ESTIMATION OF NONPARAMETRIC MODELS WITH SIMULTANEITY

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

First version: May 2010
This version: April 2012

Incomplete and unfinished version. Please do not distribute.

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 two models and two estimators, and develop one estimator for each model. In our first model, to each structural equation there corresponds an exclusive regressor. In our second model, there are two equations and only one instrument, which is excluded from the equation of interest. For the first model, we develop an estimator for the matrix of derivatives of the structural function which 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 develop indirect estimators, which are calculated paralleling constructive identification results. These estimators 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. Both estimators are based on new observational equivalence results, which we also introduce in the paper, and which can be used to develop new estimators for these models. We show that our estimators are consistent and asymptotically normal.

*The support of NSF is gratefully acknowledged. I have benefitted from detailed comments and suggestions from several readers of the first version of this paper, and from comments of participants at a large number of conferences and seminars.
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.


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 for models with nonadditive random terms, Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), and Chernozhukov, Chen, Lee, and Newey (2011). Identification in these models has been studied in terms of conditions on the reduced form for the endogenous regressors. The estimators are defined as solutions to integral equations, which may suffer from ill-posed inverse problems.

In this paper, we make assumptions and construct nonparametric estimators in ways that are significantly different from those nonparametric methods for models with simultaneity. In particular, our estimators are closely tied to pointwise identification conditions on the structural model. Our conditions 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 other words, the goal of this paper is to develop estimators that can be expressed in closed form. 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). The set of simultaneous equations that satisfy the conditions

¹See Koenker (2005) for other quantile methods.
required to employ these methods is very restrictive. (See Blundell and Matzkin (2010) for a characterization of simultaneous equation models that can be estimated using a control function approach.) The goal of this paper is to fill this important gap.

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.

We present two estimation approaches and focus on two models. For each of the two models, we present in this paper one estimation approach. However, the approaches can be used to estimate either model, as well as extensions and various modifications of both 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 estimator that we present for this model is an average-derivative type method. In this method, the calculation of the estimator for the derivatives of the reaction functions of the players 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

²Existent 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.
Our second model is a two equation model with one instrument. Consider for example 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 develop a two-step, indirect type estimator for the derivative of the demand function with respect to price. 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. The estimator is developed by substituting the joint density of price, quantity, and cost, by a nonparametric estimator for it. In the first-step, the density of the observable variables is estimated nonparametrically. In the second step, the estimator for the derivative of the demand function with respect to price is read off this nonparametric estimator of the density. The estimators that we develop are consistent and asymptotically normal.

For both models, the exclusive regressors model and the two equations with one instrument model, we present new results that can be used to develop new identification results and related estimators. The results immediately indicate ways for showing constructive identification of some or all of the elements in these simultaneous equations models. In particular, the results can be used to develop average derivative estimators for the second model, indirect estimators for the first model, and to develop estimators for only a few elements of the system of simultaneous equations.

Besides the identification results for simultaneous equation models in Matzkin (2008), the closest identification results that were developed prior to this paper were Matzkin (2007b) for a model of consumer demand, Chiappori and Komunjer (2009) for a multinomial model, and Berry and Haile (2009) for a model of demand for differentiated products. Matzkin (2007b) derived constructive identification results for a model with exclusive regressors, assuming that the density of the unobservable variables has a unique known mode. Chiappori and Komunjer (2009) derived identification results in a multinomial model, by creating a mapping between the second order derivatives of the log density of the observable variables and the second order derivatives of the log density of the unobservable variables. They employed this mapping to show generic identification. One of our identification results is also based on a mapping between second order derivatives of the log density of the observable variables and the log density of the unobservable variables. However, in contrast to Chiappori and Komunjer (2009), we use this relationship to find the particular values of the exogenous variables that correspond to the values of the unobservable variables from which we can derive constructive identification results. Berry and Haile (2009) presented a constructive identification result based on marginal integration, for the exclusive regressor model in Matzkin (2007b), under different assumptions. A recent paper by Berry and Haile
(2011) deals with constructive identification in the model analyzed in Matzkin (2008, Section 4.2).

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 (2012) for identification of some such models and Matzkin (2012a) 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. The Appendix contains some of the proofs.

2. Nonadditive simultaneous equations

Our basic model can be described as

\[(2.1) \quad s(Y, X, \varepsilon) = 0\]

where \(Y\) denotes a \(G\)-dimensional vector of observable endogenous variables, \(X\) denotes a \(K\)-dimensional vector of observable exogenous variables, \(\varepsilon\) denotes a \(G\)-dimensional vector of unobserved variables, and \(s : \mathbb{R}^{G+K+G} \rightarrow \mathbb{R}^G\) is an unknown function.

We assume that the function \(s\) is such that for any value \((x, \varepsilon)\) of \((X, \varepsilon)\), 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 \(\varepsilon\) such that \(s(y, x, \varepsilon) = 0\). We will denote the function that assigns the values of \(y\) that satisfies (2.1) for \((x, \varepsilon)\) by \(h(x, \varepsilon)\) and we will denote the function that assigns the value of \(\varepsilon\) 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 \(\mathbb{R}^G\). The vector of unobservable variables \(\varepsilon\) will be assumed to be distributed independently of \(X\) and to possess an everywhere positive and continuously differentiable density \(f\varepsilon\). 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,

\[(2.2) \quad Q = D(P, I, \varepsilon_D)\]
\[P = S(Q, W, \varepsilon_S)\]

where \(Q\) is observed quantity, \(P\) is observed price, \(I\) is observed income of the consumers, \(W\) is observed production costs, \(\varepsilon_D\) and \(\varepsilon_S\) 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, \(\varepsilon_D\) and \(\varepsilon_S\),

\[
Q = \beta_0 + \beta_1 P + \beta_2 I + \varepsilon_D \\
P = \alpha_0 + \alpha_1 Q + \alpha_2 W + \varepsilon_S
\]

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

\[
P = \gamma_0 + \gamma_1 I + \gamma_2 W + v_P \\
Q = \delta_0 + \delta_1 I + \delta_2 W + v_Q
\]

each with one additive unobservable variable. The linearity and additivity of the structural equations guarantees that the effect of the two unobservable variables, \(\varepsilon_D\) and \(\varepsilon_S\), 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 \(\varepsilon_D\) and \(\varepsilon_S\) 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, \(\varepsilon_D\) and \(\varepsilon_S\). In other words, we will only be able to establish that for some nonparametric functions \(h_D\) and \(h_S\),

\[
Q = h_D(I, W, \varepsilon_D, \varepsilon_S) \\
P = h_S(I, W, \varepsilon_D, \varepsilon_S)
\]

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, \((X'X)^{-1}XY\), 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

\[
\begin{align*}
Q &= D(P, \epsilon_D) \\
P &= S(Q, W, \epsilon_S)
\end{align*}
\]

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

#### 3.1. The model

We consider in this section the model

\[
\begin{align*}
Y_1 &= m^1(Y_2, Y_3, ..., Y_G, Z, X_1, \epsilon_1) \\
Y_2 &= m^2(Y_1, Y_3, ..., Y_G, Z, X_2, \epsilon_2) \\
&\quad \quad \quad \quad \cdot \\
Y_G &= m^G(Y_1, Y_2, ..., Y_{G-1}, Z, X_G, \epsilon_G)
\end{align*}
\]

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 \((\epsilon_1, ..., \epsilon_G)\) is a vector of unobservable variables. The observable vector \( Z \) has the effect of decreasing the rates of convergence of nonparametric estimators of model (3.1) but does not add complications for identification, as all our assumptions and identification conclusions can be interpreted as holding conditionally on \( Z \). Hence, for simplicity of exposition, we will omit \( Z \) from the model.

Since for each \( g \), the function \( m^g \) is unknown and the nonadditive \( \epsilon_g \) is unobservable, we will at most be able to identify the values of \( \epsilon_g \) up to an invertible transformation.\(^3\) Hence, for each \( g \), we may normalize \( m^g \) to be either strictly increasing in \( \epsilon_g \), as it is assumed in models additive in \( \epsilon_g \), or we may normalize \( m^g \) to be strictly decreasing in \( \epsilon_g \). The invertibility of

---

\(^3\)See Matzkin (1999, 2003, 2007) for discussion of this nonidentifiability result in the one equation model, \( y = m(x, \varepsilon) \), with \( x \) and \( \varepsilon \) independently distributed.
$m^g$ in $\varepsilon_g$ implies that, for any fixed values of the other arguments of the function $m^g$, there is a unique value of $\varepsilon_g$ for each value of $y_g$. We will denote the function that assigns such value of $\varepsilon_g$ by $r^g(y_1, \ldots, y_G, x_g)$. Our system of indirect structural equations, denoting the mapping from the vectors of observable variables to the vector of unobservable variables, is expressed as

$$
\begin{align*}
\varepsilon_1 &= r^1(Y_1, \ldots, Y_G, X_1) \\
\varepsilon_2 &= r^2(Y_1, \ldots, Y_G, X_2) \\
&\quad \vdots \\
\varepsilon_G &= r^G(Y_1, \ldots, Y_G, X_G)
\end{align*}
$$

(3.2)

The derivatives of the function $m^g$ can be calculated by substituting (3.2) into (3.1) and differentiating with respect to the various arguments. The derivative of $m^g$ with respect to $y_j$, when $j \neq g$ and $y_g$ is a specified value is

$$
\frac{\partial m^g(y_{-g}, x_g, \varepsilon_g)}{\partial y_j} |_{\varepsilon_g=r^g(y_1, \ldots, y_G, x_g)} = - \left[ \frac{\partial r^g(y_g, y_{-g}, x_g)}{\partial y_g} \right]^{-1} \left[ \frac{\partial r^g(y_g, y_{-g}, x_g)}{\partial y_j} \right]
$$

The derivatives of $m^g$ with respect to $x_g$ is the same expression as the derivative for $y_j$, except that $\partial r^g/\partial y_j$ is substituted by $\partial r^g/\partial x_g$. The derivative of $m^g$ with respect to $\varepsilon^g$ when $y_g$ is a specified value is

$$
\frac{\partial m^g(y_{-g}, x_g, \varepsilon_g)}{\partial \varepsilon_g} |_{\varepsilon_g=r^g(y_1, \ldots, y_G, x_g)} = - \left[ \frac{\partial r^g(y_g, y_{-g}, x_g)}{\partial y_g} \right]^{-1}
$$

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 $(X_1, \ldots, X_G, \varepsilon_1, \ldots, \varepsilon_G)$, the system in (3.1) has a unique solution. We will let the values of functions $h^1, \ldots, h^G$ denote the solution to these equations,

$$
\begin{align*}
Y_1 &= h^1(X_1, \ldots, X_G, \varepsilon_1, \ldots, \varepsilon_G) \\
Y_2 &= h^2(X_1, \ldots, X_G, \varepsilon_1, \ldots, \varepsilon_G) \\
&\quad \vdots \\
Y_G &= h^G(X_1, \ldots, X_G, \varepsilon_1, \ldots, \varepsilon_G)
\end{align*}
$$

(3.3)

Our first two assumptions impose conditions guaranteeing that the mapping between the
structural elements $r$ and $f_\varepsilon$ and the conditional densities of the observable variables, $f_{Y|X=x}$ is given by the transformation of variables equation

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

where $|\partial r(y, x)/\partial x|$ denotes the Jacobian determinant of $r(y, x)$ with respect to $y$. Assumption 3.1 in addition guarantees that for each $g$, the function $r^g$ has a nonvanishing derivative with respect to its exclusive regressor, $x_g$.

**Assumption 3.1:** Conditional on $(X_1, ..., X_G)$, the functions $r = (r^1, ..., r^G)$ and $h = (h^1, ..., h^G)$ are twice continuously differentiable, $1-1$, onto $R^G$, and their Jacobian determinants are bounded away from zero. Moreover, for each $g$, the derivative of $r^g$ with respect to $x_g$ is bounded away from zero.

**Assumption 3.2:** $(\varepsilon_1, ..., \varepsilon_G)$ is distributed independently of $(X_1, ..., X_G)$ with an everywhere positive and twice continuously differentiable density, $f_\varepsilon$.

For the analysis of identification, the right-hand-side of (3.4) can be assumed known. In practice, it can be estimated nonparametrically. The left-hand-side involves the structural functions, $f_\varepsilon$ and $r$, whose features are the objects of interest. Assumptions 3.1 and 3.2 together with our next assumption guarantee that both sides of (3.4) can be differentiated with respect to $y$ and $x$. This will allow us to transform (3.4) into a system of linear equations with derivatives of known functions on one side and derivatives of unknown functions on the other side.

**Assumption 3.3:** $(X_1, ..., X_G)$ possesses a differentiable density.

If $r$ is identified, $f_\varepsilon$ is also identified, since for any given $x$, $f_\varepsilon(\varepsilon) = f_{Y|X=x}(h(x, \varepsilon)) \left| \partial h(x, \varepsilon)/\partial \varepsilon \right|$, and $h$ is identified if $r$ is identified. We will concentrate on developing estimators for the identified features of $r$. Theorem 3.1 below characterizes the features of $r$ that can be identified. Roughly, the theorem states that, under appropriate conditions on the density, $f_\varepsilon$, of $\varepsilon$, and on a vector of composite derivatives of $\log |\partial r(y, x)/\partial y|$, the ratios of derivatives, $r^g_{y_j}(y, x_g)/r^g_{x_j}(y, x_g)$, of each of the functions, $r^g$, with respect to its coordinates, are identified. The statement that the ratios of derivatives, $r^g_{y_j}(y, x_g)/r^g_{x_j}(y, x_g)$, is identified, is equivalent to the statement that for each $g$, $r^g$ is identified up to an invertible transformation. This suggests considering restrictions on the set of functions $r^g$, which guarantee that no two different functions satisfying those restrictions are invertible transformations of each
other.\textsuperscript{4} One such class can be defined by requiring that for each function in the class there exists a function \(s^g : R^G \rightarrow R\) such that for all \(x_g\),

\[
(3.5) \quad r^g (y, x_g) = s^g (y) + x_g
\]

and such that \(s^g (\overline{y}) = \alpha\), where \(\overline{y}\) and \(\alpha\) are specified and constant over all the functions in the class.

### 3.2. Observational Equivalence

To motivate the additional restrictions that we will impose, we first present an observational equivalence result for the exclusive regressor models. The result is obtained by specializing the observational equivalent results in Matzkin (2008) to the exclusive regressors model, and by expressing those results in terms of the density, \(f_{Y,X}\), of the observable variables instead of in terms of the density, \(f_\varepsilon\), of the vector of unobservable variables, \(\varepsilon\). We also modify the the results further by restricting the definition of observational equivalence to a subset of the support of the vector of observable variables.

We first introduce some notation, which will be used throughout the paper. Let \(f_{Y|X=x}(y)\) denote the conditional density of the vector of observable variables. We will denote the derivative with respect to \(y\) of the log of \(f_{Y|X=x}(y)\), \(\partial \log f_{Y|X=x}(y)/\partial y\), by \(g_y(y, x) = (g_{y1}(y, x), \ldots, g_{yG}(y, x))'\). The derivative, \(\partial \log f_{Y|X=x}(y)/\partial x\), of the log of \(f_{Y|X=x}(y)\) with respect to \(x\), will be denoted by \(g_x(y, x) = (g_{x1}(y, x), \ldots, g_{xG}(y, x))'\). The derivative of the log of the density, \(f_\varepsilon\), of \(\varepsilon\), with respect to \(\varepsilon\), \(\partial \log f_\varepsilon(\varepsilon)/\partial \varepsilon\), will be denoted by \(q_\varepsilon(\varepsilon) = (q_{\varepsilon1}(\varepsilon), \ldots, q_{\varepsilonG}(\varepsilon))'\). When \(\varepsilon = r(y, x)\), \(\partial \log f_\varepsilon(\varepsilon)/\partial \varepsilon\) will be denoted by \(q_\varepsilon(r(y, x))\) or by \(q_\varepsilon(r)\). For each \(r^g\) and each \(j\), the ratio of derivatives of \(r^g\) with respect to \(y_j\) and \(x_g\), \(r_{y_j}^g (y, x_g)/r_{x_g}^g (y, x_g)\), will be denoted by \(\overline{r}_{y_j}^g (y, x_g)\). For an alternative function \(\tilde{r}\), these ratios of derivatives will be denoted by \(\tilde{r}_{y_j}^g (y, x_g)\). The Jacobian determinants, \(|\partial r(y, x)/\partial y|\) and \(|\partial \tilde{r}(y, x)/\partial y|\), will be denoted respectively by \(|r_y(y, x)|\) and \(|\tilde{r}_y(y, x)|\). The derivatives of \(|r_y(y, x)|\) and \(|\tilde{r}_y(y, x)|\) with respect to any of their arguments, \(w \in \{y_1, \ldots, y_G, x_1, \ldots, x_G\}\), will be denoted by \(|r_y(y, x)|_w\) and \(|\tilde{r}_y(y, x)|_w\). The statement of our main observational equivalence result in this section involves functions, \(d_{yg}\), defined for each \(g\) by

\[
(3.6) \quad d_{yg}(y, x) = \frac{|r_y (y, x)|_{yg}}{|r_y (y, x)|} - \sum_{k=1}^G \left[ \frac{|r_y (y, x)|_{xk}}{|r_y (y, x)|} \frac{r_{yk}^k (y, x_k)}{r_{xk}^k (y, x_k)} \right]
\]

\textsuperscript{4} Examples of classes of nonparametric functions satisfying that no two functions in the set are invertible transformations of each other were studied in Matzkin (1992, 1993, 1994) in the context of threshold crossing, binary, and multinomial choice models, and in Matzkin (1999, 2003) in the context of a one equation model with a nonadditive random term.
The term $d_{yk}(y, x)$ represents the effect on $\log|\partial r(y, x)/\partial y|$ of a simultaneous change in $y_k$ and in $(x_1, ..., x_G)$.

