Econometrica, Vol. 83, No. 1 (January, 2015), 1–66

ESTIMATION OF NONPARAMETRIC MODELS WITH SIMULTANEITY

BY ROSA L. MATZKIN[1]

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 ratios of derivatives or averages of products of derivatives of this density. The estimators are therefore easily computed functionals of a nonparametric estimator of the density of the observable variables. We consider in detail a model where to each structural equation there corresponds an exclusive regressor and a model with one equation of interest and one instrument that is included in a second equation. For both models, we provide new characterizations of observational equivalence on a set, in terms of the density of the observable variables and derivatives of the structural functions. Based on those characterizations, we develop two estimation methods. In the first method, the estimators of the structural derivatives are calculated by a simple matrix inversion and matrix multiplication, analogous to a standard least squares estimator, but with the elements of the matrices being averages of products of derivatives of nonparametric density estimators. In the second method, the estimators of the structural derivatives are calculated in two steps. In a first step, values of the instrument are found at which the density of the observable variables satisfies some properties. In the second step, the estimators are calculated directly from the values of derivatives of the density of the observable variables evaluated at the found values of the instrument. We show that both pointwise estimators are consistent and asymptotically normal.

KEYWORDS: Simultaneous equations, instrumental variables, constructive identification, nonseparable models, kernel estimators, endogeneity, structural models.

1. INTRODUCTION

THIS PAPER PRESENTS ESTIMATORS for two nonparametric models with simultaneity. The estimators are shown to be consistent and asymptotically normally distributed. They are derived from new constructive identification results that are also presented in the paper. The nonparametric models possess nonadditive unobservable random terms. We consider a model where to each equa-

[1]The support of NSF through Grants SES-0833058 and SES-1062090 is gratefully acknowledged. I have greatly benefited from the insightful comments of a co-editor, a guest co-editor, three referees, discussions with Richard Blundell, Jim Heckman, and Whitney Newey, comments by participants at the 2008 All UC Econometrics Conference at Berkeley, 2009 Winter Meeting of the Econometric Society, 2009 BIRS Conference on Semiparametric and Nonparametric Methods in Econometrics, 2009 Conference on Consumer Demand at Brown University, 2009 SETA Conference at Kyoto University, Invited Session on Advances in Econometrics at the 2009 Latin American Meeting of the Econometric Society, 2009 Annual Minnesota Lecture, 2010 World Congress of the Econometric Society, 2010 Cowles Summer Conference, many econometric seminars including Collegio Carlo Alberto, EIFEF, Harvard/MIT, Maryland, Montreal Econometrics, Princeton, UCL, UCLA, UCSD, USC, Stanford, and the excellent research assistance of Matthew Baird, Anastasia Burkovskaya, David Kang, Ruoyao Shi, Yulong Wang, Kyle Woodward, Yong Hyeon Yang, and Federico Zincenko.

 DOI: 10.3982/ECTA9348

tion there corresponds an exclusive regressor and a model with two equations and one instrument. For both models, we develop closed form estimators of structural derivatives, by averaging over excluded instruments. For the second model, we develop also two-step indirect estimators.

Estimation of structural models has been one of the main objectives of econometrics since its early times. The analyses of counterfactuals, the evaluation of welfare, and the prediction of the evolution of markets, among others, require knowledge of primitive functions and distributions in the economy, such as technologies and distributions of preferences, which often can only be estimated using structural models. Estimation of parametric structural models dates back to the early works of Haavelmo (1943, 1944), Hurwicz (1950), Koopmans (1949), Koopmans and Reiersol (1950), Koopmans, Rubin, and Leipnik (1950), Wald (1950), Fisher (1959, 1961, 1965, 1966), Wegge (1965), Rothenberg (1971), and Bowden (1973). (See Hausman (1983) and Hsiao (1983) for early review articles.) Hurwicz (1950) considered nonseparable econometric models, where the random terms are nonadditive.

Nonparametric structural models avoid specifying the functions and distributions as known up to a finite dimensional parameter vector. Several nonparametric estimators have been developed for models with simultaneity, based on conditional moment restrictions. These include Newey and Powell (1989, 2003), Darolles, Florens, and Renault (2002), Ai and Chen (2003), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), Darolles, Fan, Florens, and Renault (2011), and for models with nonadditive random terms, Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), Horowitz and Lee (2007), Chen and Pouzo (2012), and Chen, Chernozhukov, Lee, and Newey (2014). 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).[2] For structural functions with

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

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 required to employ these methods is very restrictive. (See Blundell and Matzkin (2014) for a characterization of simultaneous equations 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 nonseparable, 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 present two estimation approaches and focus on two models. Both approaches are developed from two new characterizations of observational equivalence for simultaneous equations, which we introduce in this paper. The new characterizations, expressed in terms of the density of the observable variables and ratios of derivatives of the structural functions, immediately provide constructive methods for identifying one or more structural elements.

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, at each value of the unobservable costs, from the distribution of observable equilibrium actions and players' costs. The estimator that we present for this model is an average derivative type of estimator. The calculation of the estimator for the derivatives of the reaction functions of the players, at any given value of the observable and unobservable arguments, requires only a simple matrix inversion and a matrix multiplication, analogous to the solution of linear least squares estimators. The difference is that the elements in our matrices are calculated using nonparametric averages of products of derivatives. 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 weighted averages of functions of nonparametrically estimated derivatives of the densities of the observable variables.[3]

[3]Existent extensions of the average derivative methods of Stoker (1986) and Powell, Stock, and Stoker (1989) for models with endogeneity, such as Altonji and Matzkin (2001, 2005), Blundell and Powell (2003), Imbens and Newey (2003, 2009), and Altonji, Ichimura, and Otsu (2012), 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 function 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 for this model estimators based on our two approaches, the average over instruments derivative estimator, similar to the estimator developed for our first model, and indirect estimators. Our indirect estimators are based on a two-step procedure. In the first step, either one or two values of the observable cost are found where the density of the observable variables satisfies some conditions. In the second step, the derivative of the demand with respect to price is read off the joint density of the price, quantity, and cost evaluated at the found values of cost. The estimators are developed by substituting the joint density of price, quantity, and cost, by a nonparametric estimator for it. The estimators for the derivative of the demand, at a given price and quantity, that we develop are consistent and asymptotically normal.

The new observational equivalence and constructive identification results that we present do not require large support conditions on the observable exogenous regressors. They are based on the identification results in Matzkin (2008), which start out from the transformation of variables equation for densities. Employing this equation, Matzkin (2007b, Section 2.1.4) presented a two-step constructive identification result for an exclusive regressors model, under the assumption that the density of the unobservable variables has a unique known mode. (See Matzkin (2013) for a detailed development of that result.) Other identification results that were developed using similar equations are Berry and Haile (2009, 2011, 2014) and Chiappori and Komunjer (2009). Berry and Haile (2009, 2011, 2014) developed alternative constructive identification results starting out also from the transformation of variables equation for densities. Their results apply to the class of exclusive regressors models where each unobservable variable enters in the structural functions through an index. Chiappori and Komunjer (2009) derived generic, nonconstructive 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.

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 large dimensional unobserved heterogeneity, and models where the unobservable variables are only conditionally independent of the explanatory variables. (See Matzkin (2012) for identification results based on Matzkin (2008) in such extended models.)

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.

The structure of the paper is as follows. In the next section, we present the exclusive regressors model. We develop new observational equivalence results for such model, and use those results to develop a closed form estimator for either the derivatives or the ratios of derivatives of the structural functions in this model. We show that the estimator is consistent and asymptotically normal. In Section 3, we consider a model with two equations and one instrument. We develop new observational equivalence results for such model, and use those results to develop an estimator for the derivative of the structural function that excludes the instrument, by averaging over the instrument. We show that the estimator, which is given in closed form, is also consistent and asymptotically normal. In Section 4, we present indirect, two-step estimators for the two equation, one instrument model, and show they are also consistent and asymptotically normal. Section 5 presents results of simulations performed using some of the estimators. Section 6 concludes.

2. THE MODEL WITH EXCLUSIVE REGRESSORS

2.1. *The Model*

We consider in this section the model

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

where $(Y_1, \dots, Y_G)$ is a vector of observable endogenous variables, $(Z, X_1, \dots, X_G)$ is a vector of observable exogenous variables, and $(\varepsilon_1, \dots, \varepsilon_G)$ is a vector of unobservable variables. The observable vector Z has the effect of decreasing the rates of convergence of nonparametric estimators of model (2.1) 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 ε_g is unobservable, we will at most be able to identify the values of ε_g up to an invertible transformation.[4] Hence, for each g, either we may normalize m^g to be strictly increasing in ε^g, as it is assumed in models additive in ε^g, or we may normalize

[4]See Matzkin (1999, 2003, 2007a, 2007b) for discussion of this nonidentification result in the one equation model, $Y = m(X, \varepsilon)$, with x and ε independently distributed.

m^g to be strictly decreasing in ε_g. The invertibility of m^g in ε_g implies that, for any fixed values, y_{-g} and x_{-g}, of the other arguments of the function m^g, there is a unique value of ε_g for each value of y_g. We will denote the function that assigns such value of ε_g by $r^g(y_1, \dots, y_G, x_g)$. Our system of inverse structural equations, denoting the mapping from the vectors of observable variables to the vector of unobservable variables, is expressed as

$$\begin{aligned} \varepsilon_1 &= r^1(Y_1, \dots, Y_G, X_1), \\ \varepsilon_2 &= r^2(Y_1, \dots, Y_G, X_2), \\ &\dots \\ \varepsilon_G &= r^G(Y_1, \dots, Y_G, X_G). \end{aligned} \tag{2.2}$$

The derivatives of the function m^g can be calculated by substituting (2.2) into (2.1) 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

$$\begin{aligned} &\left.\frac{\partial m^g(y_{-g}, x_g, \varepsilon_g)}{\partial y_j}\right|_{\varepsilon_g = r^g(y_1, \dots, y_G, x_g)} \\ &\quad = -\left[\frac{\partial r^g(y_g, y_{-g}, x_g)}{\partial y_g}\right]^{-1}\left[\frac{\partial r^g(y_g, y_{-g}, x_g)}{\partial y_j}\right]. \end{aligned}$$

The derivative 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 ε^g when y_g is a specified value is

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

The estimation methods we introduce are based on the assumptions that (i) ε and X have, respectively, differentiable densities, f_ε and f_X, (ii) the functions r^g $(g = 1, \dots, G)$ are twice continuously differentiable, and (iii) for all (y, x) in a set M in the support of $(Y_1, \dots, Y_G, X_1, \dots, X_G)$, the conditional density of Y given $X = x$, evaluated at $Y = y$, is given by the transformation of variables equation

$$f_{Y|X=x}(y) = f_\varepsilon\big(r(y, x)\big)\left|\frac{\partial r(y, x)}{\partial y}\right|, \tag{2.3}$$

where $|\partial r(y, x)/\partial y|$ denotes the absolute value of the Jacobian determinant of $r(y, x)$ with respect to y. In addition, we assume that (iv) for each g, the function r^g has a nonvanishing derivative with respect to its exclusive regressor, x_g. A set of sufficient conditions for (i)–(iv) is given by Assumptions 2.1–2.3 below.

ASSUMPTION 2.1: *The function* $r = (r^1, \dots, r^G)$ *is twice continuously differentiable. For each* g, r^g *is invertible in* y_g *and the derivative of* r^g *with respect to* x_g is bounded away from zero. *Conditional on* $(X_1, \dots, X_G)$, *the function* r *is* 1–1, *onto* R^G, *and as a function of* x, *its Jacobian determinant is positive and bounded away from zero.*

ASSUMPTION 2.2: $(\varepsilon_1, \dots, \varepsilon_G)$ *is distributed independently of* $(X_1, \dots, X_G)$ *with an everywhere positive and twice continuously differentiable density*, f_ε.

ASSUMPTION 2.3: $(X_1, \dots, X_G)$ *possesses a differentiable density.*

For the analysis of identification, the left-hand side of (2.3) can be assumed known. In practice, it can be estimated nonparametrically. The right-hand side involves the structural functions, f_ε and r, whose features are the objects of interest. The differentiability assumptions on f_ε, f_X, and r imply that both sides of (2.3) can be differentiated with respect to y and x. This allows us to transform (2.3) into a system of linear equations with derivatives of known functions on one side and derivatives of unknown functions on the other side. We show in the next subsection how the derivatives of the known function can be used to identify ratios of derivatives of the unknown functions r^g.

We will develop estimators for the identified features of r. Theorem 2.1 below characterizes the features of r that can be identified. Roughly, the theorem states that, under appropriate conditions on the density, f_ε, of ε, and on a vector of composite derivatives of $\log|\partial r(y, x)/\partial y|$, the ratios of derivatives, $r^g_{y_j}(y, x_g)/r^g_{x_g}(y, x_g)$, of each of the functions, r^g, with respect to its coordinates, are identified. The statement that the ratios of derivatives, $r^g_{y_j}(y, x_g)/r^g_{x_g}(y, x_g)$, are 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.[5] One such class can be defined by requiring that for each function in the class there exists a function $s^g : R^G \to R$ such that, for all x_g,

$$r^g(y, x_g) = s^g(y) + x_g, \tag{2.4}$$

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

[5]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, 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.

2.2. *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 equivalence 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_ε, of the vector of unobservable variables, ε. We also modify Matzkin's (2008) 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 $\mathbf{g}_y(y,x) = (\mathbf{g}_{y_1}(y,x),\dots,\mathbf{g}_{y_G}(y,x))'$. 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 $\mathbf{g}_x(y,x) = (\mathbf{g}_{x_1}(y,x),\dots,\mathbf{g}_{x_G}(y,x))'$. The derivative of the log of the density, f_ε, of ε, with respect to ε, $\partial \log f_\varepsilon(\varepsilon)/\partial\varepsilon$, will be denoted by $q_\varepsilon(\varepsilon) = (q_{\varepsilon_1}(\varepsilon),\dots,q_{\varepsilon_G}(\varepsilon))'$. 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, $r^g_{y_j}(y,x_g)/r^g_{x_g}(y,x_g)$, will be denoted by $\overline{r}^g_{y_j}(y,x_g)$. For an alternative function $\widetilde{r}$, these ratios of derivatives will be denoted by $\overline{\widetilde{r}}^g_{y_j}(y,x_g)$. The Jacobian determinants, $|\partial r(y,x)/\partial y|$ and $|\partial\widetilde{r}(y,x)/\partial y|$, will be denoted respectively by $|r_y(y,x)|$ and $|\widetilde{r}_y(y,x)|$. The derivatives of $|r_y(y,x)|$ and $|\widetilde{r}_y(y,x)|$ with respect to any of their arguments, $w \in \{y_1,\dots,y_G,x_1,\dots,x_G\}$, will be denoted by $|r_y(y,x)|_w$ and $|\widetilde{r}_y(y,x)|_w$. The statement of our main observational equivalence result in this section involves functions, d_{y_g}, defined for each g by

$$d_{y_g}(y,x) = \frac{|r_y(y,x)|_{y_g}}{|r_y(y,x)|} - \sum_{k=1}^{G}\left[\frac{|r_y(y,x)|_{x_k}}{|r_y(y,x)|}\,\frac{r^k_{y_g}(y,x_k)}{r^k_{x_k}(y,x_k)}\right]. \tag{2.5}$$

The term $d_{y_g}(y,x)$ can be interpreted as the effect on $\log|\partial r(y,x)/\partial y|$ of a simultaneous change in y_g and in $(x_1,\dots,x_G)$.

Let Γ denote the set of functions r that satisfy Assumption 2.1 and let Φ denote the set of densities that satisfy Assumption 2.2. We define observational equivalence within Γ over a subset, M, in the interior of the support of the vector of observable variables.

DEFINITION 2.1: Let M denote a subset of the support of (Y,X), such that for all $(y,x)\in M$, $f_{Y,X}(y,x) > \delta$, where δ is any positive constant. A function $\widetilde{r}\in\Gamma$ is observationally equivalent to $r\in\Gamma$ on M if there exist densities f_ε and $f_{\widetilde{\varepsilon}}$ satisfying Assumption 2.2 and such that, for all $(y,x)\in M$,

$$f_\varepsilon\big(r(y,x)\big)\left|\frac{\partial r(y,x)}{\partial y}\right| = f_{\widetilde{\varepsilon}}\big(\widetilde{r}(y,x)\big)\left|\frac{\partial\widetilde{r}(y,x)}{\partial y}\right|. \tag{2.6}$$

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

THEOREM 2.1: *Suppose that (r, f_ε) generates $f_{Y|X}$ on M and that Assumptions 2.1–2.3 are satisfied. A function $\widetilde{r} \in \Gamma$ is observationally equivalent to $r \in \Gamma$ on M if and only if, for all $(y, x) \in M$,*

$$(2.7)\qquad \begin{aligned} 0 &= \big(\overline{r}^1_{y_1} - \widetilde{\overline{r}}^1_{y_1}\big)\mathbf{g}_{x_1} + \big(\overline{r}^2_{y_1} - \widetilde{\overline{r}}^2_{y_1}\big)\mathbf{g}_{x_2} + \cdots \\ &\quad + \big(\overline{r}^G_{y_1} - \widetilde{\overline{r}}^G_{y_1}\big)\mathbf{g}_{x_G} + (d_{y_1} - \widetilde{d}_{y_1}), \\ &\cdots \\ 0 &= \big(\overline{r}^1_{y_2} - \widetilde{\overline{r}}^1_{y_2}\big)\mathbf{g}_{x_1} + \big(\overline{r}^2_{y_2} - \widetilde{\overline{r}}^2_{y_2}\big)\mathbf{g}_{x_2} + \cdots \\ &\quad + \big(\overline{r}^G_{y_2} - \widetilde{\overline{r}}^G_{y_2}\big)\mathbf{g}_{x_G} + (d_{y_2} - \widetilde{d}_{y_2}), \\ &\cdots \\ 0 &= \big(\overline{r}^1_{y_G} - \widetilde{\overline{r}}^1_{y_G}\big)\mathbf{g}_{x_1} + \big(\overline{r}^2_{y_G} - \widetilde{\overline{r}}^2_{y_G}\big)\mathbf{g}_{x_2} + \cdots \\ &\quad + \big(\overline{r}^G_{y_G} - \widetilde{\overline{r}}^G_{y_G}\big)\mathbf{g}_{x_G} + (d_{y_G} - \widetilde{d}_{y_G}), \end{aligned}$$

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

The proof, presented in the Appendix, uses (2.3) to obtain an expression for the unobservable $\partial \log f_\varepsilon(r(y, x))/\partial\varepsilon$ in terms of the observable $\mathbf{g}_x(y, x) = \partial \log f_{Y|X=x}(y)/\partial x$. The expression in terms of $\mathbf{g}_x(y, x)$ 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 (2.7) is obtained after manipulating the equations resulting from such substitution.

