{y_1}}{|r_y|}, \frac{|r_y|_{y_2}}{|r_y|} \right) \left(\widetilde{r}_y^{-1} \widetilde{r}_x - r_y^{-1} r_x \right)$$

Note that this is equivalent to

$$\left(\frac{\partial \log \left(f_\varepsilon \left(r^1, r^2 \right) \ |r_y| \right)}{\partial y} \right)' \begin{pmatrix} \widetilde{h}_x^1(x, \widetilde{r}^1, \widetilde{r}^2) - h_x^1(x, r^1, r^2) \\ \\ \widetilde{h}_x^2(x, \widetilde{r}^1, \widetilde{r}^2) - h_x^2(x, r^1, r^2) \end{pmatrix}$$

$$+ \frac{\partial \left[\widetilde{h}_x^1(x, \widetilde{r}^1, \widetilde{r}^2) - h_x^1(x, r^1, r^2) \right]}{\partial y_1} + \frac{\partial \left[\widetilde{h}_x^2(x, \widetilde{r}^1, \widetilde{r}^2) - h_x^2(x, r^1, r^2) \right]}{\partial y_2} = 0$$

Substituting $\left(f_\varepsilon \left(r^1, r^2 \right) \ |r_y| \right)$ by $f_{Y|X=x}(y)$, and multiplying both sides by $f_{Y|X=x}(y_1, y_2)$, we obtain

$$\frac{\partial \left(f_{Y|X=x}(y_1, y_2) \left[h_x^1(x, r^1, r^2) - \widetilde{h}_x^1(x, \widetilde{r}^1, \widetilde{r}^2) \right] \right)}{\partial y_1}$$

$$+ \frac{\partial \left(f_{Y|X=x}(y_1, y_2) \left[h_x^2(x, r^1, r^2) - \widetilde{h}_x^2(x, \widetilde{r}^1, \widetilde{r}^2) \right] \right)}{\partial y_2} = 0$$

This shows the equivalence between (i) and (ii).

To show that equivalence between (i) and (iii), suppose first that (r, f_ε) and $(\widetilde{r}, f_{\widetilde{\varepsilon}})$ are observationally equivalent, then for all y_1, y_2, x

$$f_{Y|X=x} \left(y_1, y_2 \right) = f_\varepsilon \left(r^1, r^2 \right) \ |r_y| = f_{\widetilde{\varepsilon}} \left(\widetilde{r}^1, \widetilde{r}^2 \right) \ |\widetilde{r}_y|$$

Using this in (ii), together with the expressions for $h_x^1, h_x^2, \widetilde{h}_x^1$, and $\widetilde{h}_x^2$, in terms of r and $\widetilde{r}$, we get that

$$\frac{\partial \left(f_\varepsilon \left(r^1, r^2 \right) \ r_{y_2}^1 \ r_x^2 \right)}{\partial y_1} - \frac{\partial \left(f_\varepsilon \left(r^1, r^2 \right) \ r_{y_1}^1 \ r_x^2 \right)}{\partial y_2}$$

$$= \frac{\partial \left(f_{\widetilde{\varepsilon}} \left(\widetilde{r}^1, \widetilde{r}^2 \right) \ \widetilde{r}_{y_2}^1 \ \widetilde{r}_x^2 \right)}{\partial y_1} - \frac{\partial \left(f_{\widetilde{\varepsilon}} \left(\widetilde{r}^1, \widetilde{r}^2 \right) \ \widetilde{r}_{y_1}^1 \ \widetilde{r}_x^2 \right)}{\partial y_2},$$

which is equivalent to

$$r_{y_2}^1 \frac{\partial \left(f_\varepsilon \left(r^1, r^2 \right) \ r_x^2 \right)}{\partial y_1} - \ r_{y_1}^1 \frac{\partial \left(f_\varepsilon \left(r^1, r^2 \right) \ r_x^2 \right)}{\partial y_2}$$

$$= \widetilde{r}_{y_2}^1 \frac{\partial \left(f_{\widetilde{\varepsilon}} \left(\widetilde{r}^1, \widetilde{r}^2 \right) \ \widetilde{r}_x^2 \right)}{\partial y_1} - \ \widetilde{r}_{y_1}^1 \frac{\partial \left(f_{\widetilde{\varepsilon}} \left(\widetilde{r}^1, \widetilde{r}^2 \right) \ \widetilde{r}_x^2 \right)}{\partial y_2}.$$

Substituting $f_\varepsilon \left(r^1, r^2 \right)$ by $f_{Y|X}(y) / \ |r_y|$ and $f_{\widetilde{\varepsilon}} \left(\widetilde{r}^1, \widetilde{r}^2 \right)$ by $f_{Y|X}(y) \ / \, |\widetilde{r}_y|$ we obtain (iii). We have then shown that (i) implies (iii).