Let $\Gamma$ denote the set of functions $r$ that satisfy Assumption 3.1 and let $\Phi$ denote the set of densities that satisfy Assumption 3.2. We define observational equivalence within $\Gamma$ over a subset, $M$, of the support of the vector of observable variables.

**Definition:** Let $M$ denote a subset of the support of $(Y, X)$, such that for all $(y, x) \in M$, $f_{Y,X}(y, x) > \delta_2$, where $\delta_2$ is any positive constant. A function $\tilde{r} \in \Gamma$ is observationally equivalent to $r \in \Gamma$ on $M$ if there exist densities $f_\varepsilon$ and $f_{\tilde{\varepsilon}}$ satisfying Assumption 3.2 and such that for all $(y, x) \in M$,

$$f_\varepsilon(r(y, x)) |r_y(y, x)| = f_{\tilde{\varepsilon}}(\tilde{r}(y, x)) |\tilde{r}_y(y, x)|$$

When $(r, f_\varepsilon)$ is the pair of indirect structural function and density generating $f_{Y|X}$, the definition states that $\tilde{r}$ is observationally equivalent to $r$ if there is a density that together with $\tilde{r}$ generates $f_{Y|X}$. The following theorem provides a characterization of observational equivalence on $M$.

**Theorem 3.1:** Suppose that $(r, f_\varepsilon)$ generates $f_{Y|X}$. A function $\tilde{r} \in \Gamma$ is observationally equivalent to $r \in \Gamma$ on $M$ if and only if for all $(y, x) \in M$,

$$0 = \left(\frac{\partial}{\partial y_1} \tilde{r}_{y_1} - \frac{\partial}{\partial y_1} r_{y_1}\right) g_{x_1} + \left(\frac{\partial}{\partial y_2} \tilde{r}_{y_2} - \frac{\partial}{\partial y_2} r_{y_2}\right) g_{x_2} + \cdots + \left(\frac{\partial}{\partial y_G} \tilde{r}_{y_G} - \frac{\partial}{\partial y_G} r_{y_G}\right) g_{x_G} + (d_{y_1} - \tilde{d}_{y_1})$$

(3.8) $$0 = \left(\frac{\partial}{\partial y_1} \tilde{r}_{y_1} - \frac{\partial}{\partial y_1} r_{y_1}\right) g_{x_1} + \left(\frac{\partial}{\partial y_2} \tilde{r}_{y_2} - \frac{\partial}{\partial y_2} r_{y_2}\right) g_{x_2} + \cdots + \left(\frac{\partial}{\partial y_G} \tilde{r}_{y_G} - \frac{\partial}{\partial y_G} r_{y_G}\right) g_{x_G} + (d_{y_2} - \tilde{d}_{y_2})$$

(3.8) $$0 = \left(\frac{\partial}{\partial y_1} \tilde{r}_{y_1} - \frac{\partial}{\partial y_1} r_{y_1}\right) g_{x_1} + \left(\frac{\partial}{\partial y_2} \tilde{r}_{y_2} - \frac{\partial}{\partial y_2} r_{y_2}\right) g_{x_2} + \cdots + \left(\frac{\partial}{\partial y_G} \tilde{r}_{y_G} - \frac{\partial}{\partial y_G} r_{y_G}\right) g_{x_G} + (d_{y_G} - \tilde{d}_{y_G})$$

where for each $g$ and $j$, $\frac{\partial}{\partial y_j} \tilde{r}_{y_j} = \frac{\partial}{\partial y_j} r_{y_j}(y, x_g) = \frac{\partial}{\partial y_j} \tilde{r}_{y_j}(y, x_g)$, $\frac{\partial}{\partial x_j} \tilde{r}_{y_j} = \frac{\partial}{\partial x_j} \tilde{r}_{y_j}(y, x_g)$, $\frac{\partial}{\partial y_j} \tilde{r}_{y_j} = \frac{\partial}{\partial y_j} \tilde{r}_{y_j}(y, x_g)$, $d_{y_1} = d_{y_1}(y, x)$, $\tilde{d}_{y_1} = \tilde{d}_{y_1}(y, x)$, and $g_{x_j} = \partial \log f_{Y|X=x}(y)/\partial x_j$.

The proof, presented in the Appendix, uses (3.4) to obtain an expression for the unobservable $\partial \log f_\varepsilon(r(y, x))/\partial \varepsilon$ in terms of the observable $g_x(y, x) = \partial \log f_{Y|X=x}(y)/\partial x$. The expression in terms of $g_x(y, x)$ is used to substitute $\partial \log f_\varepsilon(r(y, x))/\partial \varepsilon$ in the observational equivalence result in Matzkin (2008, Theorem 3.2). Equation (3.8) is obtained after manipulating the equations resulting from such substitution.

The necessity of (3.8) can be shown in the following, more transparent, way. Taking logs
and differentiating both sides of (3.4) with respect to \( y_j \) gives

\[
(3.9) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_j} = \sum_{g=1}^{G} \frac{\partial \log f_{\varepsilon}(r(y,x))}{\partial \varepsilon_g} \frac{r_g^g(y,x_1)}{|r_y(y,x)|} + \frac{|r_y(y,x)|_{y_j}}{|r_y(y,x)|}
\]

Taking logs and differentiating both sides of (3.4) with respect to \( x_g \) gives

\[
(3.10) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial x_g} = \frac{\partial \log f_{\varepsilon}(r(y,x))}{\partial \varepsilon_g} \frac{r_g^g(y,x_1)}{|r_y(y,x)|} + \frac{|r_y(y,x)|_{x_g}}{|r_y(y,x)|}
\]

Note that the derivative of \( \log f_{Y|X=x}(y) \) with respect to the endogenous variable \( y_j \) depends on all the \( G \) coordinates of the derivative of \( \log f_{\varepsilon}(r) \) with respect to \( \varepsilon \). In contrast, the derivative of \( \log f_{Y|X=x}(y) \) with respect to each exclusive exogenous variable \( x_g \) depends on only one of those coordinates, \( \partial \log f_{\varepsilon}(r)/\partial \varepsilon_g \). We use the relationship between \( \partial \log f_{Y|X=x}(y)/\partial x_g \) and \( \partial \log f_{\varepsilon}(r)/\partial \varepsilon_g \) to substitute the \( G \) unknown derivatives \( \partial \log f_{\varepsilon}(r)/\partial \varepsilon \) in (3.9). Solving for \( \partial \log f_{\varepsilon}(r)/\partial \varepsilon_g \) in (3.10) and substituting the result into each of the \( \partial \log f_{\varepsilon}(r)/\partial \varepsilon_g \) terms in (3.9) gives

\[
(3.11) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_j} = \sum_{g=1}^{G} \frac{\partial \log f_{Y|X=x}(y)}{\partial x_g} \frac{r_g^g(y,x_1)}{|r_y(y,x)|} + d_{y_j}(y,x)
\]

where

\[
(3.12) \quad d_{y_j}(y,x) = \frac{|r_y(y,x)|_{y_j}}{|r_y(y,x)|} - \sum_{k=1}^{G} \left[ \left( \frac{r_k^g(y,x_k)}{r_k^j(y,x_k)} \right) \frac{|r_y(y,x)|_{x_k}}{|r_y(y,x)|} \right]
\]

Any alternative pair \((\tilde{r}, \tilde{f}_{\varepsilon})\) generating the density \( f_{Y|X=x} \) on \( M \), must satisfy

\[
(3.13) \quad f_{Y|X=x}(y) = \tilde{f}_{\varepsilon}(\tilde{r}(y,x)) \frac{\partial \tilde{r}(y,x)}{\partial y}
\]

Repeating the steps leading to (3.9), (3.10) and (3.11) with \((\tilde{r}, \tilde{f}_{\varepsilon})\) substituting for \((r, f_{\varepsilon})\) we get

\[
(3.14) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_j} = \sum_{g=1}^{G} \frac{\partial \log f_{Y|X=x}(y)}{\partial x_g} \frac{\tilde{r}_g^g(y,x_1)}{|\tilde{r}_y(y,x)|} + \tilde{d}_{y_j}(y,x)
\]

where

\[
(3.15) \quad \tilde{d}_{y_j}(y,x) = \frac{|\tilde{r}_y(y,x)|_{y_j}}{|\tilde{r}_y(y,x)|} - \sum_{k=1}^{G} \left[ \left( \frac{\tilde{r}_k^g(y,x_k)}{\tilde{r}_k^j(y,x_k)} \right) \frac{|\tilde{r}_y(y,x)|_{x_k}}{|\tilde{r}_y(y,x)|} \right]
\]
Subtracting (3.14) from (3.11), we get that for each $j = 1, ..., G$

$$0 = \sum_{g=1}^{G} \left( \frac{\tilde{r}^g_{y_j}(y, x_1)}{\tilde{r}^g_{x_j}(y, x_1)} - \frac{r^g_{y_j}(y, x_1)}{r^g_{x_j}(y, x_1)} \right) \frac{\partial \log f_{Y|X=x}(y)}{\partial x_g} + \left( \tilde{d}_{y_j}(y, x) - d_{y_j}(y, x) \right)$$

as in (3.8). This shows the necessity of (3.8).

We next describe how (3.11) and (3.8) can be used to develop estimators for the identified features of $r$. Let $(\vec{r}, d)$ denote the vector $(\vec{r}^1_{y_1}, \vec{r}^2_{y_1}; ..., \vec{r}^G_{y_1}, ..., \vec{r}^1_{y_G}; d_{y_1}, ..., d_{y_G})$ corresponding to the function $r$, and let $(\vec{\tilde{r}}, \vec{\tilde{d}})$ denote the vector $(\vec{\tilde{r}}^1_{y_1}, ..., \vec{\tilde{r}}^1_{y_G}; ..., \vec{\tilde{r}}^G_{y_1}, ..., \vec{\tilde{r}}^G_{y_G}; \vec{\tilde{d}}_{y_1}, ..., \vec{\tilde{d}}_{y_G})$ corresponding to any alternative function $\tilde{r}$, where for each $g, j$, $\vec{r}^g_{y_j} = r^g_{y_j}(y, x_g)/r^g_{x_j}(y, x_j)$ and $\vec{\tilde{r}}^g_{y_j} = \vec{\tilde{r}}^g_{y_j}(y, x_g)/\vec{\tilde{r}}^g_{x_j}(y, x_j)$ denote the ratios of derivatives of the $g$-function with respect to $y_j$ and $x_g$. The result of Theorem 3.1 together with (3.11) implies that $(\vec{r}, d)$ satisfies the following system of equations

\begin{align*}
g_{y_1} &= \vec{r}^1_{y_1} g_{x_1} + \vec{r}^2_{y_1} g_{x_2} + \cdots + \vec{r}^G_{y_1} g_{x_G} + d_{y_1} \\
&\quad \cdots \\
g_{y_2} &= \vec{r}^1_{y_2} g_{x_1} + \vec{r}^2_{y_2} g_{x_2} + \cdots + \vec{r}^G_{y_2} g_{x_G} + d_{y_2} \\
&\quad \cdots \\
g_{y_G} &= \vec{r}^1_{y_G} g_{x_1} + \vec{r}^2_{y_G} g_{x_2} + \cdots + \vec{r}^G_{y_G} g_{x_G} + d_{y_G}
\end{align*}

(3.16)

Several features of this system of equations deserve mentioning. First, the above system can be interpreted as a system of $G$ equations where only the ratios of derivatives $\vec{r}^g_{y_j} = r^g_{y_j}/r^g_{x_j}$ $(g, j = 1, ..., G)$ and the terms $d_{y_1}, ..., d_{y_G}$ are unknown. Second, note that the unknown elements in this system are elements of only the indirect function $(r^1, ..., r^G)$ . They do not depend on the unknown density of $(\varepsilon_1, ..., \varepsilon_G)$. The density $f_{\varepsilon}$ enters the system only through the known terms, $g_{y_1}, ..., g_{y_G}$ and $g_{x_1}, ..., g_{x_G}$. Moreover, the values of $f_{\varepsilon}$ depend on the values of $(r^1, ..., r^G)$ rather than on the ratios of derivatives of $(r^1, ..., r^G)$ . Hence, the density $f_{\varepsilon}$ has the potential to generate variation on the values of $g_{y_1}, ..., g_{y_G}$ and $g_{x_1}, ..., g_{x_G}$, which is independent of the unknown ratios of derivatives $\vec{r}^g_{y_j} = r^g_{y_j}/r^g_{x_j}$ $(g, j = 1, ..., G)$ and of $d_{y_1}, ..., d_{y_G}$. Third, each of the ratios of derivatives depend on only one $x_g$, while $g_{y_1}, ..., g_{y_G}$ and $g_{x_1}, ..., g_{x_G}$ depend on all the vector $(x_1, ..., x_G)$ . Hence, variation on the coordinates other than $x_g$ has the potential to generate variation on the other elements of the system, while the ratios $r^g_{y_j}/r^g_{x_j}$ stay fixed. This leads to the analysis of conditions on $(r, f_{\varepsilon})$ guaranteeing that values of $(g_{x_1}, ..., g_{x_G})$, which are observable, can be found so that (3.16) can be solved for either the whole vector $(\vec{r}, d) = (\vec{r}^1_{y_1}, ..., \vec{r}^1_{y_G}; ..., \vec{r}^G_{y_1}, ..., \vec{r}^G_{y_G}; d_{y_1}, ..., d_{y_G})$ or for some elements of it.
3.3. Derivation of average derivatives estimators for the model with exclusive regressors

In this section we develop an estimator for the ratios of derivatives, $\tau$, which is based on (3.8) and (3.16). The estimator is based on a characterization of $\tau$ of the least-squares form $(X'X)^{-1}(X'Y)$. Such characterization of $\tau$ employs the fact that (3.16) holds for all values of $(g_1,\ldots,g_G)$ over any subset of $(y, x)$ where $(\tau, d)$ is constant. The elements of the matrices are obtained by averages of cross products of derivatives of $\log f_{Y|X=x}(y)$ over a set $\mathcal{M}$ where the values of $(\tau, d)$ are constant. This characterization of $\tau$ can then be interpreted as the extension to models with simultaneity of a pointwise version of the average derivative estimators developed by Powell, Stock, and Stoker (1989) for the single index model. Estimation of $\tau$ follows by substituting, in the $X'X$ and $XY$ matrices, $f_{Y|X=x}$ by a nonparametric estimator for $f_{Y|X=x}$.

We derive our expression for $\tau$ by characterizing $(\tau, d)$ as the unique solution to the minimization of an integrated square distance between the right hand side and the left hand side of (3.16). The integration set must be a subset of the support of $(Y, X)$ where $(\tau, d)$ is constant. Hence, this set will depend on the restrictions that one assumes on the function $r$. We will provide two sets of restrictions on $r$, each leading to a different integration set. Another restriction on the integration set is that it must contain in its interior $G + 1$ values of the observable variables such that when $g_e$ is evaluated at those values, the only solution to (3.8) is the vector of 0’s. This identification condition guarantees that $(\tau, d)$ is the unique minimizer of the distance function. For each of the two sets of restrictions on $r$, we will provide conditions on $f_e$ guaranteeing that such $G + 1$ values exist. The two sets of restrictions on $r$ that we will consider are stated in Assumptions 3.4 and 3.4’.

**Assumption 3.4:** The inverse function $r^G$ is such for some function $s^G : R \to R$ and all $(y, x_G)$, $r^G(y, x_G) = s^G(y) + x_G$

**Assumption 3.4’:** For each $g = 1,\ldots,G$, the inverse function $r^g$ is such that for some function $s^g : R \to R$ and all $(y, x_g)$, $r^g(y, x_g) = s^g(y) + x_g$

Assumption 3.4 is equivalent to requiring that the units of measurement of $\varepsilon_G$ are tied to those of $x_G$ by

$$-\frac{\partial m^G(y_1,\ldots,y_{G-1},x_G;\varepsilon_G)}{\partial \varepsilon_G} = \frac{\partial m^G(y_1,\ldots,y_{G-1},x_G;\varepsilon_G)}{\partial x_G}$$

$I$ thank a referee for showing that Assumption 5.4 is equivalent to this restriction.
The sets \( \overline{M} \) on which \((\tau, d)\) is constant when Assumptions 3.4 and 3.4’ are satisfied are stated in the following propositions, which are proved in the Appendix.

**Proposition 3.1:** Let \((y, x_{-G}) = (y, x_1, ..., x_{G-1})\) be fixed and given. When Assumptions 3.1-3.3 and 3.4 are satisfied, \((\overline{\tau}, d)\) is constant over the set \( \overline{M} = \{(y, x_{-G}, t_G) \mid t_G \in R \} \).

**Proposition 3.2:** Let \(y\) be fixed and given. When Assumptions 3.1-3.3 and 3.4’ are satisfied, \((\overline{\tau}, d)\) is constant over the set \( \overline{M} = \{(y, t_1, ..., t_G) \mid (t_1, ..., t_G) \in R^G \} \).

To state our assumption on the density \(f_\epsilon\), we will denote by \(A(\epsilon^{(1)}, ..., \epsilon^{(G+1)})\) the matrix of derivatives of \(f_\epsilon\) at \(G+1\) values, \(\epsilon^{(1)}, ..., \epsilon^{(G+1)}\), of \(\epsilon\),

\[
A(\epsilon^{(1)}, ..., \epsilon^{(G+1)}) = \begin{pmatrix}
\frac{\partial \log f_\epsilon(\epsilon^{(1)})}{\partial \epsilon_1} & \frac{\partial \log f_\epsilon(\epsilon^{(1)})}{\partial \epsilon_2} & \cdots & \frac{\partial \log f_\epsilon(\epsilon^{(1)})}{\partial \epsilon_G} & 1 \\
\frac{\partial \log f_\epsilon(\epsilon^{(G)})}{\partial \epsilon_1} & \frac{\partial \log f_\epsilon(\epsilon^{(G)})}{\partial \epsilon_2} & \cdots & \frac{\partial \log f_\epsilon(\epsilon^{(G)})}{\partial \epsilon_G} & 1 \\
\frac{\partial \log f_\epsilon(\epsilon^{(G+1)})}{\partial \epsilon_1} & \frac{\partial \log f_\epsilon(\epsilon^{(G+1)})}{\partial \epsilon_2} & \cdots & \frac{\partial \log f_\epsilon(\epsilon^{(G+1)})}{\partial \epsilon_G} & 1 \\
\end{pmatrix}
\]

**Assumption 3.5:** Let \((y, x_{-G}) = (y, x_1, ..., x_{G-1})\) be given and fixed. There exist \(G+1\), not necessarily known, values \(\epsilon^{(1)}, ..., \epsilon^{(G+1)}\) of \(\epsilon\), such that \(A(\epsilon^{(1)}, ..., \epsilon^{(G+1)})\) is invertible.