Theorem 2.1 can be used with (2.3) to develop constructive identification results for features of r, and estimators for such features. Taking logs and differentiating both sides of (2.3) with respect to y_j gives

$$(2.8)\qquad \frac{\partial \log f_{Y|X=x}(y)}{\partial y_j} = \sum_{g=1}^{G} \frac{\partial \log f_\varepsilon(r(y, x))}{\partial \varepsilon_g} r^g_{y_j}(y, x_1) + \frac{|r_y(y, x)|_{y_j}}{|r_y(y, x)|},$$

and taking logs and differentiating both sides of (2.3) with respect to x_g gives

$$(2.9)\qquad \frac{\partial \log f_{Y|X=x}(y)}{\partial x_g} = \frac{\partial \log f_\varepsilon(r(y, x))}{\partial \varepsilon_g} r^g_{x_g}(y, x_1) + \frac{|r_y(y, x)|_{x_g}}{|r_y(y, x)|}.$$

Solving for $\partial \log f_\varepsilon(r)/\partial\varepsilon_g$ in (2.9) and substituting the result into each of the $\partial \log f_\varepsilon(r)/\partial\varepsilon_g$ terms in (2.8), we obtain

$$\frac{\partial \log f_{Y|X=x}(y)}{\partial y_j} = \sum_{g=1}^{G} \frac{\partial \log f_{Y|X=x}(y)}{\partial x_g} \frac{r^g_{y_j}(y, x_g)}{r^g_{x_g}(y, x_g)} + d_{y_j}(y, x), \tag{2.10}$$

where $d_{y_j}(y, x)$ is as in (2.5). Equation (2.10) implies that $(\bar{r}, d)$ satisfies the following system of equations:

$$\begin{aligned}
\mathbf{g}_{y_1} &= \bar{r}^1_{y_1}\mathbf{g}_{x_1} + \bar{r}^2_{y_1}\mathbf{g}_{x_2} + \cdots + \bar{r}^G_{y_1}\mathbf{g}_{x_G} + d_{y_1}, \\
&\cdots \\
\mathbf{g}_{y_2} &= \bar{r}^1_{y_2}\mathbf{g}_{x_1} + \bar{r}^2_{y_2}\mathbf{g}_{x_2} + \cdots + \bar{r}^G_{y_2}\mathbf{g}_{x_G} + d_{y_2}, \\
&\cdots \\
\mathbf{g}_{y_G} &= \bar{r}^1_{y_G}\mathbf{g}_{x_1} + \bar{r}^2_{y_G}\mathbf{g}_{x_2} + \cdots + \bar{r}^G_{y_G}\mathbf{g}_{x_G} + d_{y_G}.
\end{aligned} \tag{2.11}$$

Several features of this system of equations deserve mentioning. First, note that this is a system of G equations where only the ratios of derivatives $\bar{r}^g_{y_j} = r^g_{y_j}/r^g_{x_g}$ $(g, j = 1, \ldots, G)$ and the terms $d_{y_1}, \ldots, d_{y_G}$ are unknown. Second, note that the unknown elements in this system are elements of only the inverse function $(r^1, \ldots, r^G)$. They do not depend on the unknown density of $(\varepsilon_1, \ldots, \varepsilon_G)$. The density f_ε enters the system only through the known terms, $\mathbf{g}_{y_1}, \ldots, \mathbf{g}_{y_G}$ and $\mathbf{g}_{x_1}, \ldots, \mathbf{g}_{x_G}$. Moreover, the values of f_ε depend on the values of $(r^1, \ldots, r^G)$ rather than on the ratios of derivatives of $(r^1, \ldots, r^G)$. Hence, the density f_ε has the potential to generate variation on the values of $\mathbf{g}_{y_1}, \ldots, \mathbf{g}_{y_G}$ and $\mathbf{g}_{x_1}, \ldots, \mathbf{g}_{x_G}$, independently of the unknown ratios of derivatives $\bar{r}^g_{y_j} = r^g_{y_j}/r^g_{x_g}$ $(g, j = 1, \ldots, G)$ and of $d_{y_1}, \ldots, d_{y_G}$. Third, each of the ratios of derivatives depend on only one x_g, while $\mathbf{g}_{y_1}, \ldots, \mathbf{g}_{y_G}$ and $\mathbf{g}_{x_1}, \ldots, \mathbf{g}_{x_G}$ depend on all the vector $(x_1, \ldots, x_G)$. 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 $r^g_{y_j}/r^g_{x_g}$ stay fixed. This leads to the analysis of conditions on (r, f_ε) guaranteeing that values of $(\mathbf{g}_{x_1}, \ldots, \mathbf{g}_{x_G})$, which are observable, can be found so that (2.11) can be solved for either the whole vector $(\bar{r}, d) = (\bar{r}^1_{y_1}, \ldots, \bar{r}^1_{y_G}; \ldots; \bar{r}^G_{y_1}, \ldots, \bar{r}^G_{y_G}; d_{y_1}, \ldots, d_{y_G})$ or for some elements of it.

2.3. *Average Derivatives Estimators for the Model With Exclusive Regressors*

We next develop an estimator for the ratios of derivatives, $\bar{r}$, based on a characterization of $\bar{r}$ of the least squares form $(\widetilde{X}'\widetilde{X})^{-1}(\widetilde{X}'\widetilde{Y})$. Such characterization of $\bar{r}$ employs the fact that (2.11) holds for all values of $(\mathbf{g}_{x_1}, \ldots, \mathbf{g}_{x_G})$ over any subset $\overline{M}$ of M where $(\bar{r}, d)$ is constant. The elements of the matrices are obtained by averages of multiplications of derivatives of $\log f_{Y|X=x}(y)$ over a set $\overline{M}$ where the values of $(\bar{r}, d)$ are constant. Estimation of $\bar{r}$ follows by substituting, in the $(\widetilde{X}'\widetilde{X})$ and $(\widetilde{X}'\widetilde{Y})$ matrices, $f_{Y|X=x}$ by a nonparametric estimator for $f_{Y|X=x}$.

We derive our expression for $\bar{r}$ by characterizing $(\bar{r}, d)$ as the unique solution to the minimization of an integrated square distance between the left-hand side and the right-hand side of (2.11). The integration set must be over a subset of the support of (Y, X) where $(\bar{r}, d)$ 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 different integration sets.

Another restriction on the integration set is that it must contain in its interior $G+1$ values of the observable variables such that when $\mathbf{g}_x$ is evaluated at those values, the only solution to (2.7) is the vector of 0's. This identification condition guarantees that $(\bar{r}, d)$ is the unique minimizer of the distance function. For each of the two sets of restrictions on r, we will provide conditions on f_ε guaranteeing that such $G+1$ values exist. The two sets of restrictions on r that we will consider are stated in Assumptions 2.4 and 2.4′.

ASSUMPTION 2.4: *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 2.4′: *For each $g = 1, \dots, G$, 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 2.4 is equivalent to requiring that the units of measurement of ε_G are tied to those of x_G by[6]

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

The sets $\overline{M}$ on which $(\bar{r}, d)$ is constant when Assumptions 2.4 and 2.4′ are satisfied are stated in the following propositions.

PROPOSITION 2.1: *Let $(y, x_{-G}) = (y, x_1, \dots, x_{G-1})$ be fixed and given. When Assumptions 2.1 and 2.4 are satisfied, $(\bar{r}, d)$ is constant over any set $\overline{M} \subset \{(y, x_{-G}, t_G) | t_G \in R\}$.*

PROPOSITION 2.2: *Let y be fixed and given. When Assumptions 2.1 and 2.4′ are satisfied, $(\bar{r}, d)$ is constant over any set $\overline{M} \subset \{(y, t_1, \dots, t_G) | (t_1, \dots, t_G) \in R^G\}$.*

[6] I thank a referee for showing that Assumption 2.4 is equivalent to this restriction.

To state our assumption on the density f_ε, we will denote by $A(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)})$ the matrix of derivatives of $\log(f_\varepsilon)$ at $G+1$ values, $\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)}$, of ε,

$$A\big(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)}\big) = \begin{pmatrix} \dfrac{\partial \log f_\varepsilon(\varepsilon^{(1)})}{\partial \varepsilon_1} & \dfrac{\partial \log f_\varepsilon(\varepsilon^{(1)})}{\partial \varepsilon_2} & \cdots & \dfrac{\partial \log f_\varepsilon(\varepsilon^{(1)})}{\partial \varepsilon_G} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial \log f_\varepsilon(\varepsilon^{(G)})}{\partial \varepsilon_1} & \dfrac{\partial \log f_\varepsilon(\varepsilon^{(G)})}{\partial \varepsilon_2} & \cdots & \dfrac{\partial \log f_\varepsilon(\varepsilon^{(G)})}{\partial \varepsilon_G} & 1 \\ \dfrac{\partial \log f_\varepsilon(\varepsilon^{(G+1)})}{\partial \varepsilon_1} & \dfrac{\partial \log f_\varepsilon(\varepsilon^{(G+1)})}{\partial \varepsilon_2} & \cdots & \dfrac{\partial \log f_\varepsilon(\varepsilon^{(G+1)})}{\partial \varepsilon_G} & 1 \end{pmatrix}.$$

ASSUMPTION 2.5: *There exist $G+1$, not necessarily known, values $\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)}$ of ε, such that $A(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)})$ is invertible.*[7]

ASSUMPTION 2.6: *There exist $G+1$, not necessarily known, values $w^{(1)}, \dots, w^{(G+1)}$ in the set $\overline{M}$, where $(\bar{r}, d)$ is constant, such that for each $k = 1, \dots, G+1$, $\varepsilon^{(k)} = r(w^{(k)})$, where $\varepsilon^{(k)}$ is as in Assumption 2.5.*

Assumptions 2.5 and 2.6 require that there exist $G+1$ values of ε, each corresponding to the value of r at one point in the set $\overline{M}$, such that the matrix $A(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)})$ is invertible. When the set $\overline{M}$ is as in Proposition 2.1, 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 ε that can be used to satisfy the invertibility of $A(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)})$ must possess the same values of $\varepsilon_1, \dots, \varepsilon_{G-1}$. Hence, only changes in the value of ε_G must generate the $G+1$ linearly independent rows in $A(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)})$. When the set $\overline{M}$ is as in Proposition 2.2, the points in the set $\overline{M}$ may differ in their values of $(x_1, \dots, x_G)$. In this case, the $G+1$ values of ε that must satisfy the invertibility of $A(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)})$ may differ in their values of any coordinate, not just in the value of their last coordinate. Normal distributions can satisfy Assumption 2.5 in the latter case, where the values of $\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)}$ are allowed to differ in all coordinates, but not in the former case, where only the last coordinates of $\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)}$ are different.

The following propositions relate Assumptions 2.5 and 2.6 to a testable condition. Let $w^{(1)}, \dots, w^{(G+1)}$ denote $G+1$ points in $\overline{M}$. For each j and k, we will denote the values of $\mathbf{g}_{x_j}$ at $w^{(k)}$ by $\mathbf{g}_{x_j}^{(k)}$.

[7]This assumption, developed in the 2012 version of this paper, is a generalization of the assumptions in Matzkin (2008, Example 4.2) 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 $\log f_\varepsilon$. (See Brown, Wegkamp, and Deb (2007) for identification in a semiparametric version of the model in Matzkin (2007b, Section 2.1.4) and Berry and Haile (2011).)

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

$$B\big(w^{(1)}, \dots, w^{(G+1)}\big) = \begin{pmatrix} \mathbf{g}^{(1)}_{x_1} & \mathbf{g}^{(1)}_{x_2} & \cdots & \mathbf{g}^{(1)}_{x_G} & 1 \\ \vdots & \vdots & \vdots & \vdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{g}^{(G+1)}_{x_1} & \mathbf{g}^{(G+1)}_{x_2} & \cdots & \mathbf{g}^{(G+1)}_{x_G} & 1 \end{pmatrix}$$

is invertible.

The existence of $G+1$ points, $w^{(1)}, \dots, w^{(G+1)}$ in $\overline{M}$, such that $B(w^{(1)}, \dots, w^{(G+1)})$ is invertible implies by Theorem 2.1 that, when Assumptions 2.1–2.3 are satisfied, $(\bar{r}, d)$ is identified on $\overline{M}$. This is because, for each g, evaluating the equation corresponding to y_g in (2.7) at $w^{(1)}, \dots, w^{(G)}$ and $w^{(G+1)}$ generates $G+1$ linear independent equations in $G+1$ unknowns, whose unique solution is the vector of 0's. Propositions 2.3 and 2.4 below show that Assumptions 2.5 and 2.6 imply that Condition I.1 is satisfied in models satisfying Assumptions 2.1–2.4 or Assumptions 2.1–2.3 and 2.4′, when $\overline{M}$ is appropriately chosen. They also show that Condition I.1 can be employed to test Assumption 2.5. Any $G+1$ points $w^{(1)}, \dots, w^{(G+1)}$ in $\overline{M}$ for which $B(w^{(1)}, \dots, w^{(G+1)})$ is invertible are mapped to $G+1$ vectors, $\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)}$ in R^G, such that $A(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)})$ is invertible.

PROPOSITION 2.3: *Suppose that Assumptions 2.1–2.3 and 2.4 are satisfied. Let the set $\overline{M}$ be included in the set $\{(y, x_{-G}, t_G) | t_G \in R\}$. Then, for all $w^{(1)}, \dots, w^{(G+1)}$ in $\overline{M}$ and $\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)}$ in R^G such that $\varepsilon^{(k)} = r(w^{(k)})$, $A(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)})$ is invertible if and only if $B(w^{(1)}, \dots, w^{(G+1)})$ is invertible.*

PROPOSITION 2.4: *Suppose that Assumptions 2.1–2.3 and 2.4′ are satisfied. Let the set $\overline{M}$ be included in the set $\{(y, t_1, \dots, t_G) | (t_1, \dots, t_G) \in R^G\}$. Then, for all $w^{(1)}, \dots, w^{(G+1)}$ in $\overline{M}$ and $\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)}$ in R^G such that $\varepsilon^{(k)} = r(w^{(k)})$, $A(\varepsilon^{(1)}, \dots, \varepsilon^{(G+1)})$ is invertible if and only if $B(w^{(1)}, \dots, w^{(G+1)})$ is invertible.*

We next define a distance function such that $(\bar{r}, d)$ is the unique solution to the minimization of this distance function. Let $\overline{M}$ denote a compact set where, as earlier, $(\bar{r}, d)$ is constant. Let $\overline{M}^t$ be such that $\overline{M} = \{y, x_{-t}\} \times \overline{M}^t$, where x_{-t} is the, possibly empty, singleton corresponding to the coordinates of x that remain fixed on $\overline{M}$. Then, when Assumption 2.4 is satisfied, $\{y, x_{-t}\} = \{y, x_{-G}\}$ and any $t \in \overline{M}^t$ is a scalar; when Assumption 2.4′ is satisfied, $\{y, x_{-t}\} = \{y\}$ and any $t \in \overline{M}^t$ is G-dimensional. Let $\mu(y, x_{-t}, t)$ denote a specified differentiable nonnegative function defined on R^{2G} such that $\int_{\overline{M}} \mu(y, x_{-t}, t)\, dt = 1$ and $\mu(y, x_{-t}, t) = 0$ on the complement, $\overline{M}^c$, of $\overline{M}$. For any vector $(\tilde{\bar{r}}, \tilde{d})$ generated from an alternative function $\tilde{r}$ satisfying the same assumptions as r, but not necessarily observationally equivalent to r, we define the distance function $S(\widetilde{\overline{r}}, \widetilde{d})$ by

$$S(\widetilde{\overline{r}}, \widetilde{d}) = \int_{\overline{M}} \left[\sum_{g=1}^{G} \left(\mathbf{g}_{y_g} - \widetilde{\overline{r}}^{1}_{y_g} \mathbf{g}_{x_1} - \widetilde{\overline{r}}^{2}_{y_g} \mathbf{g}_{x_2} - \cdots - \widetilde{\overline{r}}^{G}_{y_g} \mathbf{g}_{x_G} - \widetilde{d}_{y_g} \right)^2 \right] \times \mu(y, x_{-t}, t)\, dt.$$

The value of $S(\widetilde{\overline{r}}, \widetilde{d})$ is the integrated square distance between the left-hand side and the right-hand side of (2.11) when $(\overline{r}, d)$ is replaced by $(\widetilde{\overline{r}}, \widetilde{d})$. Since $S(\widetilde{\overline{r}}, \widetilde{d}) \geq 0$ and $S(\overline{r}, d) = 0$, $(\overline{r}, d)$ is a minimizer of $S(\cdot)$. When μ is strictly positive at $w^{(1)}, \ldots, w^{(G+1)}$ that satisfy Condition I.1, $(\overline{r}, d)$ is the unique minimizer of $S(\cdot)$. To express the first G coordinates, $\overline{r}$, of the vector $(\overline{r}, d)$ that solves the First Order Conditions for the minimization of $S(\cdot)$, we introduce some additional notation. The average of $\mathbf{g}_{y_g}$ and $\mathbf{g}_{x_g}$ over $\overline{M}$ will be denoted, respectively, by

$$\begin{aligned}
\int_{\overline{M}} \mathbf{g}_{y_g} &= \int_{\overline{M}} \mathbf{g}_{y_g}(y, x_{-t}, t) \mu(y, x_{-t}, t)\, dt \\
&= \int_{\overline{M}} \frac{\partial \log f_{Y|X=(x_{-t},t)}(y)}{\partial y_g} \mu(y, x_{-t}, t)\, dt, \\
\int_{\overline{M}} \mathbf{g}_{x_g} &= \int_{\overline{M}} \mathbf{g}_{x_g}(y, x_{-t}, t) \mu(y, x_{-t}, t)\, dt \\
&= \int_{\overline{M}} \frac{\partial \log f_{Y|X=(x_{-t},t)}(y)}{\partial x_g} \mu(y, x_{-t}, t)\, dt.
\end{aligned}$$

The averaged centered cross products between $\mathbf{g}_{y_g}$ and $\mathbf{g}_{x_s}$, and between $\mathbf{g}_{x_j}$ and $\mathbf{g}_{x_s}$, will be denoted respectively by

$$\begin{aligned}
T_{y_g, x_s} &= \int_{\overline{M}} \left(\mathbf{g}_{y_g}(y, x_{-t}, t) - \int_{\overline{M}} \mathbf{g}_{y_g} \right) \left(\mathbf{g}_{x_s}(y, x_{-t}, t) - \int_{\overline{M}} \mathbf{g}_{x_s} \right) \\
&\quad \times \mu(y, x_{-t}, t)\, dt, \\
T_{x_j, x_s} &= \int_{\overline{M}} \left(\mathbf{g}_{x_j}(y, x_{-t}, t) - \int_{\overline{M}} \mathbf{g}_{x_j} \right) \left(\mathbf{g}_{x_s}(y, x_{-t}, t) - \int_{\overline{M}} \mathbf{g}_{x_s} \right) \\
&\quad \times \mu(y, x_{-t}, t)\, dt.
\end{aligned}$$

The matrices of centered cross products, T_{XX} and T_{YX}, will be defined by

$$T_{XX} = \begin{pmatrix} T_{x_1,x_1} & T_{x_2,x_1} & \cdots & T_{x_G,x_1} \\ \vdots & \vdots & \vdots & \vdots \\ T_{x_1,x_G} & T_{x_2,x_G} & \cdots & T_{x_G,x_G} \end{pmatrix} \quad \text{and}$$

$$T_{YX} = \begin{pmatrix} T_{y_1,x_1} & T_{y_2,x_1} & \cdots & T_{y_G,x_1} \\ \vdots & \vdots & \vdots & \vdots \\ T_{y_1,x_G} & T_{y_2,x_G} & \cdots & T_{y_G,x_G} \end{pmatrix}.$$

The matrix of ratios of derivatives, R, will be defined by

$$R(\overline{r}) = \begin{pmatrix} \overline{r}^1_{y_1} & \overline{r}^1_{y_2} & \cdots & \overline{r}^1_{y_G} \\ \vdots & \vdots & \vdots & \vdots \\ \overline{r}^G_{y_1} & \overline{r}^G_{y_2} & \cdots & \overline{r}^G_{y_G} \end{pmatrix}.$$

The solution of the First Order Conditions for $\overline{r}$ results in the expression $T_{XX}R(\overline{r}) = T_{YX}$. Since $(\overline{r}, d)$ is the unique minimizer, the matrix T_{XX} must be invertible. It follows that $R(\overline{r})$ will be given by

$$R(\overline{r}) = T^{-1}_{XX} T_{YX}. \tag{2.12}$$

The following theorems establish the conditions under which $(\overline{r}, d)$ is the unique minimizer of $S(\cdot)$, which imply that $R(\overline{r})$ is given by (2.12) for the definitions of T_{XX} and T_{YX} that correspond in each case to the definition of the set $\overline{M}$.