We next show that (iii) implies (iv). For this, we express f_{ε} and $f_{\widetilde{\varepsilon}}$ as the multiplication of a marginal and a conditional density,

$$f_{\varepsilon_1,\varepsilon_2}\left(r^1,r^2\right)=f_{\varepsilon_1}\left(r^1\right)\ f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right),\ \text{ and } f_{\widetilde{\varepsilon}_1,\widetilde{\varepsilon}_2}\left(\widetilde{r}^1,\widetilde{r}^2\right)=f_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)\ f_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)$$

Expression (iii) is equivalent to

$$r^1_{y_2}\frac{\partial\left(f_{\varepsilon_1}\left(r^1\right)\ f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_x\right)}{\partial y_1}-\ r^1_{y_1}\frac{\partial\left(f_{\varepsilon_1}\left(r^1\right)\ f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_x\right)}{\partial y_2}$$
$$=\widetilde{r}^1_{y_2}\frac{\partial\left(f_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)\ f_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)\ \widetilde{r}^2_x\right)}{\partial y_1}-\ \widetilde{r}^1_{y_1}\frac{\partial\left(f_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)\ f_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)\ \widetilde{r}^2_x\right)}{\partial y_2},$$

which can be shown to be equal to

$$\frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_2}\frac{\partial\left(f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_x\right)}{\partial y_1}-\frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_1}\frac{\partial\left(f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_x\right)}{\partial y_2}$$
$$=\frac{\partial F_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)}{\partial y_2}\frac{\partial\left(\ f_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)\ \widetilde{r}^2_x\right)}{\partial y_1}-\frac{\partial F_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)}{\partial y_1}\frac{\partial\left(f_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)\ \widetilde{r}^2_x\right)}{\partial y_2}.$$

Integrating both sides with respect to x, under assumptions allowing to exchange the the order of differentiation and integration, we get that

$$\frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_2}\frac{\partial\left(\int\ f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_x\ dx\right)}{\partial y_1}-\frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_1}\frac{\partial\left(\int\ f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_x\ dx\right)}{\partial y_2}$$
$$=\frac{\partial F_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)}{\partial y_2}\frac{\partial\left(\int\ f_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)\ \widetilde{r}^2_x\ dx\right)}{\partial y_1}-\frac{\partial F_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)}{\partial y_1}\frac{\partial\left(\int\ f_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)\ \widetilde{r}^2_x\ dx\right)}{\partial y_2}.$$

Using the transformation $w=r^2\left(y_1,y_2,x\right)$ for the left hand side, and the transformation $\widetilde{w}=\widetilde{r}^2\left(y_1,y_2,x\right)$ for the right-hand-side, we get

$$\frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_2}\frac{\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)}{\partial y_1}-\frac{\partial F_{\varepsilon_1}\left(r^1\right)}{\partial y_1}\frac{\partial F_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)}{\partial y_2}$$
$$=\frac{\partial F_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)}{\partial y_2}\frac{\partial F_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)}{\partial y_1}-\frac{\partial F_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)}{\partial y_1}\frac{\partial F_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)}{\partial y_2}.$$

Hence, we have shown that (iii) implies (iv).