**Assumption 3.6:** There exists \(G+1\), not necessarily known values, \(w^{(1)}, ..., w^{(G+1)}\) in the interior of the set \(\overline{M}\), where \((\overline{\tau}, d)\) is constant, such that for each \(k = 1, ..., G+1\), \(\epsilon^{(k)} = r( w^{(k)} )\), where \(\epsilon^{(k)}\) is as in Assumption 3.5.

Assumptions 3.5 and 3.6 require that there exist \(G+1\) values of \(\epsilon\), each corresponding to the value of \(r\) at one point in the set \(\overline{M}\), such that the matrix \(A(\epsilon^{(1)}, ..., \epsilon^{(G+1)})\) is invertible. When the set \(\overline{M}\) is as in Proposition 3.1, the points in the set \(\overline{M}\) can differ only by their value of \(x_G\). Since \(x_G\) enters only in \(r^G\), the \(G+1\) values of \(\epsilon\) that can be used to satisfy the invertibility of \(A(\epsilon^{(1)}, ..., \epsilon^{(G+1)})\) must possess the same values of \(\epsilon_1, ..., \epsilon_{G-1}\). Hence, only changes in the value of \(\epsilon_G\) must generate the \(G+1\) linearly independent rows.

---

6 This assumption is a generalization of the assumptions in Matzkin (2008) and Matzkin (2010), which imposed zero values on some of the elements of this matrix, guaranteeing invertibility. Invertibility conditions on an exclusive regressor model were imposed, in previous works, on the matrix of second order derivatives of \(\log f_\epsilon\). (See Brown, Deb, and Wegkamp (2007) for identification in a semiparametric version of the model in Matzkin (2007b) and Berry and Haile (2011).)
in \( A \left( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \right) \). When the set \( \overline{M} \) is as in Proposition 3.2, the points in the set \( \overline{M} \) may differ in their values of \( (x_1, ..., x_G) \). In this case, the \( G+1 \) values of \( \varepsilon \) that must satisfy the invertibility of \( A \left( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \right) \) may differ in their values of any coordinate, not just in the value of their last coordinate. Normal distributions can satisfy Assumption 3.5 in the latter case, where the values of \( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \) are allowed to differ in all coordinates, but not in the former case, where only the last coordinates of \( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \) are different.

The following propositions relate Assumptions 3.5-3.6 to a testable condition. Let \( w^{(1)}, ..., w^{(G+1)} \) denote \( G+1 \) points in \( \overline{M} \). For each \( j \) and \( k \), we will denote the values of \( g_{x_j} \) at \( w^{(k)} \) by \( g_{x_j}^{(k)} \).

**Condition I.1:** There exist \( w^{(1)}, ..., w^{(G+1)} \) in \( \overline{M} \) such that the matrix

\[
B \left( w^{(1)}, ..., w^{(G+1)} \right) = \begin{pmatrix}
g_{x_1}^{(1)} & g_{x_2}^{(1)} & \cdots & g_{x_G}^{(1)} & 1 \\
g_{x_1}^{(2)} & g_{x_2}^{(2)} & \cdots & g_{x_G}^{(2)} & 1 \\
\vdots & \vdots & & \vdots & \vdots \\
g_{x_1}^{(G+1)} & g_{x_2}^{(G+1)} & \cdots & g_{x_G}^{(G+1)} & 1
\end{pmatrix}
\]

is invertible.

The existence of \( G+1 \) points, \( w^{(1)}, ..., w^{(G+1)} \) in \( \overline{M} \), such that \( B \left( w^{(1)}, ..., w^{(G+1)} \right) \) is invertible implies, by Theorem 3.1, that \( (\tau, d) \) is identified, since for each \( g \), evaluating the equation corresponding to \( y_g \) in (3.8) at \( w^{(1)}, ..., w^{(G)} \) and \( w^{(G+1)} \) generates \( G+1 \) linear independent equations in \( G+1 \) unknowns, whose unique solution is the vector of 0’s. Propositions 3.3 and 3.4 show that Assumptions 3.5 and 3.6 imply that Condition I.1 is satisfied in models satisfying Assumptions 3.1-3.4 or Assumptions 3.1-3.3 and 3.4’, when \( \overline{M} \) is appropriately chosen. They also show that Condition I.1 can be employed to test Assumption 3.5. Any \( G+1 \) points \( w^{(1)}, ..., w^{(G+1)} \) in \( \overline{M} \) for which \( B \left( w^{(1)}, ..., w^{(G+1)} \right) \) is invertible are mapped to \( G+1 \) vectors, \( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \) in \( R^G \), such that \( A \left( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \right) \) is invertible.

**Proposition 3.3:** Suppose that Assumptions 3.1-3.3 and 3.4 are satisfied on \( M \). Let the set \( \overline{M} \) be included in the set \( \{ (y, x_{-G}, t_G) \mid t_G \in R \} \). Then, for all \( w^{(1)}, ..., w^{(G+1)} \) in \( \overline{M} \) and \( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \) in \( R^G \) such that \( \varepsilon^{(k)} = r \left( w^{(k)} \right) \), \( A \left( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \right) \) is invertible if and only if \( B \left( w^{(1)}, ..., w^{(G+1)} \right) \) is invertible.

**Proposition 3.4:** Suppose that Assumptions 3.1-3.3 and 3.4’ are satisfied on \( M \). Let the set \( \overline{M} \) be included in the set \( \{ (y, t_1, ..., t_G) \mid (t_1, ..., t_G) \in R^G \} \). Then, for all \( w^{(1)}, ..., w^{(G+1)} \)
in \( \mathbf{M} \) and \( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \) in \( R^G \) such that \( \varepsilon^{(k)} = r \left( w^{(k)} \right) \), \( A \left( \varepsilon^{(1)}, ..., \varepsilon^{(G+1)} \right) \) is invertible if and only if \( B \left( w^{(1)}, ..., w^{(G+1)} \right) \) is invertible.

We next define a distance function such that \( (\tilde{\tau}, d) \) is the unique solution to the minimization of this distance function. Let \( \mathbf{M} \) denote as before a set where \( (\tau, d) \) is constant. Let \( \mu(y, x) \) denote a specified differentiable nonnegative function such that \( \int_{\mathbf{M}} \mu(y, x) \, d(y, x) = 1 \). For any vector \( (\tilde{\tau}, \tilde{d}) \) generated from an alternative function \( \tilde{\tau} \) satisfying the same assumptions as \( r \), we define the distance function \( S \left( \tilde{\tau}, \tilde{d} \right) \) by

\[
S \left( \tilde{\tau}, \tilde{d} \right) = \int_{\mathbf{M}} \left[ \sum_{g=1}^{G} \left( g_{y_g} - \tilde{\tau}_{y_g} \right) g_{x_1} + \tilde{\tau}_{y_g}^2 g_{x_2} + \cdots + \tilde{\tau}_{y_g}^G g_{x_G} + \tilde{d}_{y_g} \right]^2 \mu(y, x) \, d(y, x)
\]

The value of \( S \left( \tilde{\tau}, \tilde{d} \right) \) is the integrated square distance between the right-hand-side and the left-hand-side of (3.16) when \( (\tau, d) \) is replaced by \( (\tilde{\tau}, \tilde{d}) \). Since \( S \left( \tilde{\tau}, \tilde{d} \right) \geq 0 \) and \( S (\tau, d) = 0 \), \( (\tau, d) \) is a minimizer of \( S (\cdot) \). When Condition I.1 is satisfied for \( w^{(1)}, ..., w^{(G+1)} \) at which \( \mu(y, x) \) is strictly positive, \( (\tau, d) \) is the unique minimizer of \( S (\cdot) \). To express the first \( G \) coordinates, \( \tilde{\tau} \), of the vector \( (\tau, d) \) that solves the First Order Conditions for the minimization of \( S (\cdot) \), we introduce some additional notation. The average of \( g_{y_g} \) and \( g_{x_g} \) over \( \mathbf{M} \) will be denoted, respectively, by

\[
\int_{\mathbf{M}} q_{y_g} = \int_{\mathbf{M}} q_{y_g}(y, x) \, \mu(y, x) \, d(y, x) = \int_{\mathbf{M}} \frac{\partial \log f_{Y|X=x}(y)}{\partial y_g} \, \mu(y, x) \, d(y, x)
\]

\[
\int_{\mathbf{M}} q_{x_g} = \int_{\mathbf{M}} q_{x_g}(y, x) \, \mu(y, x) \, d(y, x) = \int_{\mathbf{M}} \frac{\partial \log f_{Y|X=x}(y)}{\partial x_g} \, \mu(y, x) \, d(y, x)
\]

The averaged centered cross products between \( g_{y_g} \) and \( g_{x_g} \), and between \( g_{x_j} \) and \( g_{x_s} \), will be denoted respectively by

\[
T_{y_g, x_s} = \int_{\mathbf{M}} \left( \int_{\mathbf{M}} q_{y_g}(y, x) \, d(y, x) - \int_{\mathbf{M}} q_{y_g} \right) \left( \int_{\mathbf{M}} q_{x_s}(y, x) \, d(y, x) - \int_{\mathbf{M}} q_{x_s} \right) \mu(y, x) \, d(y, x) \quad \text{and}
\]

\[
T_{x_j, x_s} = \int_{\mathbf{M}} \left( \int_{\mathbf{M}} q_{x_j}(y, x) \, d(y, x) - \int_{\mathbf{M}} q_{x_j} \right) \left( \int_{\mathbf{M}} q_{x_s}(y, x) \, d(y, x) - \int_{\mathbf{M}} q_{x_s} \right) \mu(y, x) \, d(y, x).
\]

The matrices of centered cross products, \( \Pi \) and \( \Gamma \), will be defined by
The matrix of ratios of derivatives, \( \Pi \), will be defined by

\[
\Pi = \begin{pmatrix}
T_{x_1, x_1} & T_{x_2, x_1} & \cdots & T_{x_G, x_1} \\
\vdots & \vdots & \ddots & \vdots \\
T_{x_1, x_G} & T_{x_2, x_G} & \cdots & T_{x_G, x_G}
\end{pmatrix}
\]

and \( \Gamma = \begin{pmatrix}
T_{y_1, x_1} & T_{y_2, x_1} & \cdots & T_{y_G, x_1} \\
\vdots & \vdots & \ddots & \vdots \\
T_{y_1, x_G} & T_{y_2, x_G} & \cdots & T_{y_G, x_G}
\end{pmatrix} \).

The solution of the First Order Conditions for \( \pi \) results in the expression \( \Pi R(\pi) = \Gamma \).

Since \((\pi, d)\) is the unique minimizer, the matrix \( \Pi \) must be invertible, since otherwise one can find two vectors satisfying the First Order Conditions, which must both minimize \( S(\cdot) \).

It follows that \( R(\pi) \) will be given by

\[
(3.17) \quad R(\pi) = \Pi^{-1} \Gamma
\]

The following theorems establish the conditions under which \((\pi, d)\) is the unique minimizer of \( S(\cdot) \), which imply that \( R(\pi) \) is given by (3.17).

**Theorem 3.2:** Let \((y, x_{-G})=(y, x_1, \ldots, x_{G-1})\) be given and let \( \mathcal{M} \) be included in the set \{ \( (y, x_{-G}, t_G) \mid t_G \in R \) \}. Suppose that Assumptions 3.1-3.3 and 3.4-3.6 are satisfied, and that \( \mu(t_G) \) is strictly positive at least at one set of points \( w^{(1)}, \ldots, w^{(G+1)} \) satisfying Condition I.1.

Then, \((\pi, d)\) is the unique minimizer of

\[
S(\tilde{\pi}, \tilde{d}) = \int_{\mathcal{M}} \left[ \sum_{g=1}^{G} \left( g_{y_g} - \tilde{r}_{y_g}^1 g_{x_1} + \tilde{r}_{y_g}^2 g_{x_2} + \cdots + \tilde{r}_{y_g}^G g_{x_G} + \tilde{d}_{y_g} \right)^2 \right] \mu(t_G) \, dt_G
\]

and \( R(\pi) \) is given by (3.17), where the integrals in the definitions of \( \Pi \) and \( \Gamma \) are over \( \mathcal{M} \) and with respect to \( \mu(t_G) \).

**Theorem 3.3:** Let \( y \) be given and let \( \mathcal{M} \) be included in the set \{ \( (y, t_1, \ldots, t_G) \mid (t_1, \ldots, t_G) \in R^G \) \}.
Suppose that Assumptions 3.1-3.3, 3.4’, and 3.5-3.6 are satisfied, and that \( \mu(t_1, ..., t_G) \) is strictly positive is strictly positive at least at one set of points \( w^{(1)}, ..., w^{(G+1)} \) satisfying Condition I.1. Then, \((\tau, d)\) is the unique minimizer of

\[
S(\tilde{\tau}, \tilde{d}) = \int_{\mathcal{M}} \left[ \sum_{g=1}^{G} \left( g_{y_g} - \frac{1}{\tau_{y_g}} g_{x_1} + \frac{2}{\tau_{y_g}^2} g_{x_2} + \cdots + \frac{G}{\tau_{y_g}^G} g_{x_G} + d_{y_g} \right)^2 \right] \mu(t_1, ..., t_G) \, d(t_1, ..., t_G)
\]

and \( R(\tau) \) is given by (3.17), where the integrals in the definitions of \( \Pi \) and \( \Gamma \) are over \( \mathcal{M} \) and with respect to \( \mu(t_1, ..., t_G) \).

To obtain the estimators for \( R(\tau) \), we note that each of the elements, \( T_{y_g,x_s} \) and \( T_{x_j,x_s} \), in the matrices \( \Pi \) and \( \Gamma \), can be estimated from the distribution of the observable variables, by substituting \( f_{Y|X=x}(y) \) in all the expressions by a nonparametric estimator, \( \hat{f}_{Y|X=x}(y) \), for \( f_{Y|X=x}(y) \). Denote such estimators by \( \hat{T}_{y_g,x_s} \) and \( \hat{T}_{x_j,x_s} \), and let \( \hat{\Pi} \) and \( \hat{\Gamma} \) denote the matrices whose elements are, respectively, \( \hat{T}_{y_g,x_s} \) and \( \hat{T}_{x_j,x_s} \). Then, the estimator for the matrix of ratios of derivatives \( R(\tau) \) is defined as

\[\hat{R}(\tau) = \hat{\Pi}^{-1} \hat{\Gamma}\]

### 3.4. Asymptotic Distribution for an average derivative estimator of \( \tau \)

In this section, we derive the asymptotic distributions of the average derivative estimators presented in Subsection 3.3, for the case when the estimator \( \hat{f}_{Y|X=x}(y) \) for the conditional density of \( Y \) given \( X \) is estimated by kernel methods, and the model satisfies Assumptions 3.1-3.3, 3.4’, and 3.5-3.6. Let \( \{Y^i, X^i\}_{i=1}^{N} \) denote \( N \) iid observations generated from \( f_{Y,X} \). The kernel estimator is

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

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

\[
\int \left( \hat{q}_{x_i}(y, x) - \int \hat{q}_{x_i}(y) \right) \left( \hat{q}_{x_j}(y, x) - \int \hat{q}_{x_j}(y) \right) \mu(x) \, dx
\]
where for \( k = 1, \ldots, G \)
\[
\hat{q}_{x_k}(y, x) = \frac{\partial \log \hat{f}_{Y|X=x}(y)}{\partial x_k} \quad \text{and} \quad \int \hat{q}_{x_k}(y) = \int \frac{\partial \log \hat{f}_{Y|X=x}(y)}{\partial x_k} \mu(x) \, dx
\]

Similarly, the element in the \( j \)-th row, \( i \)-th column of our estimator for \( \hat{T}_{YX} \) is
\[
\int \left( \hat{q}_{y_j}(y, x) - \int \hat{q}_{y_j}(y) \right) \left( \hat{q}_{x_i}(y, x) - \int \hat{q}_{x_i}(y) \right) \mu(x) \, dx
\]

where for \( g = 1, \ldots, G \)
\[
\hat{q}_{y_g}(y, x) = \frac{\partial \log \hat{f}_{Y|X=x}(y)}{\partial y_g} \quad \text{and} \quad \int \hat{q}_{y_g}(y) = \int \frac{\partial \log \hat{f}_{Y|X=x}(y)}{\partial y_g} \mu(x) \, dx
\]

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, \( \overline{M}^e \), of \( R^G \), and zero outside \( \overline{M}^e \). We will let \( M^y \) denote a compact set such that the value \( y \) at which we estimate \( r_y \) is an interior point of \( M^y \). Our results use the following assumptions.

**Assumption 3.7:** The density \( f_{Y,X} \) generated by \( f_\varepsilon \) 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 \( x \in \overline{M}^e \), \( f_X(x) > \delta \) and \( f_{Y,X}(y,x) > \delta \).

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

**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 \( \Delta \), and its derivatives of order \( \Delta \) are Lipschitz, where \( \Delta \geq 2 \).