THEOREM 2.2: *Let $(y, x_{-G}) = (y, x_1, \ldots, x_{G-1})$ be given and let the compact set $\overline{M}$ be included in the set $\{(y, x_{-G}, t_G) | t_G \in R\}$. Suppose that Assumptions 2.1–2.3 and 2.4–2.6 are satisfied, and that the nonnegative and continuous function $\mu(y, x_{-G}, t_G)$ is strictly positive at least at one set of points $w^{(1)}, \ldots, w^{(G+1)}$ satisfying Condition I.1. Then, $(\overline{r}, d)$ is the unique minimizer of*

$$S(\widetilde{\widetilde{r}}, \widetilde{d}) = \int_{\overline{M}} \left[\sum_{j=1}^{G} \left(\mathbf{g}_{y_j} - \widetilde{\widetilde{r}}^1_{y_j} \mathbf{g}_{x_1} - \widetilde{\widetilde{r}}^2_{y_j} \mathbf{g}_{x_2} - \cdots - \widetilde{\widetilde{r}}^G_{y_j} \mathbf{g}_{x_G} - \widetilde{d}_{y_j} \right)^2 \right] \times \mu(y, x_{-G}, t_G)\, dt_G$$

and $R(\overline{r})$ is given by (2.12).

THEOREM 2.3: *Let y be given and let the compact set $\overline{M}$ be included in the set $\{(y, t_1, \ldots, t_G) | (t_1, \ldots, t_G) \in R^G\}$. Suppose that Assumptions 2.1–2.3, 2.4′, and 2.5–2.6 are satisfied, and that the nonnegative and continuous function $\mu(y, t_1, \ldots, t_G)$ is strictly positive at least at one set of points $w^{(1)}, \ldots, w^{(G+1)}$ satisfying Condition I.1. Then, $(\overline{r}, d)$ is the unique minimizer of*

$$S(\widetilde{\widetilde{r}}, \widetilde{d}) = \int_{\overline{M}} \left[\sum_{j=1}^{G} \left(\mathbf{g}_{y_j} - \widetilde{\widetilde{r}}^1_{y_j} \mathbf{g}_{x_1} - \widetilde{\widetilde{r}}^2_{y_j} \mathbf{g}_{x_2} - \cdots - \widetilde{\widetilde{r}}^G_{y_j} \mathbf{g}_{x_G} - \widetilde{d}_{y_j} \right)^2 \right] \times \mu(y, t_1, \ldots, t_G)\, d(t_1, \ldots, t_G)$$

and $R(\overline{r})$ is given by (2.12).

To obtain the estimators for $R(\overline{r})$, we note that each of the elements, T_{x_j,x_s} and T_{y_j,x_s}, in the matrices T_{XX} and T_{YX}, 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, $\widehat{f}_{Y|X=x}(y)$, for $f_{Y|X=x}(y)$. Denote such estimators by $\widehat{T}_{x_j,x_s}$ and $\widehat{T}_{y_g,x_s}$, and let $\widehat{T_{XX}}$ and $\widehat{T_{YX}}$ denote the matrices whose elements are, respectively, $\widehat{T}_{x_j,x_s}$ and $\widehat{T}_{y_g,x_s}$. Then, the estimator for the matrix of ratios of derivatives $R(\overline{r})$ is defined as

$$\widehat{R(\overline{r})} = \widehat{T_{XX}}^{-1}\widehat{T_{YX}}.$$

2.4. *Asymptotic Properties of the Estimator*

In this section, we develop asymptotic properties for the estimator presented in Section 2.3, for the case when the estimator $\widehat{f}_{Y|X=x}(y)$ for the conditional density of Y given X is obtained by kernel methods. We assume that, for any $t \in \overline{M}^t$, $t = x \in R^G$. Thus, $\overline{M} = \{y\} \times \overline{M}^t$. Let $\{Y^i, X^i\}_{i=1}^N$ denote N independent and identically distributed (i.i.d.) observations generated from $f_{Y,X}$. The kernel estimator is

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

where K is a kernel function and σ_N is a bandwidth. The element in the kth row, ith column of our estimator for T_{XX} is

$$\int_{\overline{M}} \left(\widehat{\mathbf{g}}_{x_i}(y, x) - \int_{\overline{M}} \widehat{\mathbf{g}}_{x_i}\right)\left(\widehat{\mathbf{g}}_{x_k}(y, x) - \int_{\overline{M}} \widehat{\mathbf{g}}_{x_k}\right)\mu(y, x)\, dx,$$

where, for $k = 1, \ldots, G$,

$$\widehat{\mathbf{g}}_{x_k}(y, x) = \frac{\partial \log \widehat{f}_{Y|X=x}(y)}{\partial x_k} \quad \text{and}$$

$$\int \widehat{\mathbf{g}}_{x_k} = \int_{\overline{M}} \frac{\partial \log \widehat{f}_{Y|X=x}(y)}{\partial x_k}\mu(y, x)\, dx.$$

Similarly, the element in the kth row, ith column of our estimator for T_{YX} is

$$\int \left(\widehat{\mathbf{g}}_{y_i}(y, x) - \int_{\overline{M}} \widehat{\mathbf{g}}_{y_i}\right)\left(\widehat{\mathbf{g}}_{x_k}(y, x) - \int_{\overline{M}} \widehat{\mathbf{g}}_{x_k}\right)\mu(y, x)\, dx,$$

where, for $g = 1, \dots, G$,

$$\widehat{\mathbf{g}}_{y_g}(y, x) = \frac{\partial \log \widehat{f}_{Y|X=x}(y)}{\partial y_g} \quad \text{and}$$
$$\int \widehat{\mathbf{g}}_{y_g}(y) = \int_{\overline{M}} \frac{\partial \log \widehat{f}_{Y|X=x}(y)}{\partial y_g} \mu(y, x)\, dx.$$

We will let $\overline{M}^y$ denote a convex and compact set such that the value y at which we estimate r_y is an interior point of $\overline{M}^y$, and we will let $\overline{M}^x$ be a convex and compact set such that $\overline{M}^t$ is strictly in the interior of $\overline{M}^x$. Our result uses the following assumptions.

ASSUMPTION 2.7: *The density $f_{Y,X}$ generated by f_ε, f_X, and r is bounded 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 $(y, x) \in \overline{M}^y \times \overline{M}^x$, $f_X(x) > \delta$ and $f_{Y,X}(y, x) > \delta$.*

ASSUMPTION 2.8: *The set $\overline{M}^t$ is compact. The function $\mu(y, \cdot)$ is bounded and continuously differentiable. It has strictly positive values at all x belonging to the interior of $\overline{M}^t$, values and derivatives equal to zero when the value of any coordinate of x lies on the boundary of $\overline{M}^t$ and equal to zero when x belongs to the complement of $\overline{M}^t$. The set $\{y\} \times \overline{M}^t$ contains at least one set of points $w^{(1)}, \dots, w^{(G+1)}$ satisfying Condition I.1 and such that μ is strictly positive at each of those points.*

ASSUMPTION 2.9: *The kernel function K is of order s, where $s+2 \leq d$. It attains the value zero outside a compact set, integrates to 1, is differentiable of order Δ, and its derivatives of order Δ are Lipschitz, where $\Delta \geq 2$.*

ASSUMPTION 2.10: *The sequence of bandwidths, σ_N, is such that $\sigma_N \to 0$, $N\sigma_N^{G+2} \to \infty$, $\sqrt{N}\sigma_N^{(G/2)+1+s} \to 0$, $[N\sigma_N^{2G+2}/\ln(N)] \to \infty$, and $\sqrt{N}\sigma_N^{(G/2)+1} \times [\sqrt{\ln(N)/N\sigma_N^{2G+2}} + \sigma_N^s]^2 \to 0$.*

To describe the asymptotic behavior of our estimator, we will denote by rr the vector in R^{G^2} formed by stacking the columns of $R(\overline{r})$, so that $rr = \operatorname{vec}(R(\overline{r})) = (\overline{r}_{y_1}^1, \dots, \overline{r}_{y_1}^G; \overline{r}_{y_2}^1, \dots, \overline{r}_{y_2}^G; \dots; \overline{r}_{y_G}^1, \dots, \overline{r}_{y_G}^G)'$. Let $\widehat{rr}$ denote the estimator for rr. Accordingly, we will denote the matrix TT_{XX} by $I_G \otimes T_{XX}$ and its estimator $\widehat{TT_{XX}} = I_G \otimes \widehat{T_{XX}}$. The vector TT_{YX} will be the vector formed by stacking the columns of T_{YX}: $TT_{YX} = (T_{y_1,x_1}, \dots, T_{y_1,x_G}; T_{y_2,x_1}, \dots, T_{y_2,x_G}; \dots;$

$T_{y_G,x_1}, \ldots, T_{y_G x_G})'$, with its estimator defined by substituting each coordinate by its estimator. For each s, denote

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

and for each j, k, denote

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

In the proof of Theorem 2.4, which we present in the Appendix, we show that under our assumptions

$$\sqrt{N\sigma_N^{G+2}}(\widehat{TT_{YX}} - TT_{YX}) \underset{d}{\to} N(0, V_{T_{YX}}),$$

where the element in $V_{T_{YX}}$ corresponding to the covariance between $T_{y_j x_s}$ and $T_{y_k x_e}$ is

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

Denote by $\widehat{V}_{T_{YX}}$ the matrix whose elements are

$$\left\{ \int_{\overline{M}} \big(\Delta\partial_{x_s} \log \widehat{f}_{Y|X=x}(y)\big)\big(\Delta\partial_{x_e} \log \widehat{f}_{Y|X=x}(y)\big) \left(\frac{\mu(y, x)^2}{\widehat{f}_{Y,X}(y, x)} \right) dx \right\} \widetilde{KK}_{y_j,y_k}.$$

The following theorem is proved in the Appendix.

THEOREM 2.4: *Suppose that the model satisfies Assumptions 2.1–2.3, 2.4′, 2.5–2.10. Then,*

$$\sqrt{N\sigma_N^{G+2}}(\widehat{rr} - rr) \underset{d}{\to} N\big(0, (TT_{XX})^{-1} V_{T_{YX}} (TT_{XX})^{-1}\big)$$

and $(\widehat{TT_{XX}})^{-1} \widehat{V}_{T_{YX}} (\widehat{TT_{XX}})^{-1}$ *is a consistent estimator for* $(TT_{XX})^{-1} V_{T_{YX}} \times (TT_{XX})^{-1}$.

If Assumption 2.4′ is substituted by Assumption 2.4, it can be shown by adapting the assumptions and proofs of Theorems 2.4 and 3.2 that the rate of convergence of the estimator for rr is $\sqrt{N\sigma_N^{2G+1}}$.

3. THE MODEL WITH TWO EQUATIONS AND ONE INSTRUMENT

3.1. *The Model*

The model considered in this and the following section is

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

where (Y_1, Y_2, X) is observable and $(\varepsilon_1, \varepsilon_2)$ is unobservable. In this section, we will develop an estimator for $\partial m^1(y_2, \varepsilon_1)/\partial y_2$, analogous to the estimator in Section 2. In the next section, we will develop a two-step procedure to estimate $\partial m^1(y_2, \varepsilon_1)/\partial y_2$. We assume that m^1 is either strictly increasing or strictly decreasing in ε_1, to guarantee the existence of a function r^1 such that, for all (y_1, y_2), $y_1 = m^1(y_2, r^1(y_1, y_2))$. Differentiating this expression with respect to y_1 and y_2, it follows that, for any given (y_1, y_2),

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

where ε_1 is the unknown but unique value satisfying $y_1 = m^1(y_2, \varepsilon_1)$ and $r^1_{y_1}$ and $r^1_{y_2}$ denote the partial derivatives of r^1 with respect to y_1 and y_2. Similarly, we assume that m^2 is either strictly increasing or strictly decreasing in ε_2, to guarantee the existence of r^2 such that, for all (y_1, y_2, x), $y_2 = m^2(y_1, x, r^2(y_1, y_2, x))$. The additional assumptions guarantee that, for all (y_1, y_2, x) on a set M' in the support of $(Y, X) = (Y_1, Y_2, X)$,

$$(3.2)\qquad f_{Y_1,Y_2|X=x}(y_1, y_2) = f_{\varepsilon_1,\varepsilon_2}\big(r^1(y_1, y_2), r^2(y_1, y_2, x)\big)\left|\frac{\partial r(y_1, y_2, x)}{\partial(y_1, y_2)}\right|.$$

ASSUMPTION 3.1: *The function r^1 is invertible in y_1 and the function r^2 is invertible in y_2. The vector function r is twice continuously differentiable. The derivative of r^2 with respect to x is bounded away from zero. Conditional on X, the function r is* 1–1, *onto R^2, and as a function of x, the Jacobian determinant is positive and bounded away from zero.*

ASSUMPTION 3.2: $(\varepsilon_1, \varepsilon_2)$ *is distributed independently of X with an everywhere positive and twice continuously differentiable density*, f_ε.

ASSUMPTION 3.3: *X possesses a differentiable density.*

3.2. *Observational Equivalence*

Our observational equivalence result for model (3.1) involves functions, $c(y, x)$ and $\widetilde{c}(y, x)$, analogous to the functions d_{y_j} and $\widetilde{d}_{y_j}$ in Section 3. These are defined by

$$c(y,x) = -\frac{|r_y(y,x)|_x}{r^1_{y_1}(y) r^2_x(y,x)} - \frac{r^1_{y_2}(y)}{r^1_{y_1}(y)} \frac{|r_y(y,x)|_{y_1}}{|r_y(y,x)|} + \frac{|r_y(y,x)|_{y_2}}{|r_y(y,x)|} \quad \text{and} \tag{3.3}$$

$$\widetilde{c}(y,x) = -\frac{|\widetilde{r}_y(y,x)|_x}{\widetilde{r}^1_{y_1}(y) \widetilde{r}^2_x(y,x)} - \frac{\widetilde{r}^1_{y_2}(y)}{\widetilde{r}^1_{y_1}(y)} \frac{|\widetilde{r}_y(y,x)|_{y_1}}{|\widetilde{r}_y(y,x)|} + \frac{|\widetilde{r}_y(y,x)|_{y_2}}{|\widetilde{r}_y(y,x)|}.$$

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

DEFINITION 3.1: 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 δ_2 is any positive constant. Function $\widetilde{r} \in \Gamma'$ is observationally equivalent to $r \in \Gamma'$ on M' if there exist densities f_ε and $f_{\widetilde{\varepsilon}}$ satisfying Assumption 3.2 and such that, for all $(y,x) \in M'$,

$$f_\varepsilon\big(r(y,x)\big)\left|\frac{\partial r(y,x)}{\partial y}\right| = f_{\widetilde{\varepsilon}}\big(\widetilde{r}(y,x)\big)\left|\frac{\partial \widetilde{r}(y,x)}{\partial y}\right|. \tag{3.4}$$

Our observational equivalence for the model (3.1) is given in the following theorem.

THEOREM 3.1: *Suppose that* (r, f_ε) *generates* $f_{Y|X}$ *and that Assumptions 3.1–3.3 are satisfied. A function* $\widetilde{r} \in \Gamma'$ *is observationally equivalent to* $r \in \Gamma'$ *on* M' *if and only if, for all* $(y,x) \in M'$,

$$0 = \left(\frac{r^1_{y_2}}{r^1_{y_1}} - \frac{\widetilde{r}^1_{y_2}}{\widetilde{r}^1_{y_1}}\right)\mathbf{g}_{y_1} + \left(\frac{|r_y|}{r^2_x r^1_{y_1}} - \frac{|\widetilde{r}_y|}{\widetilde{r}^2_x \widetilde{r}^1_{y_1}}\right)\mathbf{g}_x + (c - \widetilde{c}), \tag{3.5}$$

where, for $g = 2$ *and all* j, $r^g_{y_j} = r^g_{y_j}(y,x)$, $\widetilde{r}^g_{y_j} = \widetilde{r}^g_{y_j}(y,x)$, $r^1_{y_j} = r^1_{y_j}(y)$, $\widetilde{r}^1_{y_j} = \widetilde{r}^1_{y_j}(y)$, $r^2_x = r^2_x(y,x)$, $\widetilde{r}^2_x = \widetilde{r}^2_x(y,x)$, $|r_y| = |r_y(y,x)|$, $|\widetilde{r}_y| = |\widetilde{r}_y(y,x)|$, $\mathbf{g}_{y_1} = \mathbf{g}_{y_1}(y,x) = \partial \log f_{Y|X=x}(y)/\partial y_1$, $\mathbf{g}_x = \mathbf{g}_x(y,x) = \partial \log f_{Y|X=x}(y)/\partial x$, *and where* $c = c(y,x)$ *and* $\widetilde{c} = \widetilde{c}(y,x)$ *are as defined in* (3.3).

When comparing Theorem 3.1 with Theorem 2.1, note that the lack of one exclusive regressor in the first equation has reduced the number of equations by one. The derivative of $\log f_{Y|X=x}(y)$ with respect to y_1 has taken up the place that the derivative of $\log f_{Y|X=x}(y)$ with respect to x_1 would have taken. The ratio of derivatives of r^1 appears as a coefficient of $\partial \log f_{Y|X=x}(y)/\partial y_1$. The two ratios of derivatives of r^2, which in Theorem 2.1 appeared separately, each as a coefficient of a different derivative of $\log f_{Y|X=x}$, appear in (3.5) in one coefficient, of the form $[((r^2_{y_2}/r^2_x) - (r^2_{y_1}/r^2_x)(r^1_{y_2}/r^1_{y_1}))]$.

The proof of Theorem 3.1, presented in the Appendix, proceeds in a way similar to the one used to prove Theorem 2.1. Equation (3.2) 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 3.2). After manipulation of the equations, the resulting expression is (3.5). 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.

Theorem 3.1 together with (3.2) can be used to develop constructive identification results for features of r. Differentiating both sides of (3.2) with respect to y_1, y_2, and x gives

$$\frac{\partial \log f_{Y|X=x}(y)}{\partial y_1} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_1} r^1_{y_1} + \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_2} r^2_{y_1} + \frac{|r_y|_{y_1}}{|r_y|}, \tag{3.6}$$

$$\frac{\partial \log f_{Y|X=x}(y)}{\partial y_2} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_1} r^1_{y_2} + \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_2} r^2_{y_2} + \frac{|r_y|_{y_2}}{|r_y|}, \tag{3.7}$$

$$\frac{\partial \log f_{Y|X=x}(y)}{\partial x} = \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_2} r^2_{x} + \frac{|r_y|_{x}}{|r_y|}. \tag{3.8}$$

Equation (3.8) reflects the fact that x is exclusive to r^2. Solving for $\partial \log f_\varepsilon(r(y,x))/\partial\varepsilon_2$ from (3.8) and substituting into (3.6) and (3.7) gives

$$\begin{aligned}\frac{\partial \log f_{Y|X=x}(y)}{\partial y_1} &= \frac{\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1,r^2)}{\partial \varepsilon_1} r^1_{y_1} + \frac{\partial \log f_{Y|X=x}(y)}{\partial x}\frac{r^2_{y_1}}{r^2_x} \\ &\quad + \frac{|r_y|_{y_1}}{|r_y|} - \frac{r^2_{y_1}}{r^2_x}\frac{|r_y|_{x}}{|r_y|},\end{aligned} \tag{3.9}$$

$$\begin{aligned}\frac{\partial \log f_{Y|X=x}(y)}{\partial y_2} &= \frac{\partial \log f_{Y|X=x}(y)}{\partial \varepsilon_1} r^1_{y_2} + \frac{\partial \log f_{Y|X=x}(y)}{\partial x}\frac{r^2_{y_2}}{r^2_x} \\ &\quad + \frac{|r_y|_{y_2}}{|r_y|} - \frac{r^2_{y_2}}{r^2_x}\frac{|r_y|_{x}}{|r_y|}.\end{aligned} \tag{3.10}$$

Solving for $\partial \log f_\varepsilon(r(y,x))/\partial\varepsilon_1$ from (3.9) and substituting into (3.10) gives

$$\begin{aligned}\frac{\partial \log f_{Y|X=x}(y)}{\partial y_2} &= \frac{r^1_{y_2}}{r^1_{y_1}}\frac{\partial \log f_{Y|X=x}(y)}{\partial y_1} \\ &\quad - \left[\frac{r^1_{y_2}}{r^1_{y_1}}\frac{r^2_{y_1}}{r^2_x} - \frac{r^2_{y_2}}{r^2_x}\right]\frac{\partial \log f_{Y|X=x}(y)}{\partial x} + c(y,x).\end{aligned} \tag{3.11}$$

Rearranging terms, and substituting $\partial \log f_{Y|X=x}(y)/\partial y_j$ by $\mathbf{g}_{y_j}$ and $\partial \log f_{Y|X=x}(y)/\partial x$ by $\mathbf{g}_x$, we get that

$$\mathbf{g}_{y_2}(y,x) = \frac{r^1_{y_2}}{r^1_{y_1}}\mathbf{g}_{y_1}(y,x) + \left[\frac{|r_y(y,x)|}{r^1_{y_1} r^2_x}\right]\mathbf{g}_x(y,x) + c(y,x). \tag{3.12}$$

Equation (3.12) together with (3.5) will be referred to in later sections, to build upon them estimators for $\partial m^1(y_2,\varepsilon_1)/\partial y_2 = -r^1_{y_2}/r^1_{y_1}$.