It only remains to show that (iv) implies (i). But this easily follows by noticing that taking the derivatives in (iv), we get

$$f_{\varepsilon_1}\left(r^1\right)\ r^1_{y_2}\ f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_{y_1}-f_{\varepsilon_1}\left(r^1\right)\ r^1_{y_1}\ f_{\varepsilon_2|\varepsilon_1=r^1}\left(r^2\right)\ r^2_{y_2}$$

$$= f_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)\ \widetilde{r}^1_{y_2}\ f_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)\ \widetilde{r}^2_{y_1} - f_{\widetilde{\varepsilon}_1}\left(\widetilde{r}^1\right)\ \widetilde{r}^1_{y_1}\ f_{\widetilde{\varepsilon}_2|\widetilde{\varepsilon}_1=\widetilde{r}^1}\left(\widetilde{r}^2\right)\ \widetilde{r}^2_{y_2}.$$

which is equivalent to

$$f_{\widetilde{\varepsilon}_1,\widetilde{\varepsilon}_2}\left(\widetilde{r}^1,\widetilde{r}^2\right)\ \left|\widetilde{r}_y\right|\ =\ f_{\varepsilon_1,\varepsilon_2}\left(r^1,r^2\right)\ \left|r_y\right|$$

Hence, (iv) implies (i). This completes the proof of Theorem 5.2.

8. References

AI, C. and X. CHEN (2003), "Efficient Estimation of Models with Conditional Moments Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1843.

ALTONJI, J.G. AND H. ICHIMURA (2000): "Estimating Derivatives in Nonseparable Models with Limited Dependent Variables," Mimeo, Northwestern University.

ALTONJI, J.G. AND R.L. MATZKIN (2001): "Panel Data Estimators for Nonseparable Models with Endogenous Regressors", NBER Working Paper T0267.

ALTONJI, J.G. AND R.L. MATZKIN (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica,* 73, 3, 1053-1102.

BHATTACHARYA, P.K. (1963), "On an Analog of Regression Analysis," *The Annals of Mathematical Statistics*, Vol. 34, No. 4, pp. 1459-1473.

BENKARD, C.L. and S. BERRY (2006) "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on B. Brown (1983) and Roehrig (1988), *Econometrica, 74, 5.*

BERRY, S.T. and P.A. HAILE (2009) "Identification in Differentiated Products Markets Using Market Level Data," working paper, Yale University.

BLUNDELL, R. and R.L. MATZKIN (2010) "Conditions for the Existence of Control Functions in Nonadditive Simultaneous Equation Models," mimeo, UCL and UCLA.

BLUNDELL, R. and J. L. POWELL (2003a) "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress,* Volume II, edited by M. Dewatripont, L.P. Hansen, and S.J.Turnovsky, Cambridge University Press, Cambridge, U.K.

BOWDEN, R. (1973) "The Theory of Parametric Identification," *Econometrica*, 41, 1069-1074.

BROWN B.W. (1983) "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51, 175-196.

BROWN, D.J. and R.L. MATZKIN (1998) "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," Cowles Foundation Discussion Paper #1175.

CHERNOZHUKOV, V. AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

CHERNOZHUKOV, V., G. IMBENS, AND W. NEWEY (2007): "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions," *Journal of Econometrics*.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, Vol. 71, No. 5.

CHIAPPORI, P-A, and I. KOMUNJER (2009) "On the Nonparametric Identification of Multiple Choice Models," discussion paper, UCSD.

DAROLLES, S., J.P. FLORENS, and E. RENAULT (2002), "Nonparametric Instrumental Regression," mimeo, IDEI, Toulouse.

FISHER, F.M. (1959). "Generalization of the rank and order conditions for identifiability". Econometrica 27, (3), 431–447.

FISHER, F.M. (1961) "Identifiability Criteria in Nonlinear Systems," *Econometrica*, 29, 574-590.

FISHER, F.M. (1965) "Identifiability Criteria in Nonlinear Systems: A Further Note," *Econometrica*, 33, 197-205.

FISHER, F.M. (1966) *The Identification Problem in Econometrics.* New York: McGraw-Hill.

HALL, P. and J.L. HOROWITZ (2003), "Nonparametric Methods for Inference in the Presence of Instrumental Variables," Center for Microdata Methods and Practice, WP 102/03.

HAUSMAN, J.A. (1983) “Specification and Estimation of Simultaneous Equation Models,” in *Handbook of Econometrics,* Vol. I, edited by Z. Griliches and M.D. Intriligator.

HAAVELMO, T.M. (1943). “The statistical implications of a system of simultaneous equations”. *Econometrica,* 11, (1), 1–12.