**Assumption 3.10:** The sequence of bandwidths, \( \sigma_N \), is such that \( \sigma_N \to 0 \), \( N\sigma_N^{G+2} \to \infty \), \( \sqrt{N}\sigma_N^{(G/2)+1+s} \to 0 \), \( [N\sigma_N^{2G+2}/\ln(N)] \to \infty \), 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 \( r_r(y) \) the vector in \( R^{G^2} \) formed by stacking the columns of \( r_y(y) \), so that \( rr_y(y) = \text{vec}(r_y(y)) = \)
where the element in 

\( \text{7.7-7.10. Then,} \)

\( r^1(y), \ldots, r^G(y); r^1(y), \ldots, r^G(y); \ldots; r^1_G(y), \ldots, r^G_G(y) \)'.

Let \( \hat{T}_{yy}(y) \) 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 \( \hat{T}T_{xx}(y) = I_G \otimes \hat{T}_{xx}(y) \). The vector \( TT_{yx}(y) \) will be the vector formed by stacking the columns of \( T_{yx}(y) : TT_{yx}(y) = (T_{y_1,x_1}(y), \ldots, T_{y_1,x_1}(y), \ldots, T_{y_G,x_1}(y), \ldots, T_{y_G,x_1}(y), \ldots, T_{y_G,x_1}(y))' \), with its estimator defined by substituting each coordinate by its estimator. For each \( s \), denote

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

and for each \( j, k \), denote

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

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

\[
\sqrt{N} \sigma_{N}^{G+2} \left( \hat{T}T_{yx}(y) - TT_{yx}(y) \right) \rightarrow d N(0, V(y))
\]

where the element in \( V(y) \) corresponding to the covariance between \( T_{y_1,x_s} \) and \( T_{y_k,x_s} \) is

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

Denote by \( \hat{V}(y) \) the matrix whose elements are

\[
\left\{ \int \left( \Delta \partial_x \log \hat{f}_{Y|X=x}(y) \right) \left( \Delta \partial_x \log \hat{f}_{Y|X=x}(y) \right) \left( \frac{\mu(x)^2}{f_{Y,X}(y,x)} \right) dx \right\} \widetilde{KK}_{y_j,y_k}.
\]

The following theorem is proved in the Appendix.

**Theorem 3.4:** Suppose that the model satisfies Assumptions 3.1-3.3, 3.4', 3.5-3.6, and 7.7-7.10. Then,

\[
\sqrt{N} \sigma_{N}^{G+2} (\hat{r}r(y) - rr(y)) \rightarrow d N(0, (TT_{xx}(y))^{-1} V(y) (TT_{xx}(y))^{-1})
\]

and \( (\hat{T}T_{xx}(y))^{-1} \hat{V}(y) (\hat{T}T_{xx}(y))^{-1} \) is a consistent estimator for \( (TT_{xx}(y))^{-1} V(y) (TT_{xx}(y))^{-1} \).
4. A system with two equations and one instrument

4.1. The model

In this section, we turn attention to the model

\[(4.1) \quad Y_1 = m^1(Y_2, \varepsilon_1) \]
\[Y_2 = m^2(Y_1, X_2, \varepsilon_2) \]

where \((Y_1, Y_2, X)\) is observable and \((\varepsilon_1, \varepsilon_2)\) is unobservable. We will provide results on which estimators for the derivative \(\partial m^1(y_2, \varepsilon_1)/\partial y_2\) can be developed. We employ the same assumptions as in the previous model, with the obvious modifications needed to reflect the fact that the first function, \(m^1\), has no regressor \(X_1\). In particular, in the Assumptions 4.1-4.3 below, we assume that \(m^1\) is invertible in \(\varepsilon_1\), \(m^2\) is invertible in \(\varepsilon_2\), that there exist reduced form functions \(h^1\) and \(h^2\), and that the functions are differentiable, to guarantee that for all \((y_1, y_2, x_2)\) on a set \(\mathcal{M}'\) in the support of \((Y, X) = (Y_1, Y_2, X_2)\),

\[(4.2) \quad f_{Y_1,Y_2|X_2=x_2}(y_1, y_2) = f_{\varepsilon_1,\varepsilon_2}(r^1(y_1, y_2), r^1(y_1, y_2, x_2)) \left| \frac{\partial r(y_1, y_2, x_2)}{\partial (y_1, y_2)} \right| \]

Assumption 4.1: Conditional on \(X_2\), the functions \(r = (r^1, r^2)\) and \(h = (h^1, h^2)\) are twice continuously differentiable, 1–1, onto \(\mathbb{R}^2\), and their Jacobian determinants are bounded away from zero. Moreover, the derivative of \(r^2\) with respect to \(x_2\) is bounded away from zero.

Assumption 4.2: \((\varepsilon_1, \varepsilon_2)\) is distributed independently of \(X_2\) with an everywhere positive and twice continuously differentiable density, \(f_\varepsilon\).

Assumption 4.3: \(X_2\) possesses a differentiable density.

4.2. Observational equivalence

We employ the same notation as in the previous section, with \(X = X_2\). Our observational equivalence result for model (4.1) involves functions, \(c(y, x_2)\) and \(\tilde{c}(y, x_2)\), analogous to the
functions $d_{y_j}$ and $\tilde{d}_{y_j}$ in Section 3. These are defined by

\[
c(y, x_2) = -\frac{|r_y(y, x_2)|_{x_2}}{r_{y_1}(y) r_{x_2}(y, x_2)} - \frac{r_{y_2}^1(y) |r_y(y, x_2)|_{y_1}}{r_{y_1}^1(y) |r_y(y, x_2)|} + \frac{|r_y(y, x_2)|_{y_2}}{|r_y(y, x_2)|}, \quad \text{and}
\]
\[
\tilde{c}(y, x_2) = -\frac{|\tilde{r}_y(y, x_2)|_{x_2}}{r_{y_1}(y) r_{x_2}(y, x_2)} - \frac{\tilde{r}_{y_2}^1(y) |\tilde{r}_y(y, x_2)|_{y_1}}{\tilde{r}_{y_1}^1(y) |\tilde{r}_y(y, x_2)|} + \frac{|\tilde{r}_y(y, x_2)|_{y_2}}{|\tilde{r}_y(y, x_2)|}.
\]

(4.3)

Let $\Gamma'$ denote the set of functions $r$ that satisfy Assumption 3.1. We define observational equivalence within $\Gamma'$ over a subset, $M'$, of the support of the vector of observable variables.

**Definition:** Let $M'$ denote a subset of the support of $(Y, X)$, such that for all $(y, x) \in M'$, $f_{Y \mid X}(y, x) > \delta_2$, where $\delta_2$ is any positive constant. Function $\tilde{r} \in \Gamma'$ is observationally equivalent to $r \in \Gamma'$ on $M'$ if there exist densities $f_\varepsilon$ and $f_{\tilde{\varepsilon}}$ satisfying Assumption 4.2 and such that for all $(y, x) \in M,$

\[
f_\varepsilon(r(y, x)) |r_y(y, x)| = f_{\tilde{\varepsilon}}(\tilde{r}(y, x)) |\tilde{r}_y(y, x)|.
\]

(4.4)

Our observational equivalence for the model (4.1) is given in the following theorem.

**Theorem 4.1:** Suppose that $(r, f_\varepsilon)$ generates $f_{Y \mid X}$. A function $\tilde{r} \in \Gamma'$ is observationally equivalent to $r \in \Gamma'$ on $M'$ if and only if for all $(y, x) \in M'$,

\[
0 = \left( \frac{r_{y_2}^1(y)}{r_{y_1}^1(y)} - \frac{\tilde{r}_{y_2}^1(y)}{\tilde{r}_{y_1}^1(y)} \right) g_{y_1} - \left( \frac{|r_y(y, x_2)|_{x_2}}{r_{x_2}^2 r_{y_1}^1} - \frac{|\tilde{r}_y(y, x_2)|_{x_2}}{\tilde{r}_{x_2}^2 \tilde{r}_{y_1}^1} \right) g_{x_2} + (c - \tilde{c})
\]

(4.5)

where for each $g$ and $j$, $r_{y_j}^0(y, x_g) = \tilde{r}_{y_j}^0(y, x_g) = \tilde{r}_{y_j}^0(y, x_g)$, $r_{y_1}^1(y, x_2) = \tilde{r}_{y_1}^1(y, x_2)$, $r_{x_2}^2 = \tilde{r}_{x_2}^2(y, x_2) \mid |r_y(y, x_2)|$, $\mid |\tilde{r}_y(y, x_2)|$, $g_{y_1} = g_{y_1}(y, x_2) = \partial \log f_{Y \mid X=x}(y) / \partial y_1$, $g_{x_2} = g_{x_2}(y, x_2) = \partial \log f_{Y \mid X=x}(y) / \partial x_2$, and where $c = c(y, x_2)$ and $\tilde{c} = \tilde{c}(y, x_2)$ are as defined in (4.3).

When comparing Theorem 4.1 with Theorem 3.1, note that the lack of one exclusive regressor, $X_1$, has reduced the number of equations by one. The derivative of $\log f_{Y \mid X=x}(y)$ with respect to $y_1$ has taken up the place that the derivative of $\log f_{Y \mid X=x}(y)$ with respect to $x_1$ would have taken. The ratio of derivatives of $r^1$ appears as a coefficient of $\partial \log f_{Y \mid X=x}(y) / \partial y_1$. The two ratios of derivatives of $r^2$, which in Theorem 3.1 appeared separately, each as a coefficient of a different derivative of $\log f_{Y \mid X=x}$, appear in (4.5) in one
coefficient, of the form \[ \left[ (r_{y_2}^2 / r_{x_2}^2) - \left( r_{y_1}^2 / r_{x_2}^2 \right) \left( r_{y_1}^1 / r_{y_1}^1 \right) \right]. \]

The proof of Theorem 4.1, presented in the Appendix, proceeds in a way similar to the one used to prove Theorem 3.1. Equation (4.2) is used to obtain an expression for the unobservable \( \partial \log f_\varepsilon (r(y, x)) / \partial \varepsilon \) in terms of the derivatives of the observable \( \log(f_{Y|X=x}(y)) \). The resulting expression is used to substitute \( \partial \log f_\varepsilon (r(y, x)) / \partial \varepsilon \) in the observational equivalence result in Matzkin (2008, Theorem 4.3). After manipulation of the equations, the resulting expression is (4.5). The main difference between both proofs is that in the two equations, one instrument model, the expression for \( \partial \log f_\varepsilon (r(y, x)) / \partial \varepsilon \) involves not only the derivative of \( \log(f_{Y|X=x}(y)) \) with respect to the exogenous variable, \( X \), but also the derivative of \( \log(f_{Y|X=x}(y)) \) with respect to the endogenous variable \( Y_1 \).

The necessity of (4.5) can be shown in a more transparent way as follows. Differentiating both sides of (4.2) with respect to \( y_j \) gives

\[
(4.6) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_1} = \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(r_1, r_2)}{\partial \varepsilon_1} y_1 + \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(r_1, r_2)}{\partial \varepsilon_2} y_1 + \frac{|r_y| |y_1|}{|y|} \\
(4.7) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_2} = \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(r_1, r_2)}{\partial \varepsilon_1} y_2 + \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(r_1, r_2)}{\partial \varepsilon_2} y_2 + \frac{|r_y| |y_2|}{|y|} \\
(4.8) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial x_2} = \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(r_1, r_2)}{\partial \varepsilon_2} x_2 + \frac{|r_y| x_2}{|y|}
\]

Equation (4.8) reflects the fact that \( x_2 \) is exclusive to \( r_2 \). Solving for \( \partial \log f_\varepsilon (r(y, x)) / \partial \varepsilon_2 \) from (4.8) and substituting into (4.6) and (4.7) gives

\[
(4.9) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_1} = \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(r_1, r_2)}{\partial \varepsilon_1} y_1 + \frac{\partial \log f_{Y|X=x}(y)}{\partial x_2} y_2 + \frac{|r_y| y_1}{|y|} - \frac{r_2 x_2}{|y|} \frac{|r_y| x_2}{|y|} \\
(4.10) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_2} = \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(r_1, r_2)}{\partial \varepsilon_1} y_2 + \frac{\partial \log f_{Y|X=x}(y)}{\partial x_2} y_2 + \frac{|r_y| y_2}{|y|} - \frac{r_2 x_2}{|y|} \frac{|r_y| x_2}{|y|}
\]

Solving for \( \partial \log f_\varepsilon (r(y, x)) / \partial \varepsilon_1 \) from (4.9) and substituting into (4.10) gives

\[
(4.11) \quad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_2} = \frac{r_{y_2}^1}{r_{y_1}^1} \frac{\partial \log f_{Y|X=x}(y)}{\partial y_1} + \left[ \frac{r_{y_2}^2}{r_{y_1}^1} \frac{r_{y_1}^2}{r_{x_2}^2} - \frac{r_{y_2}^2}{r_{x_2}^2} \right] \frac{\partial \log f_{Y|X=x}(y)}{\partial x_2} + c(y, x_2)
\]

Rearranging terms, and substituting \( \partial \log f_{Y|X=x}(y) / \partial y_j \) by \( g_{y_j} \) and \( \partial \log f_{Y|X=x}(y) / \partial x_2 \) by \( g_{x_2} \), we get that,

\[
(4.12) \quad g_{y_2}(y, x_2) = \frac{r_{y_2}^1}{r_{y_1}^1} g_{y_1}(y, x_2) + \left[ \frac{r_{y_2}^2}{r_{y_1}^1} \frac{r_{y_1}^2}{r_{x_2}^2} \right] g_{x_2}(y, x_2) + c(y, x_2)
\]
Equation (4.12) together with (4.5) will be referred to in later sections, to build upon them estimators for \( \partial m^1 (y_2, \varepsilon_1) / \partial y_2 = r^1_2 / r^1_1 \). To continue the proof of necessity of (4.5), we note that any alternative pair \((\tilde{r}, f_\tilde{e})\) generating the density \( f_{Y|X=x} \) on \( M \), must satisfy

\[
(4.13) \quad f_{Y|X=x}(y) = f_\tilde{e} (\tilde{r} (y, x)) \left| \frac{\partial \tilde{r}}{\partial y} (y, x) \right|
\]

Repeating the above steps with \((\tilde{r}, f_\tilde{e})\) substituting for \((r, f_e)\) we get

\[
(4.14) \quad g_{y_2}(y, x_2) = \frac{\tilde{r}^1_{y_2}(y)}{\tilde{r}^1_{y_1}(y)} g_{y_1}(y, x_2) + \left[ \frac{|\tilde{r}^1_y(y, x_2)|}{\tilde{r}^1_{y_1} \tilde{r}^2_{x_2}} \right] g_{x_2}(y, x_2) + \tilde{c}(y, x_2)
\]

Subtracting (4.14) from (4.12) gives (4.5). This shows the necessity of (4.5).

The result of Theorem 4.1 can be employed to determine the identification of \( r^1_{y_2} / r^1_{y_1} \) as well as of \( |\tilde{r}^1_y(y, x_2)| / (\tilde{r}^1_{y_1} \tilde{r}^2_{x_2}) \) and of \( c(y, x_2) \). An average derivative estimator for \( r^1_{y_2} / r^1_{y_1} \), analogous to the estimator developed in Section 3, is developed in Matzkin (2012b). We present below two indirect estimators for \( r^1_{y_2} / r^1_{y_1} \).

### 4.3. Indirect estimators in the two equation one instrument model

In this section, we develop estimators, based on equations such as (3.16) and (4.12), which instead of being calculated by inversion and multiplications of matrices, are calculated by finding values of the exogenous variables at which the density of the observable variables provides direct information about the objects of interest. The estimators are based on two steps. In the first step, the value or values of the exogenous variables are found, where the density of the observable variables satisfies some conditions. In the second step, the objects of interest are read off the density of the observable variables at the found values of the exogenous variables. We develop this approach for the two equations, one instrument model,

\[
\begin{align*}
y_1 &= m^1 (y_1, \varepsilon_1) \\
y_2 &= m^2 (y_2, x_2, \varepsilon_2)
\end{align*}
\]

where interest lies on the derivative \( \partial m^1 (y_2, \varepsilon_1) / \partial y_2 \), for the value of \( \varepsilon_1 = r^1 (y_1, y_2) \). We will assume, as in Section 3, that Assumption 3.4 is satisfied. This implies that the value of the vector \( \tilde{b} = (\tilde{b}_1, \tilde{b}_2, \tilde{b}_3) \), with \( \tilde{b}_1 = (r^1_{y_2}(y) / r^1_{y_1}(y)), \tilde{b}_2 = |r^1_y(y, x_2)| / (r^1_{y_1}(y, x_2) r^2_{x_2}(y, x_2)) \), and \( \tilde{b}_3 = c(y, x_2) \), is constant over the set \( \mathcal{M} \subset \{(y, t) | t \in R\} \), where \( c(y, x_2) \) is as defined in (4.3). (See Proposition 3.1.) We will consider two sets of assumptions, which substitute for Assumptions 3.5. Assumption 4.5 implies invertibility of a \( 2 \times 2 \) matrix whose two rows are the gradients of log \( f_\varepsilon \) at two points \((\varepsilon_1, \varepsilon_2^* )\) and \((\varepsilon_1, \varepsilon_2^{**} )\). Assumption 4.5' imposes a condition on the second
order derivatives of $\log f_\varepsilon$ at one point $(\varepsilon_1, \varepsilon_2)$. Assumptions 4.6 and 4.6’ guarantee that there exist points, $x^*$ and $x^{**}$, with $f_{Y|X=x^*}(y) > 0$ and $f_{Y|X=x^{**}}(y) > 0$ and such that the value of $r^2$ at those values of $x$ are mapped into the values $\varepsilon$ satisfying Assumptions 4.5 or 4.5’. In Propositions 4.1 and 4.2, we provide characterizations of these assumptions in terms of conditions on the observable density $f_{Y|X=x}$. We next employ these characterizations together with Theorem 4.1 to obtain expressions for $\partial m^1(y_2; \varepsilon_1)/\partial y_2$ in terms of the values of the derivatives or second derivatives of $f_{Y|X=x}$ at particular values of $x_2$.