3.3. *Average Derivatives Estimator for the Two Equations, One Instrument Model*

The result of the previous subsection can be used to obtain an estimator for the unknown values of the coefficients in equation (3.12), in analogy to the estimator developed for the exclusive regressors model. Let y be a given value of the endogenous variables. Suppose it is known that Assumption 2.4 is satisfied or, more generally, that on a subset $\overline{M}^t$ in the support of X it is the case that $|r_y(y,x)|/[r^1_{y_1} r^2_x]$ and $c(y,x)$ are constant over x. Suppose also that the following condition is satisfied.

CONDITION I.2: There exist $x^{(1)}$, $x^{(2)}$, and $x^{(3)}$ in $\overline{M}^t$ such that the rank of the matrix

$$\begin{pmatrix} \mathbf{g}^{(1)}_{y_1} & \mathbf{g}^{(1)}_{x_2} & 1 \\ \mathbf{g}^{(2)}_{y_1} & \mathbf{g}^{(2)}_{x_2} & 1 \\ \mathbf{g}^{(3)}_{y_1} & \mathbf{g}^{(3)}_{x_2} & 1 \end{pmatrix}$$

is 3, where for $i = 1,2,3$, $\mathbf{g}^{(i)}_{y_1} = \partial \log f_{Y|X=x^{(i)}}(y)/\partial y_1$ and $\mathbf{g}^{(i)}_x = \partial \log f_{Y|X=x^{(i)}}(y)/\partial x$.

Let $\mu(y_1,y_2,\cdot)$ be a continuous, nonnegative function with positive values at the points $x^{(1)}$, $x^{(2)}$, and $x^{(3)}$ and such that $\int_{\overline{M}} \mu(y_1,y_2,x)\,dx = 1$, where $\overline{M} = \{(y_1,y_2)\} \times \overline{M}^t$. Denote $\beta(y) = r^1_{y_2}(y)/r^1_{y_1}(y)$, $\gamma(y) = |r_y(y,x)|/[r^1_{y_1} r^2_x]$, and $\nu(y) = c(y,x)$. In analogy to the development in the previous section, and to the proof of Theorem 2.2, the rank condition and the continuity of μ imply that the vector $(\widetilde{\beta},\widetilde{\gamma},\widetilde{\nu}) = (\beta(y),\gamma(y),\nu(y))$ is the unique minimizer of the function $S(\widetilde{\beta},\widetilde{\gamma},\widetilde{\nu})$ defined by

$$S(\widetilde{\beta},\widetilde{\gamma},\widetilde{\nu}) = \int_{\overline{M}} \left(\mathbf{g}_{y_2}(y,x) - \widetilde{\beta}\mathbf{g}_{y_1}(y,x) - \widetilde{\gamma}\mathbf{g}_x(y,x) - \widetilde{\nu}\right)^2 \mu(y,x)\,dx.$$

The first order conditions of this minimization are given by

$$\begin{bmatrix} \int \mathbf{g}_{y_1}\mathbf{g}_{y_1} & \int \mathbf{g}_{y_1}\mathbf{g}_{x} & \int \mathbf{g}_{y_1} \\ \int \mathbf{g}_{y_1}\mathbf{g}_{x} & \int \mathbf{g}_{x}\mathbf{g}_{x} & \int \mathbf{g}_{x} \\ \int \mathbf{g}_{y_1} & \int \mathbf{g}_{x} & 1 \end{bmatrix} \begin{bmatrix} \beta(y) \\ \gamma(y) \\ \nu(y) \end{bmatrix} = \begin{bmatrix} \int \mathbf{g}_{y_2}\mathbf{g}_{y_1} \\ \int \mathbf{g}_{y_2}\mathbf{g}_{x} \\ \int \mathbf{g}_{y_2} \end{bmatrix},$$

where for $w, z \in \{y_1, y_2, x\}$, $\int \mathbf{g}_w\mathbf{g}_z = \int_{\overline{M}} \mathbf{g}_w(y,x)\mathbf{g}_z(y,x)\mu(y,x)\,dx$, and $\int \mathbf{g}_w = \int_{\overline{M}} \mathbf{g}_w(y,x)\mu(y,x)\,dx$. The 3×3 matrix is the Hessian of the function S, which is constant over $(\widetilde{\beta}, \widetilde{\gamma}, \widetilde{\nu})$. By the convexity of S and the uniqueness of a minimizer, this matrix is positive definite and therefore invertible. Solving for $\nu(y)$ and substituting into the first equation, we get

$$\begin{bmatrix} T_{y_1,y_1} & T_{y_1,x} \\ T_{y_1,x} & T_{x,x} \end{bmatrix} \begin{bmatrix} \beta(y) \\ \gamma(y) \end{bmatrix} = \begin{bmatrix} T_{y_1,y_2} \\ T_{y_2,x} \end{bmatrix},$$

where the 2×2 matrix is positive definite. Solving for $\beta(y)$, we get

$$\beta(y) = \frac{T_{y_1,y_2}T_{x,x} - T_{y_1,x}T_{y_2,x}}{T_{y_1,y_1}T_{x,x} - (T_{y_1,x})^2},$$

where the denominator is strictly positive. Replacing the terms in the expression for $\beta(y)$ by nonparametric estimators, we obtain the following nonparametric estimator for $\beta(y)$:

$$\widehat{\beta}(y) = \frac{\widehat{T}_{y_1,y_2}\widehat{T}_{x,x} - \widehat{T}_{y_1,x}\widehat{T}_{y_2,x}}{\widehat{T}_{y_1,y_1}\widehat{T}_{x,x} - (\widehat{T}_{y_1,x})^2}.$$

In the next subsection, we develop the asymptotic properties of $\widehat{\beta}(y)$ when the estimators for $\widehat{T}_{y_1,y_2}$, $\widehat{T}_{x,x}$, $\widehat{T}_{y_1,x}$, $\widehat{T}_{y_2,x}$, $\widehat{T}_{y_1,y_1}$, and $\widehat{T}_{x,x}$ are obtained by replacing $f_{Y|X}(y)$ by a kernel estimator for $f_{Y|X}(y)$. By the relationship between the derivatives of r^g and m^g, shown in Section 2.1, it follows that, since $\beta = r^1_{y_2}(y)/r^1_{y_1}(y)$,

$$\frac{\partial \widehat{m^1(y_2, \varepsilon_1)}}{\partial y_2} = -\widehat{\beta}(y) = -\frac{\widehat{T}_{y_1,y_2}\widehat{T}_{x,x} - \widehat{T}_{y_1,x}\widehat{T}_{y_2,x}}{\widehat{T}_{y_1,y_1}\widehat{T}_{x,x} - (\widehat{T}_{y_1,x})^2}.$$

Moreover, since $m^2_x(y_1, x, \varepsilon_2)/m^2_{\varepsilon_2}(y_1, x, \varepsilon_2) = r^2_x(y_1, y_2, x)$, one can consider restrictions on m^2 guaranteeing that the coefficients $\gamma(y)$ and $\nu(y)$ are constant over x.

If the derivative $r_x^2(y,x)$ equals 1 for all x, $r^2(y,x)$ is of the form $r^2(y,x) = s(y)+x$, and then the coefficients $\gamma(y)$ and $\nu(y)$ are constant over x. When $r^2(y,x)=s(y)+x$, m^2 is of the form $y_2 = m^2(y_1, \varepsilon_2 - x)$. In analogy to Proposition 2.3, Condition I.2 is satisfied in this case if and only if, for $\varepsilon_1 = r^1(y_1,y_2)$, and for $\varepsilon_2^{(1)} = s(y)+x^{(1)}$, $\varepsilon_2^{(2)} = s(y)+x^{(2)}$, and $\varepsilon_2^{(3)} = s(y)+x_2^{(3)}$, the following matrix has rank 3:

$$\begin{pmatrix} \dfrac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1,\varepsilon_2^{(1)})}{\partial \varepsilon_1} & \dfrac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1,\varepsilon_2^{(1)})}{\partial \varepsilon_2} & 1 \\ \dfrac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1,\varepsilon_2^{(2)})}{\partial \varepsilon_1} & \dfrac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1,\varepsilon_2^{(2)})}{\partial \varepsilon_2} & 1 \\ \dfrac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1,\varepsilon_2^{(3)})}{\partial \varepsilon_1} & \dfrac{\partial \log f_{\varepsilon_1,\varepsilon_2}(\varepsilon_1,\varepsilon_2^{(3)})}{\partial \varepsilon_2} & 1 \end{pmatrix}.$$

3.4. *Asymptotic Properties of the Estimator*

Suppose that when calculating $\widehat{\beta}(y)$, $\widehat{T}_{y_1,y_2}$, $\widehat{T}_{x,x}$, $\widehat{T}_{y_1,x}$, $\widehat{T}_{y_2,x}$, $\widehat{T}_{y_1,y_1}$, and $\widehat{T}_{x,x}$ are obtained by replacing $f_{Y|X}(y)$ by a kernel estimator for $f_{Y|X}(y)$. As in Section 2.4, define

$$\Delta\partial_x \log f_{Y|X=x}(y) = \frac{\partial \log f_{Y|X=x}(y)}{\partial x} - \int_{\overline{M}} \frac{\partial \log f_{Y|X=x}(y)}{\partial x}\mu(y,x)\,dx,$$

and for $j=1,2$, define

$$\Delta\partial_{y_j} \log f_{Y|X=x}(y) = \frac{\partial \log f_{Y|X=x}(y)}{\partial y_j} - \int_{\overline{M}} \frac{\partial \log f_{Y|X=x}(y)}{\partial y_j}\mu(y,x)\,dx.$$

Let

$$\widetilde{\omega}_a(y,x) = \big(\Delta\partial_{y_2}\log f_{Y|X=x}(y)\big)\left(\frac{\mu(y,x)}{f}\right)\left[\frac{T_{x,x}}{T_{y_1,y_1}T_{x,x} - T_{y_1,x}^2}\right],$$

$$\widetilde{\omega}_b(y,x) = \big(\Delta\partial_{y_1}\log f_{Y|X=x}(y)\big)\left(\frac{\mu(y,x)}{f}\right)\left[\frac{T_{x,x}}{T_{y_1,y_1}T_{x,x} - T_{y_1,x}^2}\right],$$

$$\widetilde{\omega}_c(y,x) = -\big(\Delta\partial_{x}\log f_{Y|X=x}(y)\big)\left(\frac{\mu(y,x)}{f}\right)\left[\frac{T_{y_2,x}}{T_{y_1,y_1}T_{x,x} - T_{y_1,x}^2}\right],$$

$$\widetilde{\omega}_d(y,x) = -\big(\Delta\partial_{x}\log f_{Y|X=x}(y)\big)\left(\frac{\mu(y,x)}{f}\right)\left[\frac{T_{y_1,x}}{T_{y_1,y_1}T_{x,x} - T_{y_1,x}^2}\right],$$

$$\begin{aligned}\widetilde{\omega}_e(y,x) &= -2\big(\Delta\partial_{y_1}\log f_{Y|X=x}(y)\big)\\ &\quad\times\left(\frac{\mu(y,x)}{f}\right)\frac{T_{x,x}[T_{y_1,y_2}T_{x,x} - T_{y_1,x}T_{y_2,x}]}{[T_{y_1,y_1}T_{x,x} - T_{y_1,x}^2]^2},\end{aligned}$$

$$\widetilde{\omega}_f(y,x) = 2\big(\Delta\partial_x \log f_{Y|X=x}(y)\big) \times \left(\frac{\mu(y,x)}{f}\right)\frac{T_{y_1,x}[T_{y_1,y_2}T_{x,x} - T_{y_1,x}T_{y_2,x}]}{[T_{y_1,y_1}T_{x,x} - T^2_{y_1,x}]^2}.$$

Define

$$\omega_1(x) = \big[\widetilde{\omega}_a(y,x) + \widetilde{\omega}_c(y,x) + \widetilde{\omega}_e(y,x) + \widetilde{\omega}_f(y,x)\big],$$
$$\omega_2(x) = \big[\widetilde{\omega}_b(y,x) + \widetilde{\omega}_d(y,x)\big],$$
$$\omega(x) = \big(\omega_1(x), \omega_2(x)\big).$$

Define the estimator $\widehat{\omega}(x) = (\widehat{\omega}_1(x), \widehat{\omega}_2(x))$ for $\omega(x)$ by substituting f by $\widehat{f}$ and T by $\widehat{T}$ in the definitions of $\widetilde{\omega}_a, \widetilde{\omega}_b, \dots, \widetilde{\omega}_f$, For $u = (u_1, u_2)$, denote

$$\overline{K}_y = \left[\int\left[\int\left(\frac{\partial K(u,x)}{\partial u}\right)dx\right]\left[\int\left(\frac{\partial K(u,x)}{\partial u}\right)dx\right]' du\right],$$
$$V_\beta(y) = \left[\int_{\overline{M}} \omega(x)\overline{K}_y\omega(x)' f(y,x)\,dx\right], \quad \text{and}$$
$$\widehat{V}_\beta(y) = \int_{\overline{M}} \widehat{\omega}(x)\overline{K}_y\widehat{\omega}(x)'\widehat{f}(y,x)\,dx.$$

We will make the following assumptions:

ASSUMPTION 3.4: *There exists a known convex and compact set $\overline{M}^t$ in the interior of the support of X on which the values of $|r_y(y,x)|/[r^1_{y_1}(y)r^2_x(y,x)]$ and $c(y,x)$ are constant.*

ASSUMPTION 3.5: *There exist at least one set of points $x^{(1)}$, $x^{(2)}$, $x^{(3)}$ in $\overline{M}^t$ satisfying Condition I.2. The function $\mu(y,x) = \mu(y_1,y_2,x)$ is bounded and continuously differentiable, with values and derivatives equal to zero when x is on the boundary and on the complement of $\overline{M}^t$ and with strictly positive values at $x^{(1)}$, $x^{(2)}$, $x^{(3)}$.*

The asymptotic behavior of $\widehat{\beta}(y)$ defined in the previous section is established in the following theorem.

THEOREM 3.2: *Suppose that model (3.1) satisfies Assumptions 3.1–3.5, 2.7, 2.9, and 2.10. Then*

$$\sqrt{N\sigma_N^4}\big(\widehat{\beta}(y) - \beta(y)\big) \underset{d}{\to} N\big(0, V_\beta(y)\big)$$

and $\widehat{V}_\beta(y) \to V_\beta(y)$ in probability.

4. INDIRECT ESTIMATORS

The estimators developed in Sections 2 and 3 averaged the information provided by the conditional density over a set of values of the exogenous observable variables. In contrast, the indirect estimators developed in this section focus only on one or two values of the exogenous variables. Those values are such that when the conditional density of the endogenous variables is evaluated at those values, one can directly read off the value of the elements of interest. Since the conditions for identification are not be necessarily the same, indirect estimators may provide in some cases a useful alternative to the minimum distance estimators developed in Sections 2 and 3. In addition, by focusing on the particular values of the observable exogenous variables at which one can read off a structural element of interest, the estimators developed here allow one to obtain more insight into the relation between both the exogenous and endogenous variables. Although we develop indirect estimators for the two equations, one instrument model, analogous two-step indirect estimators can be obtained for the exclusive regressors model of Section 2, employing (2.11).

4.1. *Indirect Estimators for the Two-Equations, One Instrument Model*

We develop in this section two indirect estimators, one based on first derivatives and a second based on second derivatives, for the two equations, one instrument model,

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

where interest lies on the derivative $\partial m^1(y_2, \varepsilon_1)/\partial y_2$, for the value of $\varepsilon_1 = r^1(y_1, y_2)$. Both estimators are derived from equation (3.12). 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 will assume, as in Section 3, that Assumption 2.4 is satisfied. This implies that the value of the vector $\overline{b} = (\overline{b}_1, \overline{b}_2, \overline{b}_3)$, with $\overline{b}_1 = (r^1_{y_2}(y)/r^1_{y_1}(y))$, $\overline{b}_2 = |r_y(y,x)|/(r^1_{y_1}(y,x) r^2_x(y,x))$, and $\overline{b}_3 = c(y,x)$, is constant over the set $\overline{M}' \subset \{(y,t) | t \in R\}$, where $c(y,x)$ is as defined in (3.3). (The proof is as that of Proposition 2.1 after substituting (y, x_{-G}) in that proposition by y.) We will consider two sets of assumptions, which substitute for Assumption 2.5. Assumption 4.5 implies invertibility of a 2×2 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 ε 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 3.1 to obtain expressions for $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ in terms of the values of the derivatives or second order derivatives of $\log f_{Y|X=x}$ at particular values of x.

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 ε_2 such that*

$$\frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_2} = \frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^{**}(\varepsilon_1))}{\partial \varepsilon_2} \quad \text{and} \tag{4.1}$$

$$\frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_1} \neq \frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^{**}(\varepsilon_1))}{\partial \varepsilon_1}. \tag{4.2}$$

ASSUMPTION 4.5′: *Let y be given and fixed and let $\varepsilon_1 = r^1(y_1, y_2)$. There exists a value $\varepsilon_2^*(\varepsilon_1)$ of ε_2 such that*

$$\frac{\partial^2 \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_2^2} = 0 \quad \text{and} \quad \frac{\partial^2 \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*(\varepsilon_1))}{\partial \varepsilon_2 \, \partial \varepsilon_1} \neq 0. \tag{4.3}$$

ASSUMPTION 4.6: *Let y be given and fixed and let $\varepsilon_1 = r^1(y_1, y_2)$. There exist distinct values x^* and x^{**} such that $(y, x^*), (y, x^{**}) \in \overline{M}'$ 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 $(y, x^*) \in \overline{M}'$ and such that, for $\varepsilon_2^*(\varepsilon_1)$ as in Assumption 4.5′, $\varepsilon_2^* = r^2(y_1, y_2, x^*)$.*

The following propositions provide characterizations of Assumptions 4.5–4.6 and 4.5′–4.6′ in terms of conditions on $f_{Y|X=x}$.

PROPOSITION 4.1: *Let y be given and fixed and let $\varepsilon_1 = r^1(y_1, y_2)$. Suppose that Assumptions 3.1–3.3 and 2.4 are satisfied. Assumptions 4.5–4.6 are satisfied if and only if there exist x^* and x^{**} such that $(y, x^*), (y, x^{**}) \in \overline{M}'$,*

$$\begin{aligned} &\frac{\partial \log f_{Y|X=x^*}(y)}{\partial x} = \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial x} \quad \text{and} \\ &\frac{\partial \log f_{Y|X=x^*}(y)}{\partial y_1} \neq \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1}. \end{aligned} \tag{4.4}$$

PROPOSITION 4.2: *Let y be given and fixed and let $\varepsilon_1 = r^1(y_1, y_2)$. Suppose that Assumptions 3.1–3.3 and 2.4 are satisfied. Assumptions 4.5′–4.6′ are satisfied if and only if there exists x^* such that $(y, x^*) \in \overline{M}'$,*

$$\frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x^2} = 0 \quad \text{and} \quad \frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x\, \partial y_1} \neq 0. \tag{4.5}$$

We next employ the implications of Propositions 4.1 and 4.2, together with Theorem 3.1 and equation (3.12), 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 order derivatives of $\log f_{Y|X=x}$. We state these expressions in Theorems 4.1 and 4.2.