HAAVELMO, T.M. (1944) "The Probability Approach in Econometrics," *Econometrica*, 12, Supplement, July.

HSIAO, C. (1983), "Identification," in *Handbook of Econometrics*, Vol. 1, edited by Z. Griliches and M.D. Intriligator, North-Holland Publishing Company.

HURWICZ, L. (1950) "Generalization of the Concept of Identification," in *Statistical Inference in Dynamic Economic Models,* Cowles Commission Monograph 10. New York: John Wiley.

IMBENS, G.W. AND W.K. NEWEY (2003) “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” mimeo, M.I.T.

IMBENS, G.W. AND W.K. NEWEY (2009) “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” Econometrica, 77, 5.

KOENKER, R. (2005) *Regression Quantiles*, Econometric Society Monograph, Cambridge University Press.

KOOPMANS, T.C. and O. REIERSOL (1950) "The Identification of Structural Characteristics," *Annals of Mathematical Statistics,* 21, 165-181.

KOOPMANS, T.C., A. RUBIN, and R.B. LEIPNIK (1950) "Measuring the Equation System of Dynamic Economics," in *Statistical Inference in Dynamic Equilibrium Models*, Cowles Commission Monograph 10, New York: John Wiley.

MANSKI, C.F. (1983). “Closest empirical distribution estimation”. Econometrica 51 (2), 305–320.

MATZKIN, R.L. (1999) "Nonparametric Estimation of Nonadditive Random Functions," mimeo.

MATZKIN, R.L. (2003) "Nonparametric Estimation of Nonadditive Random Functions," Econometrica, 71, 1339-1375.

MATZKIN, R.L. (2004) "Unobservable Instruments," mimeo, Northwestern University.

MATZKIN, R.L. (2005a) "Identification in Nonparametric Simultaneous Equations Models," mimeo, Northwestern University.

MATZKIN, R.L. (2005b) "Heterogenous Choice," invited lecture, World Congress of the Econometric Society, London, U.K.

MATZKIN, R.L. (2007a) "Nonparametric Identification," Chapter 73 in *Handbook of Econometrics,* Vol. 6b, edited by J.J. Heckman and E.E. Leamer, Elsevier B.V., 5307-5368.

MATZKIN, R.L. (2007b) "Heterogeneous Choice," in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society,* edited by R. Blundell, W. Newey, and T. Persson, Cambridge University Press.

MATZKIN, R.L. (2008) "Identification in Nonparametric Simultaneous Equations," *Econometrica,* Vol. 76, No. 5, 945-978.

MATZKIN, R.L. (2010a) "Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity," mimeo, UCLA.

MATZKIN, R.L. (2010b) "Estimation in Nonparametric Limited Dependent Variable Models with Simultaneity," in preparation, UCLA.

NADARAYA, E. (1964) "On Estimating Regression," *Theory Probab.Appl.*, 9, 141-142.

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

NEWEY, W. and J. POWELL (1989) "Instrumental Variables Estimation of Nonparametric Models," mimeo.

NEWEY, W. and J. POWELL (2003) “Instrumental Variables Estimation of Nonparametric Models,” *Econometrica*.

POWELL, J.L., J.H. STOCK, and T.M. STOKER (1989) "Semiparametric Estimation of Index Coefficients," *Econometrica*, 51, 1403-1430.

ROEHRIG, C.S. (1988) "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica,* 56:2, 433-447.

ROTHENBERG, T.J. (1971) "Identification in Parametric Models," *Econometrica*, 39, 577-592.

STOKER, T.M. (1986) "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461-1481.

STONE, C.J. (1977) "Consistent Nonparametric Regression," *The Annals of Statistics,* Vol. 5, No. 4, pp. 595-620.

WALD, A. (1950) "Note on Identification of Economic Relations," in *Statistical Inference in Dynamic Economic Models,* Cowles Commission Monograph 10. New York: John Wiley.

WEGGE, L.L. (1965) "Identifiability Criteria for a System of Equations as a Whole," *The Australian Journal of Statistics,* 7, 67-77.

WATSON, G.S. (1964) "Smooth Regression Analysis," *Sankhyā* : *The Indian Journal of Statistics,* Vol. 26, No. 4, 359-372.