**Assumption 4.5:** Let $y$ be given and fixed and let $\varepsilon^1 = r^1(y_1, y_2)$. There exist two distinct values $\varepsilon_2^*(\varepsilon_1)$ and $\varepsilon_2^{**}(\varepsilon_1)$ of $\varepsilon_2$ such that

\[
\frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_2} = \frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^{**}(\varepsilon_1))}{\partial \varepsilon_2}, \text{ and}
\]

\[
\frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_1} \neq \frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^{**}(\varepsilon_1))}{\partial \varepsilon_1}.
\]

**Assumption 4.5’:** Let $y$ be given and fixed and let $\varepsilon^1 = r^1(y_1, y_2)$. There exist a value $\varepsilon_2^*(\varepsilon_1)$ of $\varepsilon_2$ such that

\[
\frac{\partial^2 \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_2^2} = 0 \quad \text{and} \quad \frac{\partial^2 \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_2 \partial \varepsilon_1} \neq 0.
\]

**Assumption 4.6:** Let $y$ be given and fixed and let $\varepsilon^1 = r^1(y_1, y_2)$. There exists distinct values $x^*$ and $x^{**}$ such that $(y, x^*), (y, x^{**}) \in \mathcal{M}$ and 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 $\varepsilon_2^{**} = r^2(y_1, y_2, x^{**})$.

**Assumption 4.6’:** Let $y$ be given and fixed and let $\varepsilon^1 = r^1(y_1, y_2)$. There exists a value $x^*$ such that $(y, x^*) \in \mathcal{M}$ and such that for $\varepsilon_2^*(\varepsilon_1)$ as in Assumption 4.5’, $\varepsilon_2^* = r^2(y_1, y_2, x^*)$.

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 $\varepsilon_2$ given $\varepsilon_1$ when $\varepsilon_2 = 0$ coincide with the values and derivatives of a conditional density of the form $f_{\varepsilon_2|\varepsilon_1}(\varepsilon_2) = c \exp(-a(\varepsilon_1) \varepsilon_2^3)$, for some $c$ which could depend on $\varepsilon_1$. The conditional density $f_{\varepsilon_2|\varepsilon_1}(\varepsilon_2)$

---

7This assumption is critical for builing our indirect estimator, and, besides the difference in models, it is the main distinction of our identification approach from the identification approach used in Chiappori and Komunjer (2009). (Berry and Haile (2011) showed constructive identification of the model in Matzkin (2008, Section 4.2) assuming intertibility at a point of the matrix of second order derivatives of $\log(f_\varepsilon)$.)
is not restricted to possess this form for values of $\varepsilon_2$ other than $\varepsilon_2 = 0$. An example that satisfies Assumption 4.5' is where the conditional density of $\varepsilon_2$ given $\varepsilon_1$ has the form $f_{\varepsilon_2|\varepsilon_1}(\varepsilon_2) = c \exp(-a(\varepsilon_1) \varepsilon_2^3 - b \varepsilon_2^2)$ 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 $\varepsilon_2$. 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' in Subsection 4.4, to guarantee desired asymptotic properties for our estimators of $\partial m^1(y_2, \varepsilon_1) / \partial y_2$. The following propositions provide characterizations of Assumptions 4.5-4.6 and 4.5'-4.6' in terms of conditions on $\varphi$.

Proposition 4.1: Let $y$ be given and fixed and let $\varepsilon^1 = r^1(y_1, y_2)$. Suppose that Assumptions 4.1-4.3 and 3.4 are satisfied. Assumptions 4.5-4.6 are satisfied if and only if there exist $x^{(1)}$ and $x^{(2)}$ such that $(y, x^{(1)}), (y, x^{(2)}) \in \overline{M}$,

$$
(A.5) \quad \frac{\partial \log f_{Y|X=x^*}(y)}{\partial x} = \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial x} \quad \text{and} \quad \frac{\partial \log f_{Y|X=x^*}(y)}{\partial y_1} \neq \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1}
$$

Proposition 4.2: Let $y$ be given and fixed and let $\varepsilon^1 = r^1(y_1, y_2)$. Suppose that Assumptions 4.1-4.3 and 3.4 are satisfied. Assumptions 4.5'-4.6' are satisfied if and only if there exists $x^{(1)}$ such that $(y, x^{(1)}) \in \overline{M}$,

$$
(A.6) \quad \frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x^2} = 0 \quad \text{and} \quad \frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x \partial y_1} \neq 0
$$

We next employ the implications of Propositions 4.1 and 4.2, together with Theorem 4.1 and equation (4.12), to obtain expressions for $\partial m^1(y_2, \varepsilon_1) / \partial y_2$ in terms of ratios of differences of derivatives of $\log f_{Y|X=x}$ or in terms of ratios of second derivatives of $\log f_{Y|X=x}$. We state these expressions in Theorems 4.2 and 4.3.

Theorem 4.2: Let $y$ be given and fixed and let $\varepsilon^1 = r^1(y_1, y_2)$. Suppose that Assumptions 4.1-4.3, and 3.4 are satisfied. Let $x^*$ and $x^{**}$ be any distinct values of $X_2$ such that $(y, x^*), (y, x^{**}) \in \overline{M}$,
\[
\frac{\partial \log f_{Y|X=x^*}(y)}{\partial x} = \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial x}
\quad \text{and} \quad \frac{\partial \log f_{Y|X=x^*}(y)}{\partial y_1} \neq \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1}
\]

Then,
\[
\frac{\partial m_1(y_2, \varepsilon_1)}{\partial y_2} = -\frac{r_{y_2}^1(y)}{r_{y_1}^1(y)} = \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_2} - \frac{\partial \log f_{Y|X=x^*}(y)}{\partial y_1}
\]

**Proof:** Let \( \overline{b} = (\overline{b}_1, \overline{b}_2, \overline{b}_3) \), where \( \overline{b}_1 = r_{y_2}^1(y) / r_{y_1}^1(y), \overline{b}_2 = |r_y(y, x_2)| / (r_{y_1}^1(y, x_2) r_{x_2}^2(y, x_2)) \), and where \( \overline{b}_3 = c(y, x_2) \), is as defined in (4.3). By Assumptions 4.1-4.3, \( \overline{b} \) satisfies (4.12). By Assumption 6.4, \( \overline{b} \) is constant over the set \( \{(y, t) | t \in R\} \), since Assumption 3.4 implies that for all \( x_2, r_{x_2}^2(y, x_2) = 1 \) and \( |r_y(y, x_2)| \) is not a function of \( x_2 \). For any two values \( x_2^{(1)} \) and \( x_2^{(2)} \) of \( x \), such that \( (y, x_2^{(1)}), (y, x_2^{(2)}) \in \overline{M} \), let \( g_{y_2}^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial y_2, g_{y_1}^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial y_1 \), and \( g_{x_2}^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial x_2 \). By (4.12),
\[
g_{y_2}^{(1)} = \overline{b}_1 g_{y_1}^{(1)} + \overline{b}_2 g_{x_2}^{(1)} + \overline{b}_3
\]
and
\[
g_{y_2}^{(2)} = \overline{b}_1 g_{y_1}^{(2)} + \overline{b}_2 g_{x_2}^{(2)} + \overline{b}_3
\]
Subtracting one from the other, we get that
\[
(g_{y_2}^{(1)} - g_{y_2}^{(2)}) = \overline{b}_1 (g_{y_1}^{(1)} - g_{y_1}^{(2)}) + \overline{b}_2 (g_{x_2}^{(1)} - g_{x_2}^{(2)})
\]
If \( (g_{x_2}^{(1)} - g_{x_2}^{(2)}) = 0 \), the equation becomes
\[
(g_{y_2}^{(1)} - g_{y_2}^{(2)}) = \overline{b}_1 (g_{y_1}^{(1)} - g_{y_1}^{(2)})
\]
If in addition, \( (g_{y_1}^{(1)} - g_{y_1}^{(2)}) \neq 0 \), we have that
\[
\overline{b}_1 = (g_{y_1}^{(1)} - g_{y_1}^{(2)})^{-1} (g_{y_2}^{(1)} - g_{y_2}^{(2)})
\]
Note that \( \overline{b}_1 = r_{y_2}^1(y) / r_{y_1}^1(y) = -\partial m_1(y_2, \varepsilon_1) / \partial y_2 \). Hence,
\[
\frac{\partial m_1(y_2, \varepsilon_1)}{\partial y_2} = -\left(\frac{r_{y_2}^1(y)}{r_{y_1}^1(y)}\right) = -\frac{(g_{y_2}^{(1)} - g_{y_2}^{(2)})}{(g_{x_2}^{(1)} - g_{x_2}^{(2)})}
\]
Replacing $x^{(1)}$ by $x^*$ and $x^{(2)}$ by $x^{**}$, equation (8.6) follows. This completes the proof of Theorem 4.2.

**Theorem 4.3:** Let $y$ be given and fixed and let $\varepsilon^1 = r^1(y_1, y_2)$. Suppose that Assumptions 4.1-4.3, and 3.4 are satisfied. Let $x^*$ be a value of $X_2$ such that $(y, x^*) \in \overline{M}$,

\[
(A.9) \quad \frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x \partial x} = 0 \quad \text{and} \quad \frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x \partial y_1} \neq 0
\]

Then,

\[
(A.10) \quad \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = -\frac{\frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x \partial y_2}}{\frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x \partial y_1}}.
\]

**Proof:** As in the proof of Theorem 8.1, we let $\tilde{b} = (\tilde{b}_1, \tilde{b}_2, \tilde{b}_3)$, where $\tilde{b}_1 = r^1_{y_2}(y)/r^1_{y_1}(y)$, $\tilde{b}_2 = |r_y(y, x_2)| / (r^1_{y_1}(y, x_2) r^2_{x_2}(y, x_2))$, and where $\tilde{b}_3 = c(y, x_2)$, is as defined in (4.3). We note that Assumption 3.4 implies that $\tilde{b}$ is constant over the set $\{ (y, t) | t \in R \}$. Hence by Assumptions 4.1-4.3 and 3.4, it follows by (4.12) that for any value $x^{(1)}_2$ such that $(y, x^{(1)}_2) \in \overline{M}$,

\[
g_{y_2}^{(1)} = \tilde{b}_1 g_{y_1}^{(1)} + \tilde{b}_2 g_{x_2}^{(1)} + \tilde{b}_3
\]

where $g_{y_2}^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial y_2$, $g_{y_1}^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial y_1$, and $g_{x_2}^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial x_2$.

Since $\tilde{b}$ is constant over $x$, taking derivatives of this equation with respect to $x$ gives

\[
g_{x_2, y_2}^{(1)} = \tilde{b}_1 g_{x_2, y_1}^{(1)} + \tilde{b}_2 g_{x_2, x_2}^{(1)}
\]

where $g_{x_2, y_2}^{(k)} = \partial^2 \log f_{Y|X=x^{(k)}}(y)/\partial x_2 \partial y_2$, $g_{x_2, y_1}^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial x_2 \partial y_1$, and $g_{x_2, x_2}^{(k)} = \partial^2 \log f_{Y|X=x^{(k)}}(y)/\partial x_2 \partial x_2$. If $g_{x_2, x_2}^{(1)} = 0$ and $g_{x_2, y_1}^{(1)} \neq 0$, we have that

\[
\tilde{b}_1 = (g_{x_2, y_1}^{(1)})^{-1} (g_{x_2, y_2}^{(1)})
\]

Since $\tilde{b}_1 = r^1_{y_2}(y)/r^1_{y_1}(y) = -\partial m^1(y_2, \varepsilon_1)/\partial y_2$, it follows that

\[
\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = -\left( \frac{r^1_{y_2}}{r^1_{y_1}} \right) = -\frac{g_{x_2, y_2}^{(1)}}{g_{x_2, y_1}^{(1)}}
\]

Replacing $x^{(1)}$ by $x^*$, equation (A.10) follows. This completes the proof of Theorem 4.3.
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

$$\frac{\partial^2 \log f_Y|_{X=x}(y)}{\partial x \partial x} = 0$$

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

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = -\frac{\partial^2 \log f_Y|_{X=x^*}(y)}{\partial x \partial y_2} - \frac{\partial^2 \log f_Y|_{X=x^*}(y)}{\partial x \partial y_2}$$

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

$$\frac{\partial \log f_Y|_{X=x^*}(y)}{\partial x} = \frac{\partial \log f_Y|_{X=x^{**}}(y)}{\partial x} = 0$$

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

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = -\frac{\partial \log f_Y|_{X=x^*}(y)}{\partial y_2} - \frac{\partial \log f_Y|_{X=x^{**}}(y)}{\partial y_2} - \frac{\partial \log f_Y|_{X=x^*}(y)}{\partial y_1} - \frac{\partial \log f_Y|_{X=x^{**}}(y)}{\partial y_1}$$

Again, when $\partial \log f_Y|_{X=x}(y)/\partial y_1$, $\partial \log f_Y|_{X=x}(y)/\partial y_2$, and $\partial \log f_Y|_{X=x}(y)/\partial x$ 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.
4.4. Asymptotic properties of the Indirect Estimators

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

**Assumption 4.7:** The density $f_{\varepsilon}$ and the density $f_{Y,X}$ generated by $f_{\varepsilon}$ 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 $\partial^2 \log f_{\varepsilon} (r^1(y_1, y_2), r^2(y_1, y_2, x')) / \partial \varepsilon^2 = 0$, there exist a neighborhood $B'_{y,x}$ of $(y_1, y_2, x')$ and $B'_x$ of $x'$ such that the density $f_{Y,X}(x)$ and the density $f_{Y,X}(y, x) = f_{\varepsilon} (r^1(y_1, y_2), r^2(y_1, y_2, x)) | r_y (y_1, y_2, x) | f_X(x)$ are uniformly bounded away from zero on, respectively, $B'_x$ and $B'_{y,x}$ and $\partial^3 \log f_{\varepsilon} (r^1(y_1, y_2), r^2(y_1, y_2, x)) / \partial \varepsilon^3$ is bounded away from zero on those neighborhoods.

**Assumption 4.9:** For any $x'$ such that $\partial^2 \log f_{\varepsilon} (r^1(y_1, y_2), r^2(y_1, y_2, x')) / \partial \varepsilon^2 = 0$, $\partial^2 \log f_{\varepsilon} (r^1(y_1, y_2), r^2(y_1, y_2, x')) / \partial \varepsilon_1 \partial \varepsilon_2$ is uniformly bounded away from zero 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 $s + 5 \leq d$, is differentiable of order $\Delta$, and its derivatives of order $\Delta$ are Lipschitz, where $\Delta \geq 5$.

**Assumption 4.11:** The sequence of bandwidths, $\sigma_N$, is such that $\sigma_N \to 0$, $\sqrt{N \sigma_N^{7+2s}} \to 0$, $\sqrt{N \sigma_N^2} \to \infty$, $[N \sigma_N^4 / \ln(N)] \to \infty$, and $\sqrt{N \sigma_N^2 \left[ \sqrt{\ln(N)} / N \sigma_N^4 + \sigma_N^2 \right]} \to 0$.

Assumptions 4.7, 4.10 and 4.11 are standard for derivations of asymptotic results of kernel estimators. Assumptions 4.8 and 4.9 are made to guarantee appropriate asymptotic behavior of the estimator for the value $x^*$ at which $\partial m_1 (y_2, \varepsilon_1) / \partial y_2$ is calculated. Define $K_{y_1,x}(\tilde{y}_1, \tilde{y}_2, x) = \frac{\partial^2 K(\tilde{y}_1, \tilde{y}_2, x)}{\partial y_1 \partial x}$, $K_{y_2,x}(\tilde{y}_1, \tilde{y}_2, x) = \frac{\partial^2 K(\tilde{y}_1, \tilde{y}_2, x)}{\partial y_2 \partial x}$, and $K_{x,x}(\tilde{y}_1, \tilde{y}_2, x) = \frac{\partial^2 K(\tilde{y}_1, \tilde{y}_2, x)}{\partial x^2}$. Let $\bar{K}(\tilde{y}_1, \tilde{y}_2, x)$ denote the $3 \times 1$ vector $(K_{y_1,x}(\tilde{y}_1, \tilde{y}_2, x), K_{y_2,x}(\tilde{y}_1, \tilde{y}_2, x), K_{x,x}(\tilde{y}_1, \tilde{y}_2, x))'$. Define the vector $\omega(y, x^*) = (\omega_1, \omega_2, \omega_3)'$ where

$$
\omega_1 = \frac{\partial^2 \log f_{Y,X}(y^*, x^*)}{\partial y_2 \partial x} \left[ \frac{\partial^2 \log f_{Y,X}(y, x^*)}{\partial y_1 \partial x} \right]^2 f_{Y,X}(y, x^*), \\
\omega_2 = -\frac{1}{\partial^2 \log f_{Y,X}(y^*, x^*) f_{Y,X}(y, x^*)}.
$$
Theorem 4.4: Suppose that the model satisfies Assumptions 4.1-4.11. Let the estimator for \( \partial m^1 (y_2, \varepsilon_1) / \partial y_2 \) be as defined in (4.2). Then,

\[
\sqrt{N} \sigma_N^7 \left( \partial m^1 (y_2, \varepsilon_1) / \partial y_2 - \partial m^1 (y_2, \varepsilon_1) / \partial y_2 \right) \rightarrow_d N \left( 0, \tilde{V} \right)
\]

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

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

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

Assumption 4.9': For any two values \( x', x'' \) such that \( \partial \log f_\varepsilon \left( r^1 (y_1, y_2), r^2 (y_1, y_2, x') \right) / \partial \varepsilon_2 = 0 \), and \( \partial \log f_\varepsilon \left( r^1 (y_1, y_2), r^2 (y_1, y_2, x'') \right) / \partial \varepsilon_2 = 0 \), \( (\partial \log f_{\varepsilon_1, \varepsilon_2} (r^1, r^2) / \partial \varepsilon_1 - \partial \log f_{\varepsilon_1, \varepsilon_2} (r^1, r^2) / \partial \varepsilon_1) \) 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 \( \Delta \), and its derivatives of order \( \Delta \) are Lipschitz, where \( \Delta \geq 4 \).
Assumption 4.11': The sequence of bandwidths, \( \sigma_N \), is such that \( \sqrt{N \sigma_N^2} \to \infty \), \( \sqrt{N \sigma_N^2} \to 0 \), and \( \sqrt{N \sigma_N^2} \left[ \ln(N)/N \sigma_N^2 + \sigma_N^2 \right] \to 0 \).