THEOREM 4.1: *Let y be given and fixed and let $\varepsilon_1 = r^1(y_1, y_2)$. Suppose that Assumptions 3.1–3.3 and 2.4 are satisfied. Let x^* and x^{**} be any distinct values of X such that $(y, x^*), (y, x^{**}) \in \overline{M}'$,*

$$\frac{\partial \log f_{Y|X=x^*}(y)}{\partial x} = \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial x} \quad \text{and} \tag{4.6}$$
$$\frac{\partial \log f_{Y|X=x^*}(y)}{\partial y_1} \neq \frac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1}$$

hold. Then,

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = \frac{-r^1_{y_2}(y)}{r^1_{y_1}(y)} = \frac{\dfrac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_2} - \dfrac{\partial \log f_{Y|X=x^*}(y)}{\partial y_2}}{\dfrac{\partial \log f_{Y|X=x^*}(y)}{\partial y_1} - \dfrac{\partial \log f_{Y|X=x^{**}}(y)}{\partial y_1}}. \tag{4.7}$$

PROOF: Let $\overline{b} = (\overline{b}_1, \overline{b}_2, \overline{b}_3)$, where $\overline{b}_1 = r^1_{y_2}(y)/r^1_{y_1}(y)$, $\overline{b}_2 = |r_y(y, x)|/(r^1_{y_1}(y) r^2_x(y, x))$, and where $\overline{b}_3 = c(y, x)$, is as defined in (3.3). By Assumptions 3.1–3.3, $\overline{b}$ satisfies (3.12). By Assumption 2.4, $\overline{b}$ is constant over the set $\{(y, t) | t \in R\}$, since Assumption 2.4 implies that for all x, $r^2_x(y, x) = 1$ and $|r_y(y, x)|$ is not a function of x. For any two values $x^{(1)}$ and $x^{(2)}$ of x, such that $(y, x^{(1)}), (y, x^{(2)}) \in \overline{M}'$, let $\mathbf{g}^{(k)}_{y_2} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial y_2$, $\mathbf{g}^{(k)}_{y_1} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial y_1$, and $\mathbf{g}^{(k)}_x = \partial \log f_{Y|X=x^{(k)}}(y)/\partial x$ for $k = 1, 2$. By (3.12),

$$\mathbf{g}^{(1)}_{y_2} = \overline{b}_1 \mathbf{g}^{(1)}_{y_1} + \overline{b}_2 \mathbf{g}^{(1)}_x + \overline{b}_3 \quad \text{and}$$
$$\mathbf{g}^{(2)}_{y_2} = \overline{b}_1 \mathbf{g}^{(2)}_{y_1} + \overline{b}_2 \mathbf{g}^{(2)}_x + \overline{b}_3.$$

Subtracting one from the other, we get that

$$\left(\mathbf{g}^{(1)}_{y_2} - \mathbf{g}^{(2)}_{y_2}\right) = \overline{b}_1 \left(\mathbf{g}^{(1)}_{y_1} - \mathbf{g}^{(2)}_{y_1}\right) + \overline{b}_2 \left(\mathbf{g}^{(1)}_x - \mathbf{g}^{(2)}_x\right).$$

If $(\mathbf{g}_x^{(1)} - \mathbf{g}_x^{(2)}) = 0$, the equation becomes

$$\big(\mathbf{g}_{y_2}^{(1)} - \mathbf{g}_{y_2}^{(2)}\big) = \overline{b}_1\big(\mathbf{g}_{y_1}^{(1)} - \mathbf{g}_{y_1}^{(2)}\big).$$

If, in addition, $(\mathbf{g}_{y_1}^{(1)} - \mathbf{g}_{y_1}^{(2)}) \neq 0$, we have that

$$\overline{b}_1 = \big(\mathbf{g}_{y_1}^{(1)} - \mathbf{g}_{y_1}^{(2)}\big)^{-1}\big(\mathbf{g}_{y_2}^{(1)} - \mathbf{g}_{y_2}^{(2)}\big).$$

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,

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = -\left(\frac{r_{y_2}^1}{r_{y_1}^1}\right) = -\frac{(\mathbf{g}_{y_2}^{(1)} - \mathbf{g}_{y_2}^{(2)})}{(\mathbf{g}_{y_1}^{(1)} - \mathbf{g}_{y_1}^{(2)})}.$$

Replacing $x^{(1)}$ by x^* and $x^{(2)}$ by x^{**}, equation (4.7) follows. This completes the proof of Theorem 4.1. *Q.E.D.*

THEOREM 4.2: *Let y be given and fixed and let $\varepsilon_1 = r^1(y_1, y_2)$. Suppose that Assumptions 3.1–3.3 and 2.4 are satisfied. Let x^* be a value of X such that $(y, x^*) \in \overline{M}'$,*

$$\frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x^2} = 0 \quad \textit{and} \quad \frac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x\, \partial y_1} \neq 0. \tag{4.8}$$

Then,

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = \frac{-\dfrac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x\, \partial y_2}}{\dfrac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x\, \partial y_1}}. \tag{4.9}$$

PROOF: As in the proof of Theorem 4.1, we let $\overline{b} = (\overline{b}_1, \overline{b}_2, \overline{b}_3)$, where $\overline{b}_1 = r_{y_2}^1(y)/r_{y_1}^1(y)$, $\overline{b}_2 = |r_y(y, x)|/(r_{y_1}^1(y) r_x^2(y, x))$, and where $\overline{b}_3 = c(y, x)$, is as defined in (3.3). We note that Assumption 2.4 implies that $\overline{b}$ is constant over the set $\{(y, t) | t \in R\}$. Hence by Assumptions 3.1–3.3 and 2.4, it follows by (3.12) that for any value $x^{(1)}$ such that $(y, x^{(1)}) \in \overline{M}'$,

$$\mathbf{g}_{y_2}^{(1)} = \overline{b}_1 \mathbf{g}_{y_1}^{(1)} + \overline{b}_2 \mathbf{g}_x^{(1)} + \overline{b}_3,$$

where for $k = 1$, $\mathbf{g}_{y_2}^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial y_2$, $\mathbf{g}_{y_1}^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial y_1$, and $\mathbf{g}_x^{(k)} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial x$. Since $\overline{b}$ is constant over x, taking derivatives of this equation with respect to x gives

$$\mathbf{g}_{x,y_2}^{(1)} = \overline{b}_1 \mathbf{g}_{x,y_1}^{(1)} + \overline{b}_2 \mathbf{g}_{x,x}^{(1)},$$

where $\mathbf{g}^{(k)}_{x,y_2} = \partial^2 \log f_{Y|X=x^{(k)}}(y)/\partial x\,\partial y_2$, $\mathbf{g}^{(k)}_{x,y_1} = \partial \log f_{Y|X=x^{(k)}}(y)/\partial x\,\partial y_1$, and $\mathbf{g}^{(k)}_{x,x} = \partial^2 \log f_{Y|X=x^{(k)}}(y)/\partial x\,\partial x$.

If $\mathbf{g}^{(1)}_{x,x} = 0$ and $\mathbf{g}^{(1)}_{x,y_1} \neq 0$, we have that

$$\overline{b}_1 = \big(\mathbf{g}^{(1)}_{x,y_1}\big)^{-1}\big(\mathbf{g}^{(1)}_{x,y_2}\big).$$

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

$$\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2} = -\left(\frac{r^1_{y_2}}{r^1_{y_1}}\right) = -\frac{\mathbf{g}^{(1)}_{x,y_2}}{\mathbf{g}^{(1)}_{x,y_1}}.$$

Replacing $x^{(1)}$ by x^*, equation (4.9) follows. This completes the proof of Theorem 4.2. *Q.E.D.*

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 3.1–3.3, 2.4, and 4.5′–4.6′ are satisfied, 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) at which we want to estimate $\partial m^1(y_2,\varepsilon_1)/\partial y_2$. The next step consists of finding a value $\widehat{x}^*$ of x satisfying

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

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

$$\frac{\partial \widehat{m^1(y_2,\varepsilon_1)}}{\partial y_2} = \frac{-\dfrac{\partial^2 \widehat{\log f_{Y|X=\widehat{x}^*}(y)}}{\partial x\,\partial y_2}}{\dfrac{\partial^2 \widehat{\log f_{Y|X=\widehat{x}^*}(y)}}{\partial x\,\partial y_1}}. \tag{4.10}$$

We show below that 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 $\partial^2 \log f_{Y|X=x}(y)/\partial x^2$ 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 is consistent and asymptotically normal.

When instead of Assumptions 4.5′–4.6′, we make Assumptions 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 $\widehat{x}^*$ and $\widehat{x}^{**}$ of x satisfying

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

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

$$\frac{\partial \widehat{m^1(y_2, \varepsilon_1)}}{\partial y_2} = \frac{\dfrac{\partial \log \widehat{f_{Y|X=\widehat{x}^{**}}(y)}}{\partial y_2} - \dfrac{\partial \log \widehat{f_{Y|X=\widehat{x}^*}(y)}}{\partial y_2}}{\dfrac{\partial \log \widehat{f_{Y|X=\widehat{x}^*}(y)}}{\partial y_1} - \dfrac{\partial \log \widehat{f_{Y|X=\widehat{x}^{**}}(y)}}{\partial y_1}}. \tag{4.11}$$

We develop below the asymptotic behavior of this estimator when the values of x^* and x^{**} are known to be such that $\partial \log f_{Y|X=x^*}(y)/\partial x = \partial \log f_{Y|X=x^{**}}(y)/\partial x = 0$. These correspond to values ε_2^* and ε_2^{**} at which $\partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^*)/\partial \varepsilon_2 = \partial \log f_\varepsilon(\varepsilon_1, \varepsilon_2^{**})/\partial \varepsilon_2 = 0$. We show that the indirect estimator defined in this way is consistent and asymptotically normal.

4.2. *Asymptotic Properties of the Indirect Estimators for the Two Equations, One Instrument Model*

To derive the asymptotic properties of the estimator defined in (4.10), we make the following assumptions.

ASSUMPTION 4.7: *The density f_ε and the density $f_{Y,X}$ generated by f_ε, f_X, and r are bounded 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 $\partial^2 \log f_\varepsilon(r^1(y_1, y_2), r^2(y_1, y_2, x'))/\partial \varepsilon_2^2 = 0$, there exist neighborhoods $B'_{y,x}$ of (y_1, y_2, x') and B'_x of x' such that the density $f_X(x)$ and the density $f_{Y,X}(y, x) = f_\varepsilon(r^1(y_1, y_2), r^2(y_1, y_2, x))|r_y(y_1, y_2, x)|f_X(x)$ are uniformly bounded away from zero on, respectively, B'_x and $B'_{y,x}$ and $\partial^3 \log f_\varepsilon(r^1(y_1, y_2), r^2(y_1, y_2, x))/\partial \varepsilon_2^3$ is bounded away from zero on those neighborhoods.*

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

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

ASSUMPTION 4.11: *The sequence of bandwidths*, σ_N, *is such that* $\sigma_N \to 0$, $\sqrt{N\sigma_N^{7+2s}} \to 0$, $\sqrt{N\sigma_N^7} \to \infty$, $[N\sigma_N^9/\ln(N)] \to \infty$, *and* $\sqrt{N\sigma_N^7} \times [\sqrt{\ln(N)/N\sigma_N^9} + \sigma_N^s]^2 \to 0$.

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

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

$$\omega_2 = \frac{-1}{\dfrac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial y_1 \, \partial x} f_{Y,X}(y, x^*)};$$

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

Let

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

and let

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

be an estimator for $\widetilde{V}$, obtained by substituting f by $\widehat{f}$ in the definitions of ω_1, ω_2, and ω_3, and substituting x^* by $\widehat{x}^*$. In the Appendix, we prove the following.

THEOREM 4.3: *Suppose that the model satisfies Assumptions 3.1–3.3, 2.4, 4.5′–4.6′, and 4.7–4.11. Let the estimator for* $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ *be as defined in* (4.10). *Then*,

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

and $\widehat{\widetilde{V}}$ *is a consistent estimator for* $\widetilde{V}$.

To derive the asymptotic properties of the estimator defined in (4.11) under the additional assumption that $\partial \log f_{Y|X=x^*}(y)/\partial x = \partial \log f_{Y|X=x^{**}}(y)/\partial x = 0$, we make the following assumptions.

ASSUMPTION 4.7′: *The density* f_ε *and the density* $f_{Y,X}$ *generated by* f_ε, f_X, *and* r *are bounded 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', x'' *such that* $\partial \log f_\varepsilon(r^1(y_1, y_2), r^2(y_1, y_2, x'))/\partial\varepsilon_2 = 0$ *and* $\partial \log f_\varepsilon(r^1(y_1, y_2), r^2(y_1, y_2, x''))/\partial\varepsilon_2 = 0$, *there exist neighborhoods* $B'_{y,x}$ *of* (y_1, y_2, x'), $B''_{y,x}$ *of* (y_1, y_2, x''), B'_x *of* x', *and* B''_x *of* x'' *such that the density* $f_X(x)$ *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''_x, *and on* $B'_{y,x}$ *and* $B''_{y,x}$. *Moreover*, $\partial^2 \log f_\varepsilon(r^1(y_1, y_2), r^2(y_1, y_2, x))/\partial\varepsilon_2^2$ *is bounded away from zero on those neighborhoods*.

ASSUMPTION 4.9′: *For any two values* x', x'' *such that* $\partial \log f_\varepsilon(r^1(y_1, y_2), r^2(y_1, y_2, x'))/\partial\varepsilon_2 = 0$, *and* $\partial \log f_\varepsilon(r^1(y_1, y_2), r^2(y_1, y_2, x''))/\partial\varepsilon_2 = 0$, $(\partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2\prime})/\partial\varepsilon_1 - \partial \log f_{\varepsilon_1,\varepsilon_2}(r^1, r^{2''})/\partial\varepsilon_1)$ *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* Δ, *and its derivatives of order* Δ *are Lipschitz, where* $\Delta \geq 4$.

ASSUMPTION 4.11′: *The sequence of bandwidths*, σ_N, *is such that* $\sqrt{N\sigma_N^5} \to \infty$, $\sqrt{N\sigma_N^5}\sigma^s \to 0$, $[\sqrt{\ln(N)/N\sigma_N^7} + \sigma_N^s] \to 0$, *and* $\sqrt{N\sigma_N^5}[\sqrt{\ln(N)/N\sigma_N^7} + \sigma_N^s]^2 \to 0$.

Define $\overline{K}_{y_1}(\widetilde{y}_1, \widetilde{y}_2, x) = \frac{\partial K(\widetilde{y}_1, \widetilde{y}_2, x)}{\partial y_1}$, $\overline{K}_{y_2}(\widetilde{y}_1, \widetilde{y}_2, x) = \frac{\partial K(\widetilde{y}_1, \widetilde{y}_2, x)}{\partial y_2}$, and $\overline{K}_x(\widetilde{y}_1, \widetilde{y}_2, x) = \frac{\partial K(\widetilde{y}_1, \widetilde{y}_2, x)}{\partial x}$. Let $\widetilde{K}(\widetilde{y}_1, \widetilde{y}_2, x)$ denote the 3×1 vector $(\overline{K}_{y_1}(\widetilde{y}_1, \widetilde{y}_2, x), \overline{K}_{y_2}(\widetilde{y}_1, \widetilde{y}_2, x),$

$\overline{K}_x(\widetilde{y}_1, \widetilde{y}_2, x))'$. Define the vectors $\omega^1 = (\omega_1^1, \omega_2^1, \omega_3^1)$ and $\omega^2 = (\omega_1^2, \omega_2^2, \omega_3^2)$ by

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

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

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

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

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

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

Define

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

and define an estimator for $\overline{V}$ by

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

which is obtained by substituting f by $\widehat{f}$, x^* by $\widehat{x}^*$, and x^{**} by $\widehat{x}^{**}$ in the definitions of ω_1, ω_2, and ω_3. In the Appendix, we prove the following.

THEOREM 4.4: *Suppose that Assumptions* 3.1–3.3, 2.4, 4.5–4.6, *and* 4.7′–4.11′ *are satisfied. Define* $\partial \widehat{m^1(y_1, \varepsilon_1)}/\partial y_2$ *as in* (4.11), *for* $\widehat{x}^*$ *and* $\widehat{x}^{**}$ *satisfying* $\partial \log \widehat{f}_{Y|X=\widehat{x}^*}(y)/\partial x = \partial \log \widehat{f}_{Y|X=\widehat{x}^{**}}(y)/\partial x = 0$. *Then*,

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

and $\widehat{\overline{V}}$ *is a consistent estimator for* $\overline{V}$.

5. SIMULATIONS

In this section, we report our best results of limited experiments with the estimators developed in the previous sections. The estimators involve choosing many parameters, such as the bandwidths, the kernel functions, and the weight functions in Sections 2 and 3. In addition, trimming is often desirable to avoid random denominators whose values are close to zero. The results are sensitive to the values of all these parameters; we chose only rudimentary ones. Further studies are needed to determine how to select these optimally.

5.1. *Simulations for the Average Derivative Estimator for the Exclusive Regressors Model*

To obtain some indication of the performance of the estimator for the exclusive regressors model in Section 2, we generated simulated data from a linear and from a nonlinear model. The linear model was specified as

$$y_1 = 0.75y_2 - x_1 + \varepsilon_1,$$
$$y_2 = -0.5y_1 - x_2 + \varepsilon_2.$$

The inverse functions r^1 and r^2 were then

$$\varepsilon_1 = r^1(y_1, y_2, x_1) = y_1 - 0.75y_2 + x_1,$$
$$\varepsilon_2 = r^2(y_1, y_2, x_2) = 0.5y_1 + y_2 + x_2.$$

For each of 500 simulations, we obtained samples for (X_1^i, X_2^i) and for $(\varepsilon_1^i, \varepsilon_2^i)$ from a Normal distribution with mean $(0, 0)$ and with variance matrix equal to the identity matrix. We used the estimator developed in Section 2 to estimate the derivatives of the functions r^1 and r^2 with respect to y_1 and y_2 at $(y_1, y_2) = (0, 0)$. The derivatives of the inverse functions at $(y_1, y_2) = (0, 0)$ are, for all values of (x_1, x_2),

$$r_{y_1}^1 = 1; \quad r_{y_2}^1 = -0.75; \quad r_{y_1}^2 = 0.5; \quad r_{y_2}^2 = 1.$$

The weight function $\mu(y,\cdot)$ employed was the Normal density with mean $(0,0)$, diagonal variance matrix, and standard deviation for each coordinate equal to $1/3$ of the standard deviation of, respectively, x_1 and x_2. We used a Gaussian kernel of order 8, and bandwidths

$$h_k = \sigma_k N^{-1/(L+2order+2j)},$$

where h_k denotes the bandwidth for coordinate k, σ_k denotes standard deviation of the sample, *order* denotes the order of the kernel, L denotes the number of coordinates of the density; $j=0$ if the bandwidth is used in the estimation of a density, and $j=1$ if the bandwidth is used in the estimation of a derivative of a density.

The results are presented in Table I. For the part of Table I presenting the results for $\widehat{r}^1_{y_1}$, the column $\overline{\widehat{r}^1_{y_1}}$ denotes the average over simulations of the value of $\widehat{r}^1_{y_1}$, $\overline{\widehat{std}}$ denotes the average over simulations of the estimated standard deviation of $\widehat{r}^1_{y_1}$, *std* denotes the empirical standard deviation of $\widehat{r}^1_{y_1}$. The 0.95 column denotes the percentage of the simulations where the true value of $r^1_{y_1}$ was within the estimated 95% confidence interval. The other parts of Table I report the analogous results for the estimators of the other derivatives. In general, the average of the estimated standard errors, calculated with the equation for the variance stated in Theorem 2.4, was close to the empirical standard deviation of the estimates. As expected, both the bias and the estimated standard errors decreased with sample size.