Define \( \tilde{K}_{y_1}(\tilde{y}_1, \tilde{y}_2, x) = \frac{\partial K(\tilde{y}_1, \tilde{y}_2, x)}{\partial y_1} \), \( \tilde{K}_{y_2}(\tilde{y}_1, \tilde{y}_2, x) = \frac{\partial K(\tilde{y}_1, \tilde{y}_2, x)}{\partial y_2} \), and \( \tilde{K}_x(\tilde{y}_1, \tilde{y}_2, x) = \frac{\partial K(\tilde{y}_1, \tilde{y}_2, x)}{\partial x} \).

Let \( \tilde{K}(\tilde{y}_1, \tilde{y}_2, x) \) denote the \( 3 \times 1 \) vector \( (\tilde{K}_{y_1}(\tilde{y}_1, \tilde{y}_2, x), \tilde{K}_{y_2}(\tilde{y}_1, \tilde{y}_2, x), \tilde{K}_x(\tilde{y}_1, \tilde{y}_2, x))' \). Define the vectors \( \omega^1 = (\omega_1^1, \omega_1^2, \omega_1^3) \) and \( \omega^2 = (\omega_2^1, \omega_2^2, \omega_2^3) \) by

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

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

\[
\omega_1^3 = \frac{\partial f_{Y,X}(y, x_1^*)}{\partial y_1} f_{Y,X}(y, x_1^*) - \frac{\partial f_{Y,X}(y, x_1^*)}{\partial y_2} f_{Y,X}(y, x_2^*) ;
\]

\[
\omega_2^1 = \frac{\partial f_{Y,X}(y, x_1^*)}{\partial y_1} f_{Y,X}(y, x_1^*) - \frac{\partial f_{Y,X}(y, x_1^*)}{\partial y_2} f_{Y,X}(y, x_1^*) \right]^2 \frac{\partial^2 f_{Y,X}(y, x_1^*)}{\partial x^2} f_{Y,X}(y, x_1^*) ;
\]

\[
\omega_2^2 = \frac{\partial f_{Y,X}(y, x_1^*)}{\partial y_1} f_{Y,X}(y, x_1^*) - \frac{\partial f_{Y,X}(y, x_1^*)}{\partial y_2} f_{Y,X}(y, x_1^*) ;
\]

\[
\omega_2^3 = \frac{\partial f_{Y,X}(y, x_1^*)}{\partial y_1} f_{Y,X}(y, x_2^*) - \frac{\partial f_{Y,X}(y, x_1^*)}{\partial y_2} f_{Y,X}(y, x_1^*) \right] \frac{\partial^2 f_{Y,X}(y, x_1^*)}{\partial x^2} f_{Y,X}(y, x_1^*) .
\]

Define

\[
\nabla = \omega_1^1 \left[ \int \tilde{K}(\tilde{y}_1, \tilde{y}_2, x) \tilde{K}(\tilde{y}_1, \tilde{y}_2, x)' d(\tilde{y}_1, \tilde{y}_2, x) \right] \omega_1^1 f_{Y,X}(y, x_1^*) \]

\[
+ \omega_2^2 \left[ \int \tilde{K}(\tilde{y}_1, \tilde{y}_2, x) \tilde{K}(\tilde{y}_1, \tilde{y}_2, x)' d(\tilde{y}_1, \tilde{y}_2, x) \right] \omega_2^2 f_{Y,X}(y, x_2^*)
\]

In the Appendix we prove

**Theorem 4.5:** Suppose that Assumptions 4.1-4.4, and 4.5'-4.11' are satisfied. Define
\[ \frac{\partial m^1(y_1, \varepsilon_1)}{\partial y_2} \text{ as in (4.3). Then,} \]
\[ \sqrt{N} \sigma_N^2 \left( \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} - \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} \right) \to N \left( 0, \nabla \right) \]

5. 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.

6. Appendix

We first define some notation. Let \( r_y \) denote the \( G \times G \) matrix whose element in the \( i \)-th row and \( j \)-th column is \( r^i_j \). Let \( r_x \) denote the \( G \times G \) diagonal matrix whose element in the \( i \)-th diagonal place is \( r^i_i \). The \( G \times G \) matrix \( \tau_y \) whose element in the \( i \)-th row and \( j \)-th column is the ratio \( r^i_j \) will then be equal to \( (r_x)^{-1} r_y \). We will denote by \( q_e \) the \( G \times 1 \) vector whose \( i \)-th element is \( \partial \log f_e(r(y, x))/\partial \varepsilon_i \). The \( G \times 1 \) vectors \( \gamma_y \) and \( \gamma_x \) will be the vectors whose \( i \)-th element are, respectively, \( \partial \log | r_y(y, x) | / \partial y_i \) and \( \partial \log | r_y(y, x) | / \partial x_i \). For any function \( \tilde{r} \) and any density \( f_\varepsilon \), the matrices \( \tilde{r}_y \), \( \tilde{r}_x \) and \( \tilde{r}_y \) and the vectors \( \tilde{q}_y \), \( \tilde{\gamma}_y \), and \( \tilde{\gamma}_x \) will be defined analogously. We will denote the \( G \times 1 \) vectors \( (g_{y_1}, ..., g_{y_G}) \) and \( (g_{x_1}, ..., g_{x_G}) \) by, respectively, \( g_y \) and \( g_x \).

The following Lemma will be used in the proofs of Theorems 3.1 and 4.1.
Lemma (Matzkin (2008)): Suppose that \((r, r_x) \in (\Gamma \times \Phi)\) generate \(f_{Y,X}\) on \(M\) and \(\tilde{r} \in \Gamma\). Then, there exists \(r_x \in \Phi\) such that \((\tilde{r}, r_x)\) generate \(f_{Y,X}\) on \(M\) if and only if for all \((y, x) \in M\)

\[
(L.1) \quad \left( r_x' - \tilde{r}_x' \left( \tilde{r}_y' \right)^{-1} r_y' \right) q_x = -\left( \gamma_x - \tilde{\gamma}_x \right) + \tilde{r}_x' \left( \tilde{r}_y' \right)^{-1} \left( \gamma_y - \tilde{\gamma}_y \right)
\]

Proof: The proof follows by restricting the definition of observational equivalence in Matzkin (2008) to observational equivalence on \(M\). The arguments used in Matzkin (2008) to prove Theorem 3.1 and Theorem 3.2 can be applied in the same way as in Matzkin (2008) to prove the Lemma. The expression in \((L.1)\) is the transpose of \((3.8)\) in the statement of Theorem 3.1 in Matzkin (2008).

Proof of Theorem 3.1: We will show that in the exclusive regressors model, when \(r_x\) and \(r\) generate \(f_{Y,X}\), and when \(r_x\) and \(\tilde{r}_x\) are invertible, diagonal, \(G \times G\) matrices, \((3.8)\) is equivalent to \((L.1)\). Hence, by Lemma 1, it will follow that \(\tilde{r}\) is observationally equivalent to \(r\) if and only if \((3.8)\) is satisfied. For this, we first note that since \((r_x, r)\) generate \(f_{Y,X}\) on \(M\), for all \((y, x) \in M\),

\[
f_{Y|X=x}(y) = f_x(r(y, x)) |r_y(y, x)|
\]

Taking logs and differentiating both sides with respect to \(x\), without writing the arguments of the functions explicitly, we get

\[
g_x = r_x' q_x + \gamma_x
\]

Since by assumption \(r_x\) is invertible, we can solve uniquely for \(q_x\), getting

\[
(T.3.1) \quad \left( r_x' \right)^{-1} (g_x - \gamma_x) = q_x
\]

Replacing \(q_x\) in \((L.1)\) for the expression for \(q_x\) in \((T.3.1)\), we get

\[
(T.3.2) \quad \left[ r_x' - \tilde{r}_x' \left( \tilde{r}_y' \right)^{-1} r_y' \right] \left( r_x' \right)^{-1} (g_x - \gamma_x) = -\left( \gamma_x - \tilde{\gamma}_x \right) + \left[ \tilde{r}_x' \left( \tilde{r}_y' \right)^{-1} \right] \left( \gamma_y - \tilde{\gamma}_y \right)
\]

Since \(r_x' (r_x')^{-1} = I\), the RHS of \((T.3.2)\) can be expressed as

\[
\left[ 1 - \tilde{r}_x' \left( \tilde{r}_y' \right)^{-1} r_y' \left( r_x' \right)^{-1} \right] g_x - \left[ 1 - \tilde{r}_x' \left( \tilde{r}_y' \right)^{-1} r_y' \left( r_x' \right)^{-1} \right] \gamma_x
\]

34
Hence, premultiplying both sides of (T.3.2) by $-\tilde{r}_y' (\tilde{r}_x')^{-1}$, we get
\[
\begin{align*}
[r'_y (r_x')^{-1} - \tilde{r}_y' (\tilde{r}_x')^{-1}] g_x - \left[r'_y (r_x')^{-1} - \tilde{r}_y' (\tilde{r}_x')^{-1}\right] \gamma_x &= \left[(\tilde{r}_y - \tilde{r}_y) (\tilde{r}_x')^{-1} \right] (\gamma_x - \tilde{\gamma}_x) - (\gamma_y - \tilde{\gamma}_y)
\end{align*}
\]
Subtracting $[\tilde{r}_y' (\tilde{r}_x')^{-1}] \gamma_x$ from both sides of the equality sign, and rearranging terms, the last expression can be written as
\[
[ r'_y (r_x')^{-1} - \tilde{r}_y' (\tilde{r}_x')^{-1} ] g_x + [ \gamma_y - r'_y (r_x')^{-1} \gamma_x ] - [ \tilde{\gamma}_y - \tilde{r}_y' (\tilde{r}_x')^{-1} \tilde{\gamma}_x ] = 0
\]
Note that the vector $d_y = (d_{y_1}, ..., d_{y_G})' = \gamma_y - r'_y (r_x')^{-1} \gamma_x$ and $\tilde{d}_y = (\tilde{d}_{y_1}, ..., \tilde{d}_{y_G})' = \tilde{\gamma}_y - \tilde{r}_y' (\tilde{r}_x')^{-1} \tilde{\gamma}_x$. Hence, we have obtained that in the exclusive regressors model, (L.1) is equivalent to
\[
[ r'_y (r_x')^{-1} - \tilde{r}_y' (\tilde{r}_x')^{-1} ] g_x + d_y - \tilde{d}_y = 0
\]
This is exactly (3.8) in Section 3.

**Proof of Propositions 3.1 and 3.2:** Assumption 3.1 implies that for each $g$, the ratios of derivatives, $r^g_{y_j}(y, x_g)/r^g_{x_g}(y, x_g)$, ($j = 1, ..., G$) depend only of $y$ and $x_g$. When Assumption 5.4 is satisfied, the ratios of derivatives of $r^G$ are given by $s^G_{y_j}(y)$ since the derivative of $r^G$ with respect to $x_G$ is 1. Hence, $\tau$ is constant over $\overline{M} = \{(y, x_{-G}, t_G) | t_G \in R\}$. Moreover, the Jacobian determinant $|r_y|$, which is a function of the derivatives of the $r^g$ functions with respect to $y$, does not depend on $x_G$ either, since $x_G$ only affects $r^G$ and the derivative of $r^G$ with respect to $y_j$ is not a function of $x_G$. Hence, for each $g$, all the terms in $d_{y_g}$, defined in (3.6), are constant over $x_G$. It then follows that $(d_{y_1}, ..., d_{y_G})$ is constant over $\overline{M} = \{(y, x_{-G}, t_G) | t_G \in R\}$. A similar reasoning shows that when Assumption 3.4' is satisfied, $(\tau, d)$ is constant over $\overline{M} = \{(y, t_1, ..., t_G) | (t_1, ..., t_G) \in R^G\}$.

**Proof of Proposition 3.3:** Let $x^{(1)}, ..., x^{(G+1)}$ be such that $u^{(1)} = (y, x_{-G}, x_G^{(k)}), ..., u^{(G+1)} = (y, x_{-G}, x_G^{(G+1)}) \in \overline{M}$ and let $\varepsilon^{(1)}, ..., \varepsilon^{(G+1)}$ be such that for each $k = 1, ... G + 1$,
\[
\varepsilon^{(k)} = (r^1(y, x_1), ..., r^{G-1}(y, x_{G-1}), r^G(y, x_G^{(k)})).
\]
We will show that

\[ A \left( \varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)} \right) \text{ is invertible} \iff B \left( w^{(1)}, \ldots, w^{(G+1)} \right) \text{ is invertible} \]

Note that the relationship between the element in the \( j \)-th row and \( k \)-th column of \( B \left( w^{(1)}, \ldots, w^{(G+1)} \right) \) and the \( j \)-th row and \( k \)-th column of \( A = A \left( \varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)} \right) \) is given by

\[
g_{x_k}(y, x_{-G}, x_G^{(j)}) = \frac{\partial \log f_{\varepsilon}(\varepsilon_{-G}, r^G(y, x_{-G}, x_G^{(j)}))}{\partial \varepsilon_k} r_{x_k}^{k}(y, x_k) + \frac{\partial \log |r_{y}(y, x)|}{\partial x_k}.
\]

Suppose that \( A = A \left( \varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)} \right) \) is invertible. The value of \( \partial \log |r_{y}(y, x)|/\partial x_k \) is constant across the \( k \)-th column, because \( |r_{y}(y, x)| \) is constant over \( w^{(j)} \) \((j = 1, \ldots, G+1)\). Multiplying the \( G+1 \) column of \( B \), which is a column of 1’s, by \( \partial \log |r_{y}(y, x)|/\partial x_k \) and substracting it from the \( k \)-th column will result in a matrix of the same rank as \( B \). Since \( r_{x_k}^{k}(y, x_k) \) is also constant across the \( k \)-th column, dividing the resulting column \( k \) by \( r_{x_k}^{k}(y, x_k) \) will not affect the rank as well. Repeating the analogous operations in each of the first \( G \) columns of \( B \) results in the matrix \( A = A \left( \varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)} \right) \). Hence, \( B \) and \( A \) must have the same rank. Since \( A \) is invertible, \( B \) must be invertible. Conversely, starting from the matrix \( B \) and performing the same operations in reverse order, we end up with the matrix \( A \), which will be invertible if \( B \) is invertible.

**Proof of Proposition 3.4:** By Assumption 3.4’, for each \( k = 1, \ldots, G \), the derivative of \( r^k \) with respect \( x_k \) is 1 and the derivative of the Jacobian determinant \( |r_{y}(y, x)| \) with respect \( x_k \) is zero. Then, for each \( k = 1, \ldots, G \) and all \( x \)

\[
g_{x_k}(y, x) = \frac{\partial \log f_{\varepsilon}(r(y, x))}{\partial \varepsilon_k} r_{x_k}^{k}(y, x_k) + \frac{|r_{y}(y, x)|}{\partial \varepsilon_k} \frac{\partial \log f_{\varepsilon}(r(y, x))}{\partial \varepsilon_k}.
\]

Let \( w^{(j)} = (y, x^{(j)}) \) \((j = 1, \ldots, G+1)\) and \( \varepsilon^{(j)} = \begin{pmatrix} r^{1}(y, x_{1}^{(j)}) \ldots, r^{G-1}(y, x_{G-1}^{(j)}), r^{G}(y, x_{G}^{(j)}) \end{pmatrix} \). Since \( g_{x_k}(y, x^{(j)}) = \partial \log f_{\varepsilon}(r(y, x^{(j)})) /\partial \varepsilon_k \), \( B \left( w^{(1)}, \ldots, w^{(G+1)} \right) = A \left( \varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)} \right) \). Hence, \( A \left( \varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)} \right) \) is invertible if and only if \( B \left( w^{(1)}, \ldots, w^{(G+1)} \right) \) is invertible.

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

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

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

\[
\alpha_{y_j}(f + h) - \alpha_{y_j}(f)
\]

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

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

Define \( D\alpha_{y_j}(f;h) = \frac{h_{y_j} \cdot f - f_{y_j} \cdot h}{f^2} \); \( R\alpha_{y_j}(f;h) = -\frac{h \cdot h_{y_j} \cdot f - f_{y_j} \cdot h}{f^2(f + h)} \)

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

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

Then, \( \alpha_{y_j}(f + h) - \alpha_{y_j}(f) = D\alpha_{y_j}(f;h) + R\alpha_{y_j}(f;h) \),

and \( \beta_{x_s}(f + h) - \beta_{x_s}(f) = D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h) \).