TABLE I

TWO EQUATIONS–TWO INSTRUMENTS; LINEAR MODEL
AVERAGE OVER INSTRUMENTS ESTIMATOR

	$r^1_{y_1}=1$				$r^1_{y_2}=-0.75$			
N	$\overline{\widehat{r}^1_{y_1}}$	*std*	$\overline{\widehat{std}}$	0.95	$\overline{\widehat{r}^1_{y_1}}$	*std*	$\overline{\widehat{std}}$	0.95
500	0.54	0.52	0.78	0.96	−0.40	0.57	0.87	0.98
1,000	0.72	0.43	0.61	0.97	−0.56	0.49	0.68	0.98
2,000	0.90	0.39	0.47	0.98	−0.68	0.42	0.52	0.99
5,000	1.01	0.30	0.30	0.95	−0.75	0.30	0.34	0.98
10,000	1.03	0.25	0.22	0.93	−0.77	0.25	0.25	0.97
	$r^2_{y_1}=0.5$				$r^2_{y_2}=1$			
N	$\overline{\widehat{r}^1_{y_2}}$	*std*	$\overline{\widehat{std}}$	0.95	$\overline{\widehat{r}^1_{y_2}}$	*std*	$\overline{\widehat{std}}$	0.95
500	0.26	0.47	0.80	1.0	0.50	0.63	0.89	0.96
1,000	0.36	0.40	0.61	1.0	0.76	0.54	0.68	0.97
2,000	0.46	0.33	0.46	1.0	0.93	0.44	0.51	0.97
5,000	0.50	0.26	0.30	0.97	1.01	0.31	0.33	0.98
10,000	0.52	0.22	0.22	0.97	1.03	0.26	0.25	0.95

The nonlinear model was specified by

$$y_1 = 10\big[1 + \exp\big(2(y_2 - x_1 - 5 + \varepsilon_1)\big)\big]^{-1},$$
$$y_2 = 4.5 + 0.1y_1 - x_2 + \varepsilon_2.$$

The corresponding inverse functions are

$$\varepsilon_1 = r^1(y_1, y_2, x_1) = \frac{1}{2}\log\left(\frac{10}{y_1} - 1\right) - y_2 + x_1 + 5,$$
$$\varepsilon_2 = r^2(y_1, y_2, x_1) = -0.1y_1 + y_2 - 4.5 + x_2.$$

The derivatives of (r^1, r^2) were estimated at $(y_1, y_2) = (5, 5)$. At this point,

$$r^1_{y_1} = -0.2; \quad r^1_{y_2} = -1.0; \quad r^2_{y_1} = -0.1; \quad r^2_{y_2} = 1.0.$$

As with the linear model, for each of 500 simulations, we obtained samples for (X^i_1, X^i_2) and for $(\varepsilon^i_1, \varepsilon^i_2)$ from a Normal distribution with mean $(0, 0)$ and with variance matrix equal to the identity matrix. The results are presented in Table II. The average values of the estimated standard errors were again in this model close to the standard deviation of the estimates. It is worth noticing the difference between our nonparametric estimator for $-r^1_{y_2}/r^1_{y_1}$ and the corresponding least squares estimator, $-(\widetilde{r}^1_{y_2})_{\text{LS}} = -2.07$, calculated assuming

TABLE II

TWO EQUATIONS–TWO INSTRUMENTS; NONLINEAR MODEL
AVERAGE OVER INSTRUMENTS ESTIMATOR

	$r^1_{y_1} = -0.2$				$r^1_{y_2} = -1$			
N	$\overline{\widehat{r}^1_{y_1}}$	*std*	$\overline{\widehat{std}}$	0.95	$\overline{\widehat{r}^1_{y_1}}$	*std*	$\overline{\widehat{std}}$	0.95
500	−0.12	0.15	0.25	0.98	−0.37	0.58	0.75	0.87
1,000	−0.14	0.15	0.21	0.98	−0.51	0.50	0.65	0.92
2,000	−0.16	0.12	0.17	0.98	−0.70	0.47	0.51	0.92
5,000	−0.17	0.09	0.11	0.98	−0.90	0.33	0.34	0.92
10,000	−0.16	0.07	0.08	0.94	−0.98	0.26	0.25	0.94
	$r^2_{y_1} = -0.1$				$r^2_{y_2} = 1$			
N	$\overline{\widehat{r}^1_{y_2}}$	*std*	$\overline{\widehat{std}}$	0.95	$\overline{\widehat{r}^1_{y_2}}$	*std*	$\overline{\widehat{std}}$	0.95
500	−0.06	0.16	0.27	1	0.29	0.54	0.82	0.90
1,000	−0.08	0.15	0.24	1	0.41	0.51	0.72	0.92
2,000	−0.08	0.12	0.19	1	0.62	0.49	0.58	0.93
5,000	−0.11	0.09	0.13	0.99	0.85	0.38	0.40	0.92
10,000	−0.11	0.08	0.10	0.99	0.95	0.31	0.30	0.93

that the model is linear. The value -2.07 corresponds to the 0.987 quantile of the empirical distribution of $-\widehat{r}^1_{y_2}/\widehat{r}^1_{y_1}$ when $N = 10{,}000$.

5.2. *Simulations for the Average Derivative Estimator for the Two Equations, One Instrument Model*

We considered the same linear and nonlinear models as in the previous subsection, except that here the first equation has no exogenous regressor. So, the linear and nonlinear models were, respectively,

$$y_1 = 0.75y_2 + \varepsilon_1,$$
$$y_2 = -0.5y_1 - x_2 + \varepsilon_2,$$

and

$$y_1 = 10\big[1 + \exp\big(2(y_2 - 5 + \varepsilon_1)\big)\big]^{-1},$$
$$y_2 = 4.5 + 0.1y_1 - x_2 + \varepsilon_2.$$

For both the linear and nonlinear models, the distribution of $(\varepsilon_1, \varepsilon_2)$ was such that the marginal density of ε_1 was Normal with mean 0 and standard deviation 0.5, and conditional on ε_1, the density of ε_2 was a mixture of two normal distributions. The first of these normal distributions had mean $\mu_1 = -(1 + 0.1\varepsilon_1)^2$ and standard deviation $\sigma_1 = 0.1(1 + 0.5\varepsilon_1)^2$, while the second had mean $\mu_2 = (1 + 0.2\varepsilon_1)^2$ and standard deviation $\sigma_2 = 0.1(1 + 2\varepsilon_1)^2$. The variance of ε_2 generated this way was 1.09. X was sampled from a Uniform$(-1.5, 1.5)$.

The derivative with respect to y_2 in the first structural equation was estimated for 500 simulations. The derivative was estimated at $y_1 = y_2 = 0$ for the linear model and at $y_1 = y_2 = 5$ for the nonlinear model. The nonparametric densities and derivatives were estimated using again a Gaussian kernel of order 8 and for each coordinate, bandwidths $h_k = \sigma_k N^{-1/(L+2order+2j)}$. For both models and all samples, the weight to any value of x was assigned the value of zero when at (y_1, y_2, x), the estimated value of either the joint density, or the marginal density of X, or the conditional density was below 0.001. Lowering this trimming value increased coverage. The weight function $\mu(y, \cdot)$ was otherwise uniform. The results are presented in Table III.

For the data generated by the nonlinear model, the derivative with respect to y_2 has true value -5, while the least squares estimate for this derivative is -2.75. This value corresponds to the 0.93 quantile of the empirical distribution of the estimator when $N = 10{,}000$, to the 0.992 quantile when $N = 25{,}000$ and it is above the 1.0 quantile when $N = 50{,}000$.

5.3. *Simulations for an Indirect Estimator*

We performed simulations for the estimator that is based on first order derivatives. For each of 500 simulations, we generated ε_1^i from a uniform dis-

TABLE III

TWO EQUATIONS–ONE INSTRUMENT; AVERAGE OVER INSTRUMENT ESTIMATOR

	Linear Model $\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2}=0.75$				Nonlinear Model $\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2}=-5$			
N	$\overline{\widehat{m}_{y_1}^1}$	*std*	$\widehat{\overline{std}}$	0.95	$\overline{\widehat{m}_{y_1}^1}$	*std*	$\widehat{\overline{std}}$	0.95
10,000	0.76	0.30	0.32	0.93	−5.05	1.89	2.17	0.94
25,000	0.74	0.17	0.21	0.98	−4.95	1.06	1.48	0.96
50,000	0.73	0.12	0.16	0.98	−4.90	0.77	1.09	0.98

tribution with support $(-1,1)$, and for each ε_1^i, we generated ε_2^i from a conditional density satisfying

$$\log\big(f_{\varepsilon_2|\varepsilon_1}(\varepsilon_2)\big)=(1/4)(\varepsilon_2)^2-(1/2)(\varepsilon_2)^4+\varepsilon_1+\ \varepsilon_1\varepsilon_2+\varepsilon_1(\varepsilon_2)^2.$$

When $\varepsilon_1=0$, the derivative of $\log(f_{\varepsilon_2|\varepsilon_1}(\varepsilon_2))$ with respect to ε_2 equals zero at $\varepsilon_2=-0.5$, $\varepsilon_2=0$, and $\varepsilon_2=0.5$. X was generated from a Uniform$(-1,1)$ distribution. The object of interest, $\partial m^1(y_2,\varepsilon_1)/\partial y_2$, was estimated at $y_1=y_2=0$ for the linear model, and at $y_1=y_2=5$ for the nonlinear model. The results reported in Table IV were obtained with kernels of order 4 and bandwidths of size 3/4 of those determined as in the previous two subsections. Larger bandwidths generated a larger bias for both the linear and the nonlinear estimators. A higher order kernel made it difficult finding values of X at which $\partial\log f_{Y|X=x}(y)/\partial x=0$. Even with a kernel of order 4, for some simulations, such values were not always found. Moreover, even when such values were found, it was also possible that the estimator had values of different sign than the correct one, or that the random denominators in the calculation of the estimates or its variance were too low. The numbers in Table IV are calculated using only the simulations where such situations did not arise. Out of 500 simulations, $\widetilde{N}$ denotes the number of simulations so obtained. The lower

TABLE IV

TWO EQUATIONS–ONE INSTRUMENT; INDIRECT ESTIMATOR

	Linear Model $\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2}=0.75$					Nonlinear Model $\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2}=-5$				
N	$\overline{\widehat{m}_{y_2}^1}$	*std*	$\widehat{\overline{std}}$	0.95	$\widetilde{N}$	$\overline{\widehat{m}_{y_2}^1}$	*std*	$\widehat{\overline{std}}$	0.95	$\widetilde{N}$
5,000	0.84	0.38	0.61	0.99	452	−5.51	2.27	2.78	0.98	445
10,000	0.76	0.27	0.35	0.98	471	−5.32	1.80	2.12	0.98	474
25,000	0.74	0.22	0.27	0.97	488	−5.00	1.18	1.33	0.97	492

bound allowed on the denominators was 10^{-6}. For the simulations where more than two values x of X at which $\partial \log f_{Y|X=x}(y)/\partial x = 0$ were found, the estimate was calculated using the smallest and largest values of such x values. For the data generated from the nonlinear model, the least squares estimate for $\partial m^1(y_2, \varepsilon_1)/\partial y_2$ is -3.1, which corresponds to the 0.93 quantile of the empirical distribution of the indirect estimator when $N = 25{,}000$.

6. CONCLUSIONS

In this paper, we have developed two new approaches for estimation of nonparametric simultaneous equations models. Both approaches were based on new identification results that were developed in this paper for two models, the exclusive regressors model and the two equations, one instrument model. In the exclusive regressors model, to each structural equation there corresponds an exclusive observable exogenous variable. In the two equations, one instrument model, the object of interest is the equation where the instrument is excluded. For each model, we developed estimators that are calculated by a simple matrix inversion and a matrix multiplication, where the elements in each matrix are calculated by averages of products of nonparametrically estimated derivatives of the conditional density of the observable variables. For the two equations, one instrument model, we developed also two-step indirect estimators. In the first step, values of the observable exogenous variable are found where the nonparametric estimator of the conditional density of the observable variables satisfies some conditions. In the second step, the value of the object of interest is read off the estimated conditional density evaluated at those values. We developed two such two-step indirect estimators, one using first order derivatives and a second one using second order derivatives. We have shown that all the new estimators are asymptotically normal. Although we only developed two estimation approaches, the new identification results can be used to develop many other approaches.

APPENDIX

PROOF OF THEOREM 2.1: We first introduce some notation and a lemma. Let r_y denote the $G \times G$ matrix whose element in the ith row and jth column is $r^i_{y_j}(y, x_i)$. Let r_x denote the $G \times G$ diagonal matrix whose element in the ith diagonal place is $r^i_{x_i}(y, x_i)$. The $G \times G$ matrix $\overline{r}_y$ whose element in the ith row and jth column is the ratio $r^i_{y_j}(y, x_i)/r^i_{x_i}(y, x_i)$ will then be equal to $(r_x)^{-1} r_y$. We will denote by q_ε the $G \times 1$ vector whose ith element is $\partial \log f_\varepsilon(r(y, x)/\partial \varepsilon_i$. The $G \times 1$ vectors γ_y and γ_x will be the vectors whose ith elements are, respectively, $\partial \log |r_y(y, x)|/\partial y_i$ and $\partial \log |r_y(y, x)|/\partial x_i$. For any function $\widetilde{r}$ and any density $f_{\widetilde{\varepsilon}}$, the matrices $\widetilde{r}_y$, $\widetilde{r}_x$, and $\overline{\widetilde{r}}_y$ and the vectors $\widetilde{q}_\varepsilon$, $\widetilde{\gamma}_y$, and $\widetilde{\gamma}_x$ will be defined analogously. We will denote the $G \times 1$ vectors $(\mathbf{g}_{y_1}, \ldots, \mathbf{g}_{y_G})'$ and $(\mathbf{g}_{x_1}, \ldots, \mathbf{g}_{x_G})'$ by, respectively, $\mathbf{g}_y$ and $\mathbf{g}_x$. We will make use of the following lemma.

LEMMA M—Matzkin (2008): *Let M denote a convex set in the interior of the support of (Y, X). Suppose that $(r, f_\varepsilon) \in (\Gamma \times \Phi)$ generate $f_{Y|X}$ on M and $\widetilde{r} \in \Gamma$. Then, there exists $f_{\widetilde{\varepsilon}} \in \Phi$ such that $(\widetilde{r}, f_{\widetilde{\varepsilon}})$ generate $f_{Y|X}$ on M if and only if, for all $(y, x) \in M$,*

$$\text{(L.1)} \qquad \left(r'_x - \widetilde{r}'_x\left(\widetilde{r}'_y\right)^{-1} r'_y\right) q_\varepsilon = -(\gamma_x - \widetilde{\gamma}_x) + \widetilde{r}'_x\left(\widetilde{r}'_y\right)^{-1}(\gamma_y - \widetilde{\gamma}_y).$$

PROOF: First, restrict the definition of observational equivalence in Matzkin (2008) to observational equivalence on M. Then, the lemma follows by the same arguments used in Matzkin (2008) to prove her Theorems 3.1 and 3.2. The expression in (L.1) is the transpose of (2.8) in the statement of Theorem 3.1 in Matzkin (2008). *Q.E.D.*

We will next show that in the exclusive regressors model, when f_ε and r generate $f_{Y|X}$, and when r_x and $\widetilde{r}_x$ are invertible, diagonal, $G \times G$ matrices, on M, (2.7) is equivalent to (L.1). Hence, by Lemma M, it will follow that $\widetilde{r}$ is observationally equivalent to r on M if and only if (2.7) is satisfied. For this, we first note that since (r, f_ε) generate $f_{Y|X}$ on M, for all $(y, x) \in M$,

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

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

$$\mathbf{g}_x = r'_x q_\varepsilon + \gamma_x.$$

Since by assumption r_x is invertible, we can solve uniquely for q_ε, getting

$$\text{(T.2.1)} \qquad \left(r'_x\right)^{-1}(\mathbf{g}_x - \gamma_x) = q_\varepsilon.$$

Replacing q_ε in (L.1) by the expression for q_ε in (T.2.1), we get

$$\text{(T.2.2)} \qquad \left[r'_x - \widetilde{r}'_x\left(\widetilde{r}'_y\right)^{-1} r'_y\right]\left(r'_x\right)^{-1}(\mathbf{g}_x - \gamma_x) = -(\gamma_x - \widetilde{\gamma}_x) + \left[\widetilde{r}'_x\left(\widetilde{r}'_y\right)^{-1}\right](\gamma_y - \widetilde{\gamma}_y).$$

Since $r'_x(r'_x)^{-1} = I$, the left-hand side of (T.2.2) can be expressed as

$$\begin{aligned}
&\left[r'_x - \widetilde{r}'_x\left(\widetilde{r}'_y\right)^{-1} r'_y\right]\left(r'_x\right)^{-1}(\mathbf{g}_x - \gamma_x) \\
&\quad = \left[I - \widetilde{r}'_x\left(\widetilde{r}'_y\right)^{-1} r'_y\left(r'_x\right)^{-1}\right]\mathbf{g}_x - \left[I - \widetilde{r}'_x\left(\widetilde{r}'_y\right)^{-1} r'_y\left(r'_x\right)^{-1}\right]\gamma_x.
\end{aligned}$$

Hence, premultiplying both sides of (T.2.2) by $-\widetilde{r}_y'(\widetilde{r}_x')^{-1}$, we get

$$\begin{aligned}&\left[r_y'(r_x')^{-1}-\widetilde{r}_y'(\widetilde{r}_x')^{-1}\right]\mathbf{g}_x-\left[r_y'(r_x')^{-1}-\widetilde{r}_y'(\widetilde{r}_x')^{-1}\right]\gamma_x\\&\quad=\left[\widetilde{r}_y'(\widetilde{r}_x')^{-1}\right](\gamma_x-\widetilde{\gamma}_x)-(\gamma_y-\widetilde{\gamma}_y).\end{aligned}$$

Subtracting $[\widetilde{r}_y'(\widetilde{r}_x')^{-1}]\gamma_x$ from both sides of the equality sign, and rearranging terms, the last expression can be written as

$$\left[r_y'(r_x')^{-1}-\widetilde{r}_y'(\widetilde{r}_x')^{-1}\right]\mathbf{g}_x+\left[\gamma_y-r_y'(r_x')^{-1}\gamma_x\right]-\left[\widetilde{\gamma}_y-\widetilde{r}_y'(\widetilde{r}_x')^{-1}\widetilde{\gamma}_x\right]=0.$$

Note that the vector $d_y=(d_{y_1},\dots,d_{y_G})'=\gamma_y-r_y'(r_x')^{-1}\gamma_x$ and $\widetilde{d}_y=(\widetilde{d}_{y_1},\dots,\widetilde{d}_{y_G})'=\widetilde{\gamma}_y-\widetilde{r}_y'(\widetilde{r}_x')^{-1}\widetilde{\gamma}_x$. Hence, we have obtained that in the exclusive regressors model, (L.1) is equivalent to

$$\left[r_y'(r_x')^{-1}-\widetilde{r}_y'(\widetilde{r}_x')^{-1}\right]\mathbf{g}_x+d_y-\widetilde{d}_y=0.$$

This is exactly (2.7) in Section 2. *Q.E.D.*

PROOF OF PROPOSITIONS 2.1 AND 2.2: By the definition of the model in (2.2), for each g, the ratios of derivatives, $r^g_{y_j}(y,x_g)/r^g_{x_g}(y,x_g)$ $(j=1,\dots,G)$, depend only on y and x_g. When Assumption 2.4 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, $\bar{r}$ is constant over $\{(y,x_{-G},t_G)|t_G\in R\}$. 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 (2.5), are constant over x_G. It then follows that $(d_{y_1},\dots,d_{y_G})$ is constant over $\{(y,x_{-G},t_G)|t_G\in R\}$. A similar reasoning shows that when Assumption 2.4′ is satisfied, $(\bar{r},d)$ is constant over $\{(y,t_1,\dots,t_G)|(t_1,\dots,t_G)\in R^G\}$. *Q.E.D.*

PROOF OF PROPOSITION 2.3: Let $x_G^{(1)},\dots,x_G^{(G+1)}$ be such that $w^{(1)}=(y,x_{-G},x_G^{(k)}),\dots,w^{(G+1)}=(y,x_{-G},x_G^{(G+1)})\in\overline{M}$ and let $\varepsilon^{(1)},\dots,\varepsilon^{(G+1)}$ be such that, for each $k=1,\dots,G+1$,

$$\varepsilon^{(k)}=\left(r^1(y,x_1),\dots,r^{G-1}(y,x_{G-1}),r^G\left(y,x_G^{(k)}\right)\right).$$

We will show that

$$\begin{aligned}&A\left(\varepsilon^{(1)},\dots,\varepsilon^{(G+1)}\right)\text{ is invertible}\\&\quad\Longleftrightarrow\quad B\left(w^{(1)},\dots,w^{(G+1)}\right)\text{ is invertible.}\end{aligned}$$

Note that the relationship between the element in the jth row and kth column of $B(w^{(1)}, \ldots, w^{(G+1)})$ and the jth row and kth column of $A = A(\varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)})$ for $k \in \{1, \ldots, G\}$ is given by

$$\mathbf{g}_{x_k}\big(y, x_{-G}, x_G^{(j)}\big) = \frac{\partial \log f_\varepsilon(\varepsilon_{-G}, r^G(y, x_{-G}, x_G^{(j)}))}{\partial \varepsilon_k} r^k_{x_k}(y, x_k) + \frac{\partial \log |r_y(y, x)|}{\partial x_k},$$

where $\varepsilon_{-G} = (r^1(y, x_1), \ldots, r^{G-1}(y, x_{G-1}))$. Suppose that $A = A(\varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)})$ is invertible. The value of $\partial \log |r_y(y, x)|/\partial x_k$ is constant across the kth 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 subtracting it from the kth column will result in a matrix of the same rank as B. Since $r^k_{x_k}(y, x_k)$ is also constant across the kth column, dividing the resulting column k by $r^k_{x_k}(y, x_k)$ will not affect the rank as well. Repeating the analogous operations in each of the first G columns of B results in the matrix $A = A(\varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)})$. 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. *Q.E.D.*

PROOF OF PROPOSITION 2.4: By Assumption 2.4′, for each $k = 1, \ldots, G$, the derivative of r^k with respect to x_k is 1 and the derivative of the Jacobian determinant $|r_y(y, x)|$ with respect to x_k is zero. Then, for each $k = 1, \ldots, G$ and all x,

$$\begin{aligned}\mathbf{g}_{x_k}(y, x) &= \frac{\partial \log f_\varepsilon(r(y, x))}{\partial \varepsilon_k} r^k_{x_k}(y, x_k) + \frac{|r_y(y, x)|_{x_k}}{|r_y(y, x)|} \\ &= \frac{\partial \log f_\varepsilon(r(y, x))}{\partial \varepsilon_k}.\end{aligned}$$

Let $w^{(j)} = (y, x^{(j)}) \in \overline{M}$ ($j = 1, \ldots, G + 1$) and $\varepsilon^{(j)} = (r^1(y, x_1^{(j)}), \ldots, r^{G-1}(y, x_{G-1}^{(j)}), r^G(y, x_G^{(j)}))$. Since $\mathbf{g}_{x_k}(y, x^{(j)}) = \partial \log f_\varepsilon(r(y, x^{(j)}))/\partial \varepsilon_k$, $B(w^{(1)}, \ldots, w^{(G+1)}) = A(\varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)})$. Hence, $A(\varepsilon^{(1)}, \ldots, \varepsilon^{(G+1)})$ is invertible if and only if $B(w^{(1)}, \ldots, w^{(G+1)})$ is invertible. *Q.E.D.*

PROOF OF THEOREM 2.2: Let $t_G^{(1)}, \ldots, t_G^{(G+1)}$ denote the values of t_G corresponding to the points $w^{(1)}, \ldots, w^{(G+1)}$ satisfying Condition I.1. Since μ is strictly positive and continuous at $(y, x_{-G}, t_G^{(1)}), \ldots, (y, x_{-G}, t_G^{(G+1)})$, there exist $\delta > 0$ neighborhoods $B((y, x_{-G}, t_G^{(k)}), \delta)$ ($k = 1, \ldots, G + 1$), such that μ is strictly positive on those neighborhoods. Let U denote the union of those neighborhoods. For any j,

$$\int_{\overline{M}} \left(\mathbf{g}_{y_j} - \widetilde{\overline{r}}^1_{y_j}\mathbf{g}_{x_1} - \widetilde{\overline{r}}^2_{y_j}\mathbf{g}_{x_2} - \cdots - \widetilde{\overline{r}}^G_{y_j}\mathbf{g}_{x_G} - \widetilde{d}_{y_j}\right)^2 \mu(y, x_{-G}, t)\, dt$$
$$= \int_{\overline{M}\cap U} \left(\mathbf{g}_{y_j} - \widetilde{\overline{r}}^1_{y_j}\mathbf{g}_{x_1} - \widetilde{\overline{r}}^2_{y_j}\mathbf{g}_{x_2} - \cdots - \widetilde{\overline{r}}^G_{y_j}\mathbf{g}_{x_G} - \widetilde{d}_{y_j}\right)^2 \mu(y, x_{-G}, t)\, dt$$
$$+ \int_{\overline{M}\cap U^c} \left(\mathbf{g}_{y_j} - \widetilde{\overline{r}}^1_{y_j}\mathbf{g}_{x_1} - \widetilde{\overline{r}}^2_{y_j}\mathbf{g}_{x_2} - \cdots - \widetilde{\overline{r}}^G_{y_j}\mathbf{g}_{x_G} - \widetilde{d}_{y_j}\right)^2$$
$$\times \mu(y, x_{-G}, t)\, dt.$$

When $(\widetilde{\overline{r}}_{y_j}, \widetilde{d}_{y_j}) = (\widetilde{\overline{r}}^1_{y_j}, \widetilde{\overline{r}}^2_{y_j}, \ldots, \widetilde{\overline{r}}^G_{y_j}, \widetilde{d}_{y_j}) = (\overline{r}^1_{y_j}, \overline{r}^2_{y_j}, \ldots, \overline{r}^G_{y_j}, d_{y_j})$, both terms are zero. When $(\widetilde{\overline{r}}_{y_j}, \widetilde{d}_{y_j}) \neq (\overline{r}^1_{y_j}, \overline{r}^2_{y_j}, \ldots, \overline{r}^G_{y_j}, d_{y_j})$, Condition I.1 implies that for at least one $w^{(s)}$, $(\mathbf{g}_{y_j} - \widetilde{\overline{r}}^1_{y_j}\mathbf{g}_{x_1} - \widetilde{\overline{r}}^2_{y_j}\mathbf{g}_{x_2} - \cdots - \widetilde{\overline{r}}^G_{y_j}\mathbf{g}_{x_G} - \widetilde{d}_{y_j})^2 \neq 0$. The continuity of $\mathbf{g}_{y_j}$ and $\mathbf{g}_{x_j}$ on t, which is implied by Assumptions 2.1–2.3, implies then that the integral over U is strictly positive. Hence, $S(\widetilde{\overline{r}}, \widetilde{d})$ is uniquely minimized at $(\widetilde{\overline{r}}_{y_j}, \widetilde{d}_{y_j}) = (\overline{r}^1_{y_j}, \overline{r}^2_{y_j}, \ldots, \overline{r}^G_{y_j}, d_{y_j})$. Since S is a convex and differentiable function, its matrix of second order derivatives must be positive semidefinite at the minimizer $(\overline{r}, d)$. Denote $\int_{\overline{M}} \mathbf{g}_{x_j}\mathbf{g}_{x_s}\mu(y, x_{-G}, t)\, dt$ by $\int \mathbf{g}_{x_j}\mathbf{g}_{x_s}$ and denote $\int_{\overline{M}} \mathbf{g}_{x_j}\mu(y, x_{-G}, t)\, dt$ by $\int \mathbf{g}_{x_j}$. The first order conditions for any subvector $(\overline{r}^1_{y_j}, \overline{r}^2_{y_j}, \ldots, \overline{r}^G_{y_j}, d_{y_j})$ are

$$\begin{bmatrix} \int \mathbf{g}_{x_1}\mathbf{g}_{x_1} & \int \mathbf{g}_{x_1}\mathbf{g}_{x_2} & \cdots & \int \mathbf{g}_{x_1}\mathbf{g}_{x_G} & \int \mathbf{g}_{x_1} \\ \vdots & \vdots & \vdots & \vdots & \int \mathbf{g}_{x_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \int \mathbf{g}_{x_G}\mathbf{g}_{x_1} & \int \mathbf{g}_{x_G}\mathbf{g}_{x_2} & \cdots & \int \mathbf{g}_{x_G}\mathbf{g}_{x_G} & \int \mathbf{g}_{x_G} \\ \int \mathbf{g}_{x_1} & \int \mathbf{g}_{x_2} & \cdots & \int \mathbf{g}_{x_G} & 1 \end{bmatrix} \begin{bmatrix} \overline{r}^1_{y_j} \\ \overline{r}^2_{y_j} \\ \vdots \\ \overline{r}^G_{y_j} \\ d_{y_j} \end{bmatrix}$$
$$= \begin{bmatrix} \int \mathbf{g}_{y_j}\mathbf{g}_{x_1} \\ \vdots \\ \int \mathbf{g}_{y_j}\mathbf{g}_{x_G} \\ \int \mathbf{g}_{y_j} \end{bmatrix},$$

and the $(G+1) \times (G+1)$ matrix in this expression is the matrix of second order derivatives of S with respect to $(\widetilde{\overline{r}}_g, \widetilde{d}_g)$, which is constant over $(\widetilde{\overline{r}}_{y_j}, \widetilde{d}_{y_j})$.

If this matrix were not invertible, the solution would not be unique. Hence, since the matrix is positive semidefinite and invertible, it is positive definite. Solving for d_{y_j} using the last equation and substituting the resulting expression into the other equations, we get

$$\begin{bmatrix} T_{x_1x_1} & T_{x_1x_2} & \cdots & T_{x_1x_G} \\ \vdots & \vdots & \vdots & \vdots \\ T_{x_Gx_1} & T_{x_Gx_2} & \cdots & T_{x_Gx_G} \end{bmatrix} \begin{bmatrix} \overline{r}^1_{y_j} \\ \overline{r}^2_{y_j} \\ \vdots \\ \overline{r}^G_{y_j} \end{bmatrix} = \begin{bmatrix} T_{y_jx_1} \\ \vdots \\ T_{y_jx_G} \end{bmatrix},$$

where the $G \times G$ matrix is also positive definite, by properties of partitioned matrices, and therefore invertible. *Q.E.D.*

PROOF OF THEOREM 2.3: The arguments are almost identical to those in the proof of Theorem 2.2. *Q.E.D.*

PROOF OF THEOREM 2.4: Let $\mathtt{F}$ denote the set of bounded and continuously differentiable functions g on the extension of the support of (Y, X) to the whole space and such that the function $\widetilde{g}$ defined by $\widetilde{g}(x) = \int g(y, x)\, dy$ is bounded and continuously differentiable. For all functions g in $\mathtt{F}$, denote by $\|g\|$ the sum of the sup norms of the values and derivatives of g and of $\widetilde{g}$ on $\overline{M}^y \times \overline{M}^x$. By Assumption 2.7, the density $f_{Y,X}$ belongs to $\mathtt{F}$ and it is such that, for $\delta > 0$ and all $(y, x) \in \overline{M}^y \times \overline{M}^x$, $f_{Y,X}(y, x) > \delta$ and $f_X(x) > \delta$. Define the functionals $\alpha_{y_j}(\cdot)$ and $\beta_{x_s}(\cdot)$ on F by $\alpha_{y_j}(g) = \partial \log g_{Y|X=x}(y)/\partial y_j$ and $\beta_{x_s}(g) = \partial \log g_{Y|X=x}(y)/\partial x_s$. For simplicity we leave the argument (y, x) implicit. For each j, s, define the functional $\Phi_{y_j,x_s}(g)$ by

$$\begin{aligned} \Phi_{y_j,x_s}(g) = {} & \int_{\overline{M}} \alpha_{y_j}(g)\beta_{x_s}(g)\mu(y, x)\, dx \\ & - \left(\int_{\overline{M}} \alpha_{y_j}(g)\mu(y, x)\, dx\right)\left(\int_{\overline{M}} \beta_{x_s}(g)\mu(y, x)\, dx\right). \end{aligned}$$

Then, $\widehat{T}_{y_jx_s} = \Phi_{y_jx_s}(\widehat{f})$ and $T_{y_jx_s} = \Phi_{y_jx_s}(f)$. By Lemma A.2 in the Supplemental Material (Matzkin (2015)) and Assumptions 2.7 and 2.9, there exist finite $b_1, \widetilde{\delta}_1 > 0$, a linear functional $D\Phi_{y_jx_s}$, and a functional $R\Phi_{y_jx_s}$ such that, when $\|\widehat{f} - f\| \le \widetilde{\delta}_1$,

$$\begin{aligned} & \Phi_{y_jx_s}(\widehat{f}) - \Phi_{y_jx_s}(f) = D\Phi_{y_jx_s}(f; \widehat{f} - f) + R\Phi_{y_jx_s}(f; \widehat{f} - f), \\ & \left|D\Phi_{y_jx_s}(f; \widehat{f} - f)\right| \le b_1\|\widehat{f} - f\|, \quad \text{and} \\ & \left|R\Phi_{y_jx_s}(f; \widehat{f} - f)\right| \le b_1\|\widehat{f} - f\|^2. \end{aligned}$$

By Lemma A.5 in the Supplemental Material and Assumptions 2.7, 2.9, and 2.10, $\sqrt{N\sigma^{G+2}}\|\widehat{f}-f\|^2 = o_p(1)$. Hence,

$$\begin{aligned}
&\sqrt{N\sigma^{G+2}}\big(\Phi_{y_j x_s}(\widehat{f}) - \Phi_{y_j x_s}(f)\big) \\
&\quad = \sqrt{N\sigma^{G+2}} D\Phi_{y_j x_s}(f;\widehat{f}-f) + \sqrt{N\sigma^{G+2}} R\Phi_{y_j x_s}(f;\widehat{f}-f) \\
&\quad = \sqrt{N\sigma^{G+2}} D\Phi_{y_j x_s}(f;\widehat{f}-f) + o_p(1).
\end{aligned}$$

Also by Lemma A.5 in the Supplemental Material,

$$\begin{aligned}
&\sqrt{N\sigma^{G+2}} D\Phi_{y_j x_s}(f;\widehat{f}-f) \\
&\quad = \sqrt{N\sigma^{G+2}} \int_{\overline{M}} \big(\widehat{f}_{y_j}(y,x) - f_{y_j}(y,x)\big) \\
&\qquad \times \left[\frac{\left(\beta_{x_s}(f) - \int_{\overline{M}} \beta_{x_s}(f)\mu\,dx\right)\mu}{f}\right] dx + o_p(1).
\end{aligned}$$

Let $\omega(x)$ denote the $G^2 \times G$ matrix whose rows correspond in order to $\Phi_{y_1,x_1}, \Phi_{y_1,x_2}, \ldots, \Phi_{y_1,x_G}, \ldots, \Phi_{y_G x_1}, \Phi_{y_G x_2}, \ldots, \Phi_{y_G,x_G}$ and whose columns correspond in order to $y_1, y_2, \ldots, y_G$. Let the entry corresponding to $(\Phi_{y_g x_s}, y_j)$ equal to 0 if $g \neq j$ and equal to $(\beta_{x_s}(f) - \int_{\overline{M}} \beta_{x_s}(f)\mu(x)\,dx)\mu(x)/f$ otherwise. Assumptions 2.7 and 2.8 imply that each entry is bounded and continuous and equal to zero on the complement of $\overline{M}$. Hence, Assumption 5.1 in Newey (1994) is satisfied. Our Assumptions 2.7 and 2.9 imply that Assumptions K, H, and Y in Newey (1994) are also satisfied. Hence, it follows by Assumption 2.10 and Lemma 5.3 in Newey (1994) that

$$\sqrt{N\sigma_N^{G+2}}(\widehat{TT}_{yx} - TT_{yx}) \xrightarrow{d} N(0, V_{T_{YX}}), \tag{T.3}$$

where $V_{T_{YX}}$ is as defined in Section 2. Define

$$\begin{aligned}
\Phi_{x_j,x_s}(g) &= \int_{\overline{M}} \beta_{y_j}(g)\beta_{x_s}(g)\mu(y,x)\,dx \\
&\quad - \left(\int_{\overline{M}} \beta_{y_j}(g)\mu(y,x)\,dx\right)\left(\int_{\overline{M}} \beta_{x_s}(g)\mu(y,x)\,dx\right).
\end{aligned}$$

Then, $\widehat{T}_{x_j x_s} = \Phi_{x_j x_s}(\widehat{f})$ and $T_{x_j x_s} = \Phi_{x_j x_s}(f)$. By Lemma A.3 in the Supplemental Material, for a finite $b_3 > 0$, all j, s, and all $\widehat{f}$ such that $\|\widehat{f} - f\| \leq \widetilde{\delta}_3$,

$$\begin{aligned}
|\widehat{T}_{x_j x_s} - T_{x_j x_s}| &= \big|\Phi_{x_j x_s}(\widehat{f}) - \Phi_{x_j x_s}(f)\big| \\
&\leq \big|D\Phi_{x_j x_s}(f;\widehat{f}-f)\big| + \big|R\Phi_{x_j x_s}(f;\widehat{f}-f)\big| \\
&\leq b_3\|\widehat{f}-f\| + b_3\|\widehat{f}-f\|^2.
\end{aligned}$$

By Assumption 2.10 and Lemma A.5 in the Supplemental Material, $\|\widehat{f} - f\| \to 0$ in probability, and $\sqrt{N\sigma^{G+2}}|\widehat{T}_{x_j x_s} - T_{x_j x_s}| \to 0$ in probability. Hence, $\widehat{TT_{XX}}$ converges $o_p(N^{-1/2}\sigma^{-G/2-1})$ to the matrix TT_{XX}, which is positive definite by Theorem 2.3. This together with (T.3) implies by Slutsky's theorem that

$$\sqrt{N\sigma_N^{G+2}}(\widehat{rr} - rr) \underset{d}{\to} N\big(0, (TT_{XX})^{-1}V_{T_{YX}}(TT_{XX})^{-1}\big).$$

To show that $(\widehat{TT_{XX}})^{-1}\widehat{V}_{T_{YX}}(\widehat{TT_{XX}})^{-1}$ is a consistent estimator for $(TT_{XX})^{-1}V_{T_{YX}}(TT_{XX})^{-1}$, it remains to show that, for all j, s,

$$\begin{aligned}
&\int_{\overline{M}}\left[\big(\Delta\partial_{x_j}\log\widehat{f}_{Y|X=x}(y)\big)\big(\Delta\partial_{x_s}\log\widehat{f}_{Y|X=x}(y)\big)\frac{\mu(y,x)^2}{\widehat{f}(y,x)}\right]dx \\
&\quad\overset{p}{\to}\int_{\overline{M}}\left[\big(\Delta\partial_{x_j}\log f_{Y|X=x}(y)\big)\big(\Delta\partial_{x_s}\log f_{Y|X=x}(y)\big)\frac{\mu(y,x)^2}{f(y,x)}\right]dx.
\end{aligned}$$

By Lemma A.6 in the Supplemental Material, the term inside the first integral converges in probability, uniformly over $x \in \overline{M}$, to the term inside the second integral. Since by Assumption 2.7 this expression is bounded from above and below, and since $\overline{M}$ is compact, it follows that the first integral converges in probability to the second integral. *Q.E.D.*

PROOF OF THEOREM 3.1: We will show that in the model with two equations and one instrument, (3.5) is equivalent to (L.1). As in the proof of Theorem 2.1, we note that since (r, f_ε) generate $f_{Y|X}$ on M, for all $(y, x) \in M$,