Denote \( \mu(x) \ dx \) by \( \mu \), 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,

$$
\Phi_{y_j,x_s}(f+h) - \Phi_{y_j,x_s}(f) \\
= \int Da_{y_j}(f;h) \left( \beta_{x_s}(f) - \int \beta_{x_s}(f) \mu(x) \, dx \right) \mu(x) \, dx \\
+ \int Db_{x_s}(f;h) \left( \alpha_{y_j}(f) - \int \alpha_{y_j}(f) \mu(x) \, dx \right) \mu(x) \, dx \\
+ \int Ra_{y_j}(f;h) \left( \beta_{x_s}(f) - \int \beta_{x_s}(f) \mu(x) \, dx \right) \mu(x) \, dx \\
+ \int Rb_{x_s}(f;h) \left( \alpha_{y_j}(f) - \int \alpha_{y_j}(f) \mu(x) \, dx \right) \mu(x) \, dx \\
+ \int \left( Da_{y_j}(f;h) + Ra_{y_j}(f;h) \right) \left( Db_{x_s}(f;h) + Rb_{x_s}(f;h) \right) \mu(x) \, dx \\
- \left( \int \left( Da_{y_j}(f;h) + Ra_{y_j}(f;h) \right) \mu(x) \, dx \right) \left( \int \left( Db_{x_s}(f;h) + Rb_{x_s}(f;h) \right) \mu(x) \, dx \right).
$$

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 $a < \infty$,

$$
|D\Phi_{y_j,x_s}(f;h)| \leq a \|h\| \quad \text{and} \quad |R\Phi_{y_j,x_s}(f;h)| \leq a \|h\|^2.
$$
Expanding the first term in the sum, we get

$$ (T.1) \quad \int D\alpha_{y_i}(f; h) \left( \beta_{x_i}(f) - \int \beta_{x_i}(f) \mu(x) dx \right) \mu(x) dx $$

$$ = \int h_{y_j} \left[ \mu(x) (\beta_{x_i}(f) - \int \beta_{x_i}(f) \mu(x) dx) \right] dx $$

$$ - \int h \left[ f_{y_j} \mu(x) (\beta_{x_i}(f) - \int \beta_{x_i}(f) \mu(x) dx) \right] dx. $$

Expanding the second term in the sum, we get

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

$$ = \int \frac{[h_{x_i} f - f_{x_i} h]}{f^2} f_{y_j} \mu(x) dx - \int \frac{[\hat{h}_{x_i} \hat{f} - \hat{f}_{x_i} \hat{h}]}{f^2} f_{y_j} \mu(x) dx $$

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

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

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

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

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

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

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

Letting $h = \hat{f} - f$, it follows by our assumptions and standard kernel methods (see, e.g., Newey (1994)) that
Using the notation of Section 3, and applying Newey (1994), this implies that for

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

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

\[ \sqrt{N \sigma_N^{G+2}} \left( \int h \left[ \frac{f_{y,j} \mu(x) (\beta_{x_j}(f) - \int \beta_{x_j}(f) \mu(x) dx)}{\mu(x)} \right] dx \right) \stackrel{p}{\rightarrow} 0 \]

Hence, by the definition of \( \Phi \), it follows that

\[
\sqrt{N \sigma_N^{G+2}} \left( \Phi_{y_j,x_j} \left( \hat{f} \right) - \Phi_{y_j,x_j} \left( f \right) \right) \\
= \sqrt{N \sigma_N^{G+2}} \left( \int D \alpha_{y_j}(f; \hat{f} - f) \left( \beta_{x_j}(f) - \int \beta_{x_j}(f) \mu(x) dx \right) \mu(x) dx \right) + o_p(1) \\
= \sqrt{N \sigma_N^{G+2}} \left( \int \left( \frac{\partial \tilde{f}_{y,j}(y,x)}{\partial y_j} - \frac{\partial f_{y,j}(y,x)}{\partial y_j} \right) \left( \frac{\Delta \partial_{x_j} \log \tilde{f}_{y|x=x}(y)}{\tilde{f}_{y,x}(y,x)} \right) \mu(x) dx \right) + o_p(1)
\]

where \( \Delta \partial_{x_j} \log f_{y|x=x}(y) = \frac{\partial \log \tilde{f}_{y|x=x}(y)}{\partial x_j} - \int \frac{\partial \log \tilde{f}_{y|x=x}(y)}{\partial x_j} \mu(x) dx \)

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

\[ (T.3) \quad \sqrt{N \sigma_N^{G+2}} \left( \hat{T}_{yx}(y) - TT_{yx}(y) \right) \rightarrow^d N(0, V(y)). \]

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,

\[ \Upsilon_{x_j,x_s}(\hat{f}) - \Upsilon_{x_j,x_s}(f) = D \Upsilon_{x_j,x_s} \left( f; \hat{f} - f \right) + R \Upsilon_{x_j,x_s} \left( f; \hat{f} - f \right) \]

where

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

40
and for some $b < \infty$,

$$\left| D_\mathbf{x} f (\beta \hat{f} - f) \right| \leq b \left\| \hat{f} - f \right\| \quad \text{and} \quad \left| R_\mathbf{x} f (\beta \hat{f} - f) \right| \leq b \left\| \hat{f} - f \right\|^2$$

This implies, under our assumptions that $\beta \mathbf{T}_{xx} (y) \to \mathbf{T}_{xx} (y)$. The result of the theorem then follows from this, $(T.3)$, Slutsky’s Theorem, and the definition of $\hat{r}_y (y)$.

**Proof of Theorem 4.1:** We will show that in the model with two equations and one instrument, (4.5) is equivalent to $(L.1)$. As in the proof of Theorem 3.1, we note that since $(r, f_\varepsilon)$ generate $f_{Y|X}$ on $M$, for all $(y, x) \in M$,

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

Specifically for the two equation, one instrument model,

$$(T.4.1) \quad f_{Y_1, Y_2 | X_2 = X_2} (y_1, y_2) = f_{\varepsilon_1, \varepsilon_2} (r^1 (y_1, y_2), r^2 (y_1, y_2, x_2)) \left| r_y (y_1, y_2, x) \right|$$

Taking logs and differentiating both sides of $(T.4.1)$ with respect to $x_2$, we get without writing the arguments of the functions explicitly,

$$(T.4.2) \quad g_{x_2} = r^2_{x_2} q_{x_2} + \gamma_x$$

Taking logs and differentiating both sides of $(T.4.1)$ with respect to $y_1$, we get

$$(T.4.3) \quad g_{y_1} = r^1_{y_1} q_{\varepsilon_1} + r^2_{y_1} q_{\varepsilon_2} + \gamma_{y_1}$$

Since the matrix

$$
\begin{pmatrix}
0 & r^2_{x_2} \\
 r^1_{y_1} & r^2_{y_1}
\end{pmatrix}
$$

is invertible, because by assumption $r^1_{y_1}$ and $r^2_{x_2}$ are different from zero, the unique value of the vector $(q_{\varepsilon_1}, q_{\varepsilon_2})$ that solves $(T.4.2) - (T.4.3)$ is

$$(T.4.4) \quad \begin{pmatrix}
q_{\varepsilon_1} \\
q_{\varepsilon_2}
\end{pmatrix} = \begin{pmatrix}
-r^2_{x_2} r^1_{y_1} & 1 \\
1 & 0
\end{pmatrix} \begin{pmatrix}
g_{x_2} - \gamma_{x_2} \\
g_{y_1} - \gamma_{y_1}
\end{pmatrix}$$

Since $r^*_x$ in $(L.1)$ is, in our model, $(0, r^2_{x_2})$, it follows by $(T.4.2)$ (or, equivalently, by $(T.4.4)$)
that

\[(T.4.5)\quad r'_x q_x = r'_{x_2} q_{x_2} = g_{x_2} - \gamma_{x_2}\]

By the definition of \(r'_y\) and (T.4.4),

\[(T.4.6)\quad r'_y q_x = \begin{pmatrix} r'_{y_1} & r'_{y_2} \\ r'_{y_2} & r'_{y_2} \end{pmatrix} \begin{pmatrix} -\frac{r'_{y_1}}{r_{y_1}^2 r'_{x_2}} & \frac{1}{r'_{y_1}} \\ \frac{1}{r'_{x_2}} & 0 \end{pmatrix} \begin{pmatrix} g_{x_2} - \gamma_{x_2} \\ g_{y_1} - \gamma_{y_1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{r'_{y_1}}{r_{y_1}^2 r'_{x_2}} + \frac{r'_{y_2}}{r_{y_2} r'_{x_2}} & \frac{r'_{y_2}}{r_{y_2}^2} \end{pmatrix} \begin{pmatrix} g_{x_2} - \gamma_{x_2} \\ g_{y_1} - \gamma_{y_1} \end{pmatrix} = \begin{pmatrix} g_{y_1} - \gamma_{y_1} \\ \left(\frac{r'_{y_2}}{r_{y_1}^2}\right) (g_{y_1} - \gamma_{y_1}) + \left(\frac{|r'_{y_2}|}{r_{y_1}^2 r'_{x_2}}\right) (g_{x_2} - \gamma_{x_2}) \end{pmatrix}

Note that

\[(T.4.7)\quad \hat{r'}_x \left(\hat{r'}_y\right)^{-1} = \begin{pmatrix} 0, \frac{r'_{y_1}}{|r'_{y_1}|} \end{pmatrix} \begin{pmatrix} \frac{r'_{y_2}^2}{r_{y_1}^2} & -\frac{r'_{y_1}^2}{r_{y_1}^2} \\ -\frac{r'_{y_1}^2}{r_{y_1}^2} & \frac{r'_{y_2}^2}{r_{y_1}^2} \end{pmatrix} = \begin{pmatrix} -\frac{r'_{y_2}}{r_{y_1}^2} \left(\frac{r'_{y_2}^2}{|r'_{y_1}|}\right), \frac{r'_{y_1}}{r_{y_1}^2} \left(\frac{r'_{y_2}^2}{|r'_{y_1}|}\right) \end{pmatrix}

By (T.4.6) and (T.4.7), it follows that

\[(T.4.8)\quad \left[\hat{r'}_x \left(\hat{r'}_y\right)^{-1}\right] \left[r'_y q_x\right] = \left[\begin{pmatrix} -\frac{r'_{y_2}}{r_{y_1}^2} \left(\frac{r'_{y_2}^2}{|r'_{y_1}|}\right), \frac{r'_{y_1}}{r_{y_1}^2} \left(\frac{r'_{y_2}^2}{|r'_{y_1}|}\right) \end{pmatrix}\right] \begin{pmatrix} g_{y_1} - \gamma_{y_1} \\ \left(\frac{r'_{y_2}}{r_{y_1}^2}\right) (g_{y_1} - \gamma_{y_1}) + \left(\frac{|r'_{y_2}|}{r_{y_1}^2 r'_{x_2}}\right) (g_{x_2} - \gamma_{x_2}) \end{pmatrix}

\= \left[\frac{r'_{y_1}}{r_{y_1}^2} \left(\frac{r'_{y_1}^2}{|r'_{y_1}|}\right) - \frac{r'_{y_2}^2}{r_{y_1}^2} \left(\frac{r'_{y_1}^2}{|r'_{y_1}|}\right) + \frac{r'_{y_2}^2}{r_{y_1}^2} \left(\frac{r'_{y_2}}{r_{y_1}^2}\right) \right] (g_{y_1} - \gamma_{y_1}) + \left[\frac{r'_{y_1}}{r_{y_1}^2} \left(\frac{r'_{y_1}^2}{|r'_{y_1}|}\right) \left(\frac{|r'_{y_2}|}{r_{y_1}^2 r'_{x_2}}\right) \right] (g_{x_2} - \gamma_{x_2})\]
Using (T.4.5) and (T.4.8), we can express the RHS of (L.1) as

\[(T.4.9) \quad \left( r_x' - \frac{\widetilde{r}_x}{r_x} \left( \frac{r_y'}{r_y} \right)^{-1} \frac{r_y'}{r_y} \right) q_x \]

\[= \left( r_x' q_x - \left[ \frac{\widetilde{r}_x}{r_x} \left( \frac{r_y'}{r_y} \right)^{-1} \right] \left[ \frac{r_y'}{r_y} q_x \right] \right) \]

\[= \left( g_{x2} - \gamma_{x2} \right) - \left[ \frac{\widetilde{r}_y}{r_y} \left( \frac{\widetilde{r}_y}{r_y} \right)^{-1} \right] \left( \frac{r_y'}{r_y} \right) \left( g_{y1} - \gamma_{y1} \right) - \left[ \frac{\widetilde{r}_x}{r_x} \left( \frac{\widetilde{r}_x}{r_x} \right)^{-1} \right] \left( \frac{r_y'}{r_y} \right) \left( g_{y1} - \gamma_{y1} \right) \]

By (T.4.7), the LHS of (L.1) can be expressed as

\[(T.4.10) \quad -(\gamma_x - \widetilde{\gamma}_x) + \frac{r_x'}{r_x} \left( \frac{r_y'}{r_y} \right)^{-1} (\gamma_y - \widetilde{\gamma}_y) \]

\[= -(\gamma_x - \widetilde{\gamma}_x) + \left( \frac{\widetilde{r}_y}{r_y} \left( \frac{\widetilde{r}_y}{r_y} \right)^{-1} \right) \left( \frac{r_y'}{r_y} \right) \left( \frac{r_y'}{r_y} \right) \left( \gamma_y - \widetilde{\gamma}_y \right) \]

\[= -(\gamma_x - \widetilde{\gamma}_x) - \frac{\widetilde{r}_y}{r_y} \left( \frac{\widetilde{r}_y}{r_y} \right)^{-1} \left( \gamma_y - \widetilde{\gamma}_y \right) + \frac{\widetilde{r}_y}{r_y} \left( \frac{\widetilde{r}_y}{r_y} \right)^{-1} \left( \gamma_y - \widetilde{\gamma}_y \right) \]

By (T.4.9) and (T.4.10), the expression in (L.1) is then equivalent to

\[(T.4.11) \quad \left[ 1 - \frac{\widetilde{r}_y}{r_y} \left( \frac{\widetilde{r}_y}{r_y} \right)^{-1} \right] \left( g_{x2} - \gamma_{x2} \right) - \left[ \frac{\widetilde{r}_y}{r_y} \left( \frac{\widetilde{r}_y}{r_y} \right)^{-1} \right] \left( g_{y1} - \gamma_{y1} \right) \]

\[= -(\gamma_x - \widetilde{\gamma}_x) - \frac{\widetilde{r}_y}{r_y} \left( \frac{\widetilde{r}_y}{r_y} \right)^{-1} \left( \gamma_y - \widetilde{\gamma}_y \right) + \frac{\widetilde{r}_y}{r_y} \left( \frac{\widetilde{r}_y}{r_y} \right)^{-1} \left( \gamma_y - \widetilde{\gamma}_y \right) \]

Multiplying both sides of (T.4.11) by \(- \frac{\vert r_y \vert}{(\widetilde{r}_x') (\widetilde{r}_y')}\), we get

\[(T.4.12) \quad \left[ \frac{\vert r_y \vert}{(\widetilde{r}_x') (\widetilde{r}_y')} - \frac{\vert r_y \vert}{(\widetilde{r}_x') (\widetilde{r}_y')} \right] \left( g_{x2} - \gamma_{x2} \right) + \left[ \frac{\vert r_y \vert}{(\widetilde{r}_x') (\widetilde{r}_y')} - \frac{\vert r_y \vert}{(\widetilde{r}_x') (\widetilde{r}_y')} \right] \left( g_{y1} - \gamma_{y1} \right) \]

\[= \left[ \frac{\vert r_y \vert}{(\widetilde{r}_x') (\widetilde{r}_y')} \right] \left( \gamma_x - \widetilde{\gamma}_x \right) + \left[ \frac{\vert r_y \vert}{(\widetilde{r}_x') (\widetilde{r}_y')} \right] \left( \gamma_y - \widetilde{\gamma}_y \right) - \left( \gamma_y - \widetilde{\gamma}_y \right) \]

Subtracting \(\frac{\vert r_y \vert}{(\widetilde{r}_x') (\widetilde{r}_y')} \gamma_x + \frac{\vert r_y \vert}{(\widetilde{r}_x') (\widetilde{r}_y')} \gamma_y\) from both sides of (T.4.12) we get
Expansion of the form any $\eta$

**Assumption 4.8.** We apply again Newey (1994). For this, we denote the maximum of the supremum of the values and derivatives up to the fourth order of $\eta$ over a neighborhood of $\gamma$, which is defined implicitly by $\Phi (g, \gamma) = 0$ 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 $F$. For any $h$ such that $\|h\|$ is small enough, any $x$ in a neighborhood of $x^*$, and any small enough...
\[ \delta \text{ such that } |\delta| > 0, \]

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

\[ - \left( \frac{\partial^2 g(x)}{\partial x^2} \left( \frac{\partial g(x)}{\partial x} \right)^2 \left( \frac{\partial^2 g(x)}{\partial x^2} \right)^2 \right) \text{ and} \]

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

Define

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

\[ D_x \Phi (g, x; \delta) = \frac{\partial^3 g(x) |_{x = x(y)} \delta}{\partial x^3} \]

\[ R_f \Phi (g, x; h) = \Phi (g + h, x) - \Phi (g, x) - D_g \Phi (g, x; h), \text{ and} \]

\[ R_x \Phi (g, x; \delta) = \Phi (g, x + \delta) - \Phi (g, x) - D_x \Phi (g, x; \delta). \]

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

\[ \|D_x \Phi (g, x; \delta)\| \leq a |\delta| ; \quad \|R_x \Phi (g, x; \delta)\| \leq a |\delta|^2 ; \]

\[ \|D_g \Phi (g, x; h)\| \leq a \|h\| ; \quad \text{and} \quad \|R_g \Phi (g, x; h)\| \leq a \|h\|^2 \]

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 \(\kappa\) 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 \(\Phi\) is a \(C^2\) map on a neighborhood of \((f, x^*)\) and its second order derivatives are uniformly bounded on such neighborhood, \(\kappa\) 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 \(\kappa\) as an input. Define the functional \(\Psi(g, \kappa(g))\) by

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

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

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

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

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

\[
D_x \Psi (g, x^*; \delta) = \frac{\partial}{\partial x} \left( \frac{- \partial^2 \log g_{y|X=x^*}(y)/\partial y_2 \partial x}{\partial^2 \log g_{y|X=x^*}(y)/\partial y_1 \partial x} \right) \delta
\]
Then,
\[
D\Psi (f, \kappa(f); h) = D_z\Psi (f, x^*; h) + D_x\Psi (f, x^*; D\kappa(f); h) \quad \text{and}
\]
\[
R\Psi (f, \kappa(f); h) = \Psi (f + h, \kappa(f + h)) - \Psi (f, \kappa(f)) - D\Psi (f, \kappa(f); h)
\]

The properties we derived on $D\kappa$ and $R\kappa$ and our assumptions imply that for some $b < \infty$, $|D\Psi (f, \kappa(f); h)| \leq b\|h\|$ and $|R\Psi (f, \kappa(f); h)| \leq b\|h\|^2$. By standard kernel methods and our assumptions it follows that when $h = \hat{f} - f$,
\[
\sqrt{N\sigma_N^7} D\Psi (f, \kappa(f); h)
\]
\[
= \sqrt{N\sigma_N^7} \left[ \frac{\partial f(x^*)}{\partial y_1 \partial x} f(x^*) - \frac{\partial f(x^*)}{\partial y_1} \frac{\partial f(x^*)}{\partial x} \right] \frac{\partial^2 h(x^*)}{\partial y_2 \partial x} + \sqrt{N\sigma_N^7} \left[ \frac{\partial^2 f(x^*)}{\partial y_2 \partial x} f(x^*) - \frac{\partial f(x^*)}{\partial y_2} \frac{\partial f(x^*)}{\partial x} \right] \frac{\partial^2 h(x^*)}{\partial y_1 \partial x} + \sqrt{N\sigma_N^7} \frac{\partial \left( - \frac{\partial^2 \log f_Y|X=x^*}(y)}{\partial y_2 \partial x} \right)}{\partial x} \left( \frac{\partial^3 \log f_Y|X=x*}(y)}{\partial x^3} \right)^{-1} \left[ -1 f(x^*) \right] \frac{\partial^2 h(x^*)}{\partial x^2} + o_p(1)
\]

Hence, when $h = \hat{f} - f$,
\[
\sqrt{N\sigma_N^7} D\Psi (f, \kappa(f); h)
\]
\[
= \sqrt{N\sigma_N^7} \left[ \frac{\partial^2 \log f_Y|X=x^*}(y)}{\partial y_2 \partial x} \right] f_Y|X=x*(y, x^*) \frac{\partial^2 h(x^*)}{\partial y_1 \partial x} + \sqrt{N\sigma_N^7} \left[ \frac{\partial^2 \log f_Y|X=x^*}(y)}{\partial y_1 \partial x} \right] f_Y|X=x*(y, x^*) \frac{\partial^2 h(x^*)}{\partial y_2 \partial x} + \sqrt{N\sigma_N^7} \left[ \frac{\partial^3 \log f_Y|X=x*}(y)}{\partial x^3} \right] f_Y|X=x*(y, x^*) \frac{\partial^2 h(x^*)}{\partial x^2} + o_p(1)
\]

Standard results for kernel estimators imply then that
\[
\sqrt{N\sigma_N^7} D\Psi \left(f, \kappa(f); \hat{f} - f\right) \overset{d}{\to} N(0, \bar{V})
\]
where $\bar{V}$ is as defined prior to the statement of Theorem 4.3. Our assumptions imply that $\sqrt{N\sigma_N^7} R\Psi \left(f, \kappa(f); \hat{f} - f\right) = o_p(1)$. Hence,
\[
\sqrt{N\sigma_N^2} \left[ \partial m^1(y_2, \varepsilon_1) / \partial y_2 - \partial m^1(y_2, \varepsilon_1) / \partial y_2 \right] = \sqrt{N\sigma_N^2} \left[ \Psi \left( \hat{f}, \kappa(f) \right) - \Psi (f, \kappa(f)) \right] = \sqrt{N\sigma_N^2} D\Psi \left( f, \kappa(f); \hat{f} - f \right) + o_p(1) \xrightarrow{d} N \left( 0, \tilde{V} \right)
\]

**Proof of Theorem 4.5:** The proof is similar to the 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 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 \( \bar{g}(x) \), with similar notation for other functions in \( F \). 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 \( \kappa(g) = (\kappa^1(g), \kappa^2(g)) \) 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 \( (\delta_1, \delta_2) \), such that \( |\delta_1| \) and \( |\delta_2| \) are small enough

\[
\Phi \left( g + h, x_1, x_2 \right) - \Phi \left( g, x_1, x_2 \right) = \left( \begin{array}{c} \frac{\partial g(x_1)}{\partial x} + \frac{\partial h(x_1)}{\partial x} - \frac{\partial \bar{g}(x_1)}{\partial x} - \frac{\partial \bar{h}(x_1)}{\partial x} \\frac{\partial g(x_2)}{\partial x} + \frac{\partial h(x_2)}{\partial x} - \frac{\partial \bar{g}(x_2)}{\partial x} - \frac{\partial \bar{h}(x_2)}{\partial x} \end{array} \right)
\]

\[
\Phi \left( g, x_1 + \delta_1, x_2 \right) - \Phi \left( g, x_1, x_2 \right) = \left( \begin{array}{c} \frac{\partial g(x_1 + \delta_1)}{\partial x} - \frac{\partial \bar{g}(x_1 + \delta_1)}{\partial x} \\frac{\partial g(x_1)}{\partial x} + \frac{\partial \bar{g}(x_1)}{\partial x} \end{array} \right)'
\]

and

\[
\Phi \left( g, x_1, x_2 + \delta_2 \right) - \Phi \left( g, x_1, x_2 \right) = \left( \begin{array}{c} \frac{\partial g(x_2 + \delta_2)}{\partial x} - \frac{\partial \bar{g}(x_2 + \delta_2)}{\partial x} \\frac{\partial g(x_2)}{\partial x} + \frac{\partial \bar{g}(x_2)}{\partial x} \end{array} \right)'
\]

48
Define

\[ D_g \Phi(g, x_1, x_2; h) = \begin{pmatrix} \frac{\partial h(x_1)}{g(x_1)} - \frac{\partial g(x_1)}{g(x_1)^2} \frac{h(x_1)}{g(x_1)} + \frac{\partial g(x_1)^2 \gamma_1(x_1)}{2g(x_1)^2} \\ \frac{\partial h(x_2)}{g(x_2)} - \frac{\partial g(x_2)}{g(x_2)^2} \frac{h(x_2)}{g(x_2)} + \frac{\partial g(x_2)^2 \gamma_2(x_2)}{2g(x_2)^2} \end{pmatrix} \]

\[ D_{x_1} \Phi(g, x_1, x_2; \delta_1) = \frac{\partial^2 \log g_{x|x=x_1}(y)}{\partial x^2} \delta_1; \quad D_{x_2} \Phi(g, x_1, x_2; \delta_2) = \frac{\partial^2 \log g_{x|x=x_2}(y)}{\partial x^2} \delta_2 \]

\[ R_f \Phi(g, x_1, x_2) = \Phi(g + h, x_1, x_2) - \Phi(g, x_1, x_2) - D_g \Phi(g, x_1, x_2; h), \]

\[ R_x \Phi(g, x_1, x_2; \delta_1) = \Phi(g, x_1 + \delta_1, x_2) - \Phi(g, x_1, x_2) - D_x \Phi(g, x_1, x_2; \delta_1), \]

and

\[ R_{x_2} \Phi(g, x_1, x_2; \delta_2) = \Phi(g, x_1, x_2 + \delta_2) - \Phi(g, x_1, x_2) - D_{x_2} \Phi(g, x_1, x_2; \delta_2). \]

Our assumptions imply that for some \( a < \infty \),

\[ \|D_{x_j} \Phi(g, x_1, x_2; \delta_j)\| \leq a |\delta_j| \quad \text{and} \quad \|R_{x_j} \Phi(g, x_1, x_2; \delta_j)\| \leq a |\delta_j|^2 \quad \text{for} \ j = 1, 2 \]

\[ \|D_g \Phi(g, x_1, x_2; h)\| \leq a \|h\| ; \quad \text{and} \quad \|R_f \Phi(g, x_1, x_2; h)\| \leq a \|h\|^2 \]

Hence, \( D_x \Phi(g, x_1, x_2; \delta) \) is the Fréchet derivative of \( \Phi \) with respect to \( x \) and \( D_g \Phi(g, x_1, x_2; h) \) is the Fréchet derivative of \( \Phi \) 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 \((f, x_1^*, x_2^*)\). 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 \( \Phi(f, x_1^*, x_2^*) \). It then follows by the Implicit Function Theorem on Banach spaces that there exist unique functionals \( \kappa^1 \) and \( \kappa^2 \) such that \( \kappa^1(f) = x_1^*, \kappa^2(f) = x_2^* \), for all \( g \) in a neighborhood of \( f \)

\[ \Phi(g, \kappa^1(g), \kappa^2(g)) = 0 \]

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

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

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

\[ \Psi(g, x_1, x_2) = \left[ \begin{array}{c} \frac{\partial g_{y_1}(y, x_2)}{\partial y_1} - \frac{\partial g_{y_1}(y, x_1)}{\partial y_1} \\ \frac{\partial g_{y_2}(y, x_2)}{\partial y_2} - \frac{\partial g_{y_2}(y, x_1)}{\partial y_2} \end{array} \right] \left[ \begin{array}{c} \frac{\partial g_{y_2}(y, x_1)}{\partial y_1} - \frac{\partial g_{y_2}(y, x_2)}{\partial y_1} \\ \frac{\partial g_{y_1}(y, x_1)}{\partial y_1} - \frac{\partial g_{y_1}(y, x_2)}{\partial y_1} \end{array} \right]^{-1} \]
Then, \( \Psi \left( \tilde{f}, \kappa^1 \left( \tilde{f} \right), \kappa^2 \left( \tilde{f} \right) \right) = \partial m^1 (y_2, \varepsilon_1) / \partial y_2 \) and \( \Psi \left( f, \kappa^1 \left( f \right), \kappa^2 \left( f \right) \right) = \partial m^1 (y_2, \varepsilon_1) / \partial y_2 \). Denote \( f_{Y,X} (y, x_j) \) by \( f (x_j) \) and \( h_{Y,X} (y, x_j) \) by \( h (x_j) \) (\( j = 1, 2 \)). Then, for \( ||h|| \), \( |\delta_1| \), and \( |\delta_2| \) sufficiently small,

\[
\Psi \left( f + h, x_1^*, x_2^* \right) - \Psi \left( f, x_1^*, x_2^* \right) = \frac{\left[ \frac{\partial f (x_2^*)}{\partial y_2} + \frac{\partial h (x_2^*)}{\partial y_2} \right]}{\left[ \frac{\partial f (x_1^*)}{\partial y_1} + \frac{\partial h (x_1^*)}{\partial y_1} \right]} \left[ f \left( x_1^* \right) + h \left( x_1^* \right) \right] - \frac{\left[ \frac{\partial f (x_1^*)}{\partial y_2} + \frac{\partial h (x_1^*)}{\partial y_2} \right]}{\left[ \frac{\partial f (x_2^*)}{\partial y_1} + \frac{\partial h (x_2^*)}{\partial y_1} \right]} \left[ f \left( x_2^* \right) + h \left( x_2^* \right) \right] - \frac{\frac{\partial f (x_2^*)}{\partial y_2} f \left( x_1^* \right)}{\partial y_1} - \frac{\frac{\partial f (x_1^*)}{\partial y_2} f \left( x_2^* \right)}{\partial y_1}
\]

\[
\Psi \left( f, x_1^* + \delta_1, x_2^* \right) - \Psi \left( f, x_1^*, x_2^* \right) = \frac{\partial f (x_1^* + \delta_1)}{\partial y_1} f (x_2^*) - \frac{\partial f (x_1^*)}{\partial y_1} f (x_2^*) - \frac{\partial f (x_2^*)}{\partial y_2} f (x_1^*) + \frac{\partial f (x_2^*)}{\partial y_2} f (x_2^*) - \frac{\partial f (x_1^*)}{\partial y_1} f (x_1^*)
\]

\[
\Psi \left( f, x_1^*, x_2^* + \delta_2 \right) - \Psi \left( f, x_1^*, x_2^* \right) = \frac{\partial f (x_2^* + \delta_2)}{\partial y_1} f (x_1^*) - \frac{\partial f (x_1^*)}{\partial y_1} f (x_2^*) - \frac{\partial f (x_2^*)}{\partial y_2} f (x_1^*) + \frac{\partial f (x_2^*)}{\partial y_2} f (x_2^*) - \frac{\partial f (x_1^*)}{\partial y_1} f (x_1^*)
\]

Define

\[
D_f \Psi (f, x_1^*, x_2^*; h)
\]

\[
= \left[ \frac{\partial h (x_2^*)}{\partial y_2} f (x_1^*) - \frac{\partial h (x_1^*)}{\partial y_2} f (x_2^*) + \frac{\partial f (x_2^*)}{\partial y_2} h (x_1^*) - \frac{\partial f (x_2^*)}{\partial y_2} h (x_2^*) \right]
\]

\[
= \left[ \frac{\partial f (x_2^*)}{\partial y_1} f (x_1^*) - \frac{\partial f (x_1^*)}{\partial y_1} f (x_2^*) \right]
\]

\[
= \left[ \frac{\partial f (x_1^*)}{\partial y_2} f (x_2^*) - \frac{\partial f (x_2^*)}{\partial y_2} f (x_1^*) \right]
\]

\[
R_f \Psi (f, x_1^*, x_2^*; h) = \Psi \left( f + h, x_1^*, x_2^* \right) - \Psi \left( f, x_1^*, x_2^* \right) - D_f \Psi (f, x_1^*, x_2^*; h)
\]

50
\[
D_{x_1} \Psi (f, x_1^*, x_2^*; \delta_1) = \frac{\partial}{\partial x_1} \left( \frac{\partial \log f_{Y|X=x_1^*}(y)/\partial y_2 - \partial \log f_{Y|X=x_2^*}(y)/\partial y_1}{\partial \log f_{Y|X=x_1^*}(y)/\partial y_1 - \partial \log f_{Y|X=x_2^*}(y)/\partial y_1} \right) \bigg|_{x_1 = x_1^*; \delta_1}
\]
\[
D_{x_2} \Psi (f, x_1^*, x_2^*; \delta_2) = \frac{\partial}{\partial x_2} \left( \frac{\partial \log f_{Y|X=x_2^*}(y)/\partial y_2 - \partial \log f_{Y|X=x_1^*}(y)/\partial y_1}{\partial \log f_{Y|X=x_1^*}(y)/\partial y_1 - \partial \log f_{Y|X=x_2^*}(y)/\partial y_1} \right) \bigg|_{x_2 = x_2^*; \delta_2}
\]
\[
D \Psi (f, x_1^*, x_2^*; h) = D_f \Psi (f, x_1^*, x_2^*; h) + D_{x_1} \Psi (f, x_1^*, x_2^*; D_1^0(f; h)) + D_{x_2} \Psi (f, x_1^*, x_2^*; D_2^0(f; h))
\]
and \[
R \Psi (f, x_1^*, x_2^*; h) = \Psi (f + h, \kappa_1^0(f + h), \kappa_2^0(f + h)) - \Psi (f, x_1^*, x_2^*) - D \Psi (f, x_1^*, x_2^*; h).
\]

Our assumptions imply that
\[
|D \Psi (f, x_1^*, x_2^*; h)| \leq a \|h\| \quad \text{and} \quad |R \Psi (f, x_1^*, x_2^*; h)| \leq a \|h\|^2
\]

By standard properties of kernel estimators, it follows that when \( h = \hat{f} - f \),
\[
\sqrt{N \sigma_N^5} D \Psi (f, x_1^*, x_2^*; h)
\]
\[
= \sqrt{N \sigma_N^5} \left[ \frac{\partial h(x_1^*)}{\partial y_2} f(x_1^*) - \frac{\partial h(x_2^*)}{\partial y_2} f(x_2^*) \right] + \sqrt{N \sigma_N^5} \left[ \frac{\partial f(x_1^*)}{\partial y_1} f(x_1^*) - \frac{\partial f(x_2^*)}{\partial y_1} f(x_2^*) \right]
\]
\[
- \sqrt{N \sigma_N^5} \left[ \frac{\partial f(x_1^*)}{\partial y_2} f(x_1^*) - \frac{\partial f(x_2^*)}{\partial y_2} f(x_2^*) \right]^2
\]
\[
+ \sqrt{N \sigma_N^5} \left( \frac{\partial \log f_{Y|X=x_1^*}(y)/\partial y_2 - \partial \log f_{Y|X=x_1^*}(y)/\partial y_1}{\partial \log f_{Y|X=x_1^*}(y)/\partial y_1 - \partial \log f_{Y|X=x_1^*}(y)/\partial y_1} \right) \bigg|_{x_1 = x_1^*} \left( - \frac{\partial h(x_1^*)}{\partial x} \right)
\]
\[
+ \sqrt{N \sigma_N^5} \left( \frac{\partial \log f_{Y|X=x_2^*}(y)/\partial y_2 - \partial \log f_{Y|X=x_2^*}(y)/\partial y_1}{\partial \log f_{Y|X=x_2^*}(y)/\partial y_1 - \partial \log f_{Y|X=x_2^*}(y)/\partial y_1} \right) \bigg|_{x_2 = x_2^*} \left( - \frac{\partial h(x_2^*)}{\partial x} \right) + o_p(1)
\]

Hence, by standard results for kernel estimators, \( \sqrt{N \sigma_N^5} D \Psi (f, x_1^*, x_2^*; h) \to N(0, \nabla) \) where \( \nabla \) is as defined prior to the statement of Theorem 4.4. Since our assumptions guarantee that \( \sqrt{N \sigma_N^5} R \Psi (f, x_1^*, x_2^*; h) = o_p(1) \), we can conclude that
\[
\sqrt{N \sigma_N^5} \left( \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} - \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} \right)
\]
\[
= \sqrt{N \sigma_N^5} \left( \Psi \left( \hat{f}, \kappa^1 \left( \hat{f} \right), \kappa^2 \left( \hat{f} \right) \right) - \Psi (f, x_1^*, x_2^*) \right) d \sim N(0, \nabla)
\]
This concludes the proof.

8. References


CHESHER, A. (2005): “Identification of Non-additive Structural Functions," mimeo, presented at the Symposium on Nonparametric Structural Models, World Congress of the Econo-
metric Society, London.