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

Specifically, for the two equation, one instrument model,

$$f_{Y_1,Y_2|X=x}(y_1,y_2) = f_{\varepsilon_1,\varepsilon_2}\big(r^1(y_1,y_2), r^2(y_1,y_2,x)\big)\big|r_y(y_1,y_2,x)\big|. \tag{T.3.1}$$

By Lemma B.1 in the Supplemental Material, Assumptions 3.1–3.3 together with (T.3.1) imply that

$$\begin{aligned}
&\big(r_x' - \widetilde{r}_x'(\widetilde{r}_y')^{-1}r_y'\big)q_\varepsilon \\
&\quad= \left[1 - \widetilde{r}_{y_1}^1\left(\frac{\widetilde{r}_x^2}{|\widetilde{r}_y|}\right)\left(\frac{|r_y|}{r_{y_1}^1 r_x^2}\right)\right](\mathbf{g}_x - \gamma_x) \\
&\qquad - \left[\widetilde{r}_{y_1}^1\left(\frac{\widetilde{r}_x^2}{|\widetilde{r}_y|}\right)\left(\frac{r_{y_2}^1}{r_{y_1}^1}\right) - \widetilde{r}_{y_2}^1\left(\frac{\widetilde{r}_x^2}{|\widetilde{r}_y|}\right)\right](\mathbf{g}_{y_1} - \gamma_{y_1}),
\end{aligned}$$

and that

$$
\begin{aligned}
&-(\gamma_x-\widetilde{\gamma}_x)+\widetilde{r}_x'\left(\widetilde{r}_y'\right)^{-1}(\gamma_y-\widetilde{\gamma}_y)\\
&\quad=-(\gamma_x-\widetilde{\gamma}_x)-\widetilde{r}^1_{y_2}\left(\frac{\widetilde{r}^2_x}{|\widetilde{r}_y|}\right)(\gamma_{y_1}-\widetilde{\gamma}_{y_1})+\widetilde{r}^1_{y_1}\left(\frac{\widetilde{r}^2_x}{|\widetilde{r}_y|}\right)(\gamma_{y_2}-\widetilde{\gamma}_{y_2}).
\end{aligned}
$$

Hence, (L.1) can be expressed as

$$
\begin{aligned}
&\left[1-\widetilde{r}^1_{y_1}\left(\frac{\widetilde{r}^2_x}{|\widetilde{r}_y|}\right)\left(\frac{|r_y|}{r^1_{y_1}r^2_x}\right)\right](\mathbf{g}_x-\gamma_x)\\
&\qquad-\left[\widetilde{r}^1_{y_1}\left(\frac{\widetilde{r}^2_x}{|\widetilde{r}_y|}\right)\left(\frac{r^1_{y_2}}{r^1_{y_1}}\right)-\widetilde{r}^1_{y_2}\left(\frac{\widetilde{r}^2_x}{|\widetilde{r}_y|}\right)\right](\mathbf{g}_{y_1}-\gamma_{y_1})\\
&\quad=-(\gamma_x-\widetilde{\gamma}_x)-\widetilde{r}^1_{y_2}\left(\frac{\widetilde{r}^2_x}{|\widetilde{r}_y|}\right)(\gamma_{y_1}-\widetilde{\gamma}_{y_1})+\widetilde{r}^1_{y_1}\left(\frac{\widetilde{r}^2_x}{|\widetilde{r}_y|}\right)(\gamma_{y_2}-\widetilde{\gamma}_{y_2}).
\end{aligned}
\tag{T.3.2}
$$

Multiplying both sides of (T.3.2) by $[-|\widetilde{r}_y|/(\widetilde{r}^2_x\widetilde{r}^1_{y_1})]$, we get

$$
\begin{aligned}
&\left[\left(\frac{|r_y|}{r^1_{y_1}r^2_x}\right)-\left(\frac{|\widetilde{r}_y|}{\widetilde{r}^2_x\widetilde{r}^1_{y_1}}\right)\right](\mathbf{g}_x-\gamma_x)+\left[\left(\frac{r^1_{y_2}}{r^1_{y_1}}\right)-\left(\frac{\widetilde{r}^1_{y_2}}{\widetilde{r}^1_{y_1}}\right)\right](\mathbf{g}_{y_1}-\gamma_{y_1})\\
&\quad=\left[\frac{|\widetilde{r}_y|}{\widetilde{r}^2_x\widetilde{r}^1_{y_1}}\right](\gamma_x-\widetilde{\gamma}_x)+\left[\frac{\widetilde{r}^1_{y_2}}{\widetilde{r}^1_{y_1}}\right](\gamma_{y_1}-\widetilde{\gamma}_{y_1})-(\gamma_{y_2}-\widetilde{\gamma}_{y_2}).
\end{aligned}
\tag{T.3.3}
$$

Subtracting $[|\widetilde{r}_y|/(\widetilde{r}^2_x\widetilde{r}^1_{y_1})]\gamma_x+[\widetilde{r}^1_{y_2}/\widetilde{r}^1_{y_1}]\gamma_{y_1}$ from both sides of (T.3.3), we get

$$
\begin{aligned}
&\left[\left(\frac{|r_y|}{r^1_{y_1}r^2_x}\right)-\left(\frac{|\widetilde{r}_y|}{\widetilde{r}^2_x\widetilde{r}^1_{y_1}}\right)\right]\mathbf{g}_x-\left[\frac{|r_y|}{r^1_{y_1}r^2_x}\right]\gamma_x\\
&\qquad+\left[\left(\frac{r^1_{y_2}}{r^1_{y_1}}\right)-\left(\frac{\widetilde{r}^1_{y_2}}{\widetilde{r}^1_{y_1}}\right)\right]\mathbf{g}_{y_1}-\left[\frac{r^1_{y_2}}{r^1_{y_1}}\right]\gamma_{y_1}\\
&\quad=-\left[\frac{|\widetilde{r}_y|}{\widetilde{r}^2_x\widetilde{r}^1_{y_1}}\right]\widetilde{\gamma}_x-\left[\frac{\widetilde{r}^1_{y_2}}{\widetilde{r}^1_{y_1}}\right]\widetilde{\gamma}_{y_1}-(\gamma_{y_2}-\widetilde{\gamma}_{y_2}).
\end{aligned}
\tag{T.3.4}
$$

By (3.3) in Section 3.2,

$$
c(y,x)=\gamma_{y_2}-\left[\frac{r^1_{y_2}}{r^1_{y_1}}\right]\gamma_{y_1}-\left[\frac{|r_y|}{r^1_{y_1}r^2_x}\right]\gamma_x \quad \text{and}
$$

$$
\widetilde{c}(y,x)=\widetilde{\gamma}_{y_2}-\left[\frac{\widetilde{r}^1_{y_2}}{\widetilde{r}^1_{y_1}}\right]\widetilde{\gamma}_{y_1}-\left[\frac{|\widetilde{r}_y|}{\widetilde{r}^1_{y_1}\widetilde{r}^2_x}\right]\widetilde{\gamma}_x.
$$

Hence, (T.3.4) can be expressed, after rearranging it, as

$$0=\left[\left(\frac{r_{y_2}^1}{r_{y_1}^1}\right)-\left(\frac{\widetilde{r}_{y_2}^1}{\widetilde{r}_{y_1}^1}\right)\right]\mathbf{g}_{y_1}+\left[\left(\frac{|r_y|}{r_{y_1}^1 r_x^2}\right)-\left(\frac{|\widetilde{r}_y|}{\widetilde{r}_x^2\widetilde{r}_{y_1}^1}\right)\right]\mathbf{g}_x + \left[c(y,x)-\widetilde{c}(y,x)\right]. \tag{T.3.5}$$

Q.E.D.

PROOF OF THEOREM 3.2: Let F be defined as in the proof of Theorem 2.4, let the functionals $\Phi_{y_1,x}$ and $\Phi_{y_2,x}$ on F be as defined also in the proof of Theorem 2.4, and let Φ_{y_1,y_2}, $\Phi_{x,x}$, Φ_{y_1,y_1} be defined analogously. Define the functional Ξ on F by

$$\Xi(g)=\frac{\Phi_{y_1,y_2}(g)\Phi_{x,x}(g)-\Phi_{y_1,x}(g)\Phi_{y_2,x}(g)}{\Phi_{y_1,y_1}(g)\Phi_{x,x}(g)-\Phi_{y_1,x}^2(g)}.$$

By Lemma A.4 in the Supplemental Material, there exist finite $c>0$ and $\widetilde{\delta}>0$ such that, for all $h\in$ F with $\|h\|\le\widetilde{\delta}$,

$$\Xi(f+h)-\Xi(f)=D\Xi(f;h)+R\Xi(f;h),$$

where

$$\begin{aligned}D\Xi(f;h)=&\frac{[D\Phi_{y_1,y_2}(f;h)\Phi_{x,x}(f)-D\Phi_{y_1,x}(f;h)\Phi_{y_2,x}(f)]}{\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}^2(f)}\\&+\frac{[\Phi_{y_1,y_2}(f)D\Phi_{x,x}(f;h)-\Phi_{y_1,x}(f)D\Phi_{y_2,x}(f;h)]}{\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}^2(f)}\\&-\left[\Phi_{y_1,y_2}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}(f)\Phi_{y_2,x}(f)\right]\\&\times\frac{[D\Phi_{y_1,y_1}(f;h)\Phi_{x,x}+\Phi_{y_1,y_1}(f)D\Phi_{x,x}(f;h)]}{[\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}^2(f)]^2}\\&+\left[2D\Phi_{y_1,x}(f;h)\Phi_{y_1,x}(f)\right]\\&\times\frac{[\Phi_{y_1,y_2}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}(f)\Phi_{y_2,x}(f)]}{[\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}^2(f)]^2}\end{aligned}$$

and

$$\left|D\Xi(f;h)\right|\le c\|h\| \quad\text{and}\quad \left|R\Xi(f;h)\right|\le c\|h\|^2.$$

Letting $h=\widehat{f}-f$, it follows that, for $G=2$,

$$\begin{aligned}&\sqrt{N\sigma^{G+2}}D\Xi(f;\widehat{f}-f)\\&\quad=\sqrt{N\sigma^{G+2}}D\Phi_{y_1,y_2}(f;\widehat{f}-f)\left[\frac{\Phi_{x,x}(f)}{\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}^2(f)}\right]\end{aligned}$$

$$
\begin{aligned}
&-\sqrt{N\sigma^{G+2}}D\Phi_{y_1,x}(f;\widehat{f}-f)\left[\frac{\Phi_{y_2,x}(f)}{\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi^2_{y_1,x}(f)}\right] \\
&+\sqrt{N\sigma^{G+2}}D\Phi_{x,x}(f;\widehat{f}-f)\left[\frac{\Phi_{y_1,y_2}(f)}{\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi^2_{y_1,x}(f)}\right] \\
&-\sqrt{N\sigma^{G+2}}D\Phi_{y_2,x}(f;\widehat{f}-f)\left[\frac{\Phi_{y_1,x}(f)}{\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi^2_{y_1,x}(f)}\right] \\
&-\sqrt{N\sigma^{G+2}}D\Phi_{y_1,y_1}(f;\widehat{f}-f) \\
&\times\frac{\Phi_{x,x}[\Phi_{y_1,y_2}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}(f)\Phi_{y_2,x}(f)]}{[\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi^2_{y_1,x}(f)]^2} \\
&-\sqrt{N\sigma^{G+2}}D\Phi_{x,x}(f;\widehat{f}-f) \\
&\times\frac{\Phi_{y_1,y_1}[\Phi_{y_1,y_2}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}(f)\Phi_{y_2,x}(f)]}{[\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi^2_{y_1,x}(f)]^2} \\
&+\sqrt{N\sigma^{G+2}}D\Phi_{y_1,x}(f;\widehat{f}-f) \\
&\times\frac{2\Phi_{y_1,x}(f)[\Phi_{y_1,y_2}(f)\Phi_{x,x}(f)-\Phi_{y_1,x}(f)\Phi_{y_2,x}(f)]}{[\Phi_{y_1,y_1}(f)\Phi_{x,x}(f)-\Phi^2_{y_1,x}(f)]^2}.
\end{aligned}
$$

By Lemma A.5 in the Supplemental Material, for all j, s,

$$
\begin{aligned}
&\sqrt{N\sigma^{G+2}}D\Phi_{y_j,x_s}(f,\widehat{f}-f) \\
&\quad=\sqrt{N\sigma^{G+2}}\int_{\overline{M}}\big(\widehat{f}_{y_j}(y,x)-f_{y_j}(y,x)\big) \\
&\qquad\times\left[\frac{\mu\left(\beta_{x_s}(f)-\int_{\overline{M}^x}\beta_{x_s}(f)\mu\,dx\right)}{f}\right]dx+o_p(1), \\
&\sqrt{N\sigma^{G+2}}\big[\Phi_{x_j,x_s}(\widehat{f})-\Phi_{x_j,x_s}(f)\big] \\
&\quad=\sqrt{N\sigma^{G+2}}D\Phi_{x_x,x_s}(f,\widehat{f}-f)+\sqrt{N\sigma^{G+2}}R\Phi_{x_x,x_s}(f,\widehat{f}-f) \\
&\quad=o_p(1), \\
&\sqrt{N\sigma^{G+2}}D\Phi_{y_j,y_s}(f;\widehat{f}-f) \\
&\quad=\sqrt{N\sigma^{G+2}}\int_{\overline{M}}\big(\widehat{f}_{y_j}(y,x)-f_{y_j}(y,x)\big)
\end{aligned}
$$

$$\times \left[\frac{\mu\left(\alpha_{y_s}(f) - \int \alpha_{y_s}(f)\mu\, dx\right)}{f} \right] dx$$

$$+ \sqrt{N\sigma^{G+2}} \int_{\overline{M}} \left(\widehat{f}_{y_s}(y, x) - f_{y_s}(y, x)\right)$$

$$\times \left[\frac{\mu\left(\alpha_{y_j}(f) - \int \alpha_{y_j}(f)\mu\, dx\right)}{f} \right] dx + o_p(1),$$

and

$$\sqrt{N\sigma^{G+2}} \|\widehat{f} - f\|^2 \to 0 \quad \text{in probability,}$$

where α_{y_j} and β_{x_s} are as defined in the proof of Theorem 2.4. Hence,

$$\begin{aligned}
&\sqrt{N\sigma^{G+2}} D\Xi(f; \widehat{f} - f) \\
&\quad = \sqrt{N\sigma^{G+2}} \int_{\overline{M}} (\widehat{f}_{y_1} - f_{y_1}) \widetilde{\omega}_a(y, x)\, dx \\
&\qquad + \sqrt{N\sigma^{G+2}} \int_{\overline{M}} (\widehat{f}_{y_2} - f_{y_2}) \widetilde{\omega}_b(y, x)\, dx \\
&\qquad + \sqrt{N\sigma^{G+2}} \int_{\overline{M}} (\widehat{f}_{y_1} - f_{y_1}) \widetilde{\omega}_c(y, x)\, dx \\
&\qquad + \sqrt{N\sigma^{G+2}} \int_{\overline{M}} (\widehat{f}_{y_2} - f_{y_2}) \widetilde{\omega}_d(y, x)\, dx \\
&\qquad + \sqrt{N\sigma^{G+2}} \int_{\overline{M}} (\widehat{f}_{y_1} - f_{y_1}) \widetilde{\omega}_e(y, x)\, dx \\
&\qquad + \sqrt{N\sigma^{G+2}} \int_{\overline{M}} (\widehat{f}_{y_1} - f_{y_1}) \widetilde{\omega}_f(y, x)\, dx + o_p(1),
\end{aligned}$$

where $\widetilde{\omega}_a, \ldots, \widetilde{\omega}_f$ are as defined in Section 3. Thus,

$$\begin{aligned}
&\sqrt{N\sigma^{G+2}} D\Xi(f; \widehat{f} - f) \\
&\quad = \sqrt{N\sigma^{G+2}} \int_{\overline{M}} (\widehat{f}_{y_1} - f_{y_1}) \\
&\qquad \times \left[\widetilde{\omega}_a(y, x) + \widetilde{\omega}_c(y, x) + \widetilde{\omega}_e(y, x) + \widetilde{\omega}_f(y, x)\right] dx \\
&\qquad + \sqrt{N\sigma^{G+2}} \int_{\overline{M}} (\widehat{f}_{y_2} - f_{y_2}) \left[\widetilde{\omega}_b(y, x) + \widetilde{\omega}_d(y, x)\right] dx + o_p(1).
\end{aligned}$$

Let

$$\omega_1(x) = \big[\widetilde{\omega}_a(y,x) + \widetilde{\omega}_c(y,x) + \widetilde{\omega}_e(y,x) + \widetilde{\omega}_f(y,x)\big],$$
$$\omega_2(x) = \big[\widetilde{\omega}_b(y,x) + \widetilde{\omega}_d(y,x)\big],$$

and

$$\omega(x) = \big(\omega_1(x), \omega_2(x)\big).$$

By Lemma 5.3 in Newey (1994), with $k_1 = G = 2$, $k_2 = 1$, $l = 1$, and $m(\widehat{h}) - m(h_0) = D\Xi(f; \widehat{f} - f)$, it follows that

$$\sqrt{N\sigma^{G+2}}D\Xi(f; \widehat{f} - f) \to N(0, V_\beta),$$

where

$$V_\beta = \left[\int_{\overline{M}} \omega(x)\overline{K}_y\omega(x)' f(y,x)\,dx\right],$$

and where for $u = (u_1, u_2)$,

$$\overline{K}_y = \left[\int\left[\int\left(\frac{\partial K(u,x)}{\partial u}\right)dx\right]\left[\int\left(\frac{\partial K(u,x)}{\partial u}\right)dx\right]' du\right].$$

Since by Lemma A.5,

$$\sqrt{N\sigma^{G+2}}\big|R\Xi(f; \widehat{f} - f)\big| \le \big(\sqrt{N\sigma^{G+2}}\big)b_1\|\widehat{f} - f\|^2 \xrightarrow{p} 0,$$

it follows that

$$\begin{aligned}
&\sqrt{N\sigma^{G+2}}\big(\Xi(\widehat{f}) - \Xi(f)\big)\\
&\quad = \sqrt{N\sigma^{G+2}}\big(D\Xi(f; \widehat{f} - f) + R\Xi(f; \widehat{f} - f)\big)\\
&\quad = \sqrt{N\sigma^{G+2}}D\Xi(f; \widehat{f} - f) + o_p(1).
\end{aligned}$$

Hence,

$$\sqrt{N\sigma^{G+2}}\big(\Xi(\widehat{f}) - \Xi(f)\big) \to N(0, V_\beta).$$

To show that $\widehat{V}_\beta$ is a consistent estimator for V_β, we note that by Lemma A.6 in the Supplemental Material, for all j, s and all $w, z \in \{y_j, y_s, x_j, x_s\}$, for finite

$\overline{c} > 0$, and for $\|\widehat{f} - f\|$ sufficiently small,

$$\Bigg| \big(\Delta\partial_w \log \widehat{f}_{Y|X=x}(y)\big)\big(\Delta\partial_z \log \widehat{f}_{Y|X=x}(y)\big)\frac{\mu(y,x)^2}{\widehat{f}(y,x)} - \big(\Delta\partial_w \log f_{Y|X=x}(y)\big)\big(\Delta\partial_z \log f_{Y|X=x}(y)\big)\frac{\mu(y,x)^2}{f(y,x)}\Bigg|$$

is uniformly bounded on $\overline{M}$ by $\overline{c}\|\widehat{f} - f\|$. By Lemma A.5, for all j, s and all $w, z \in \{y_j, y_s, x_j, x_s\}$, $\Phi_{wz}(\widehat{f}) \xrightarrow{p} \Phi_{wz}(f)$. It then follows that for some finite $\overline{\overline{c}}$ and each $k \in \{a, b, \dots, f\}$,

$$\sup_{(y.x)\in\overline{M}} \big|\widehat{\widetilde{\omega}}_k(y,x) - \widetilde{\omega}_k(y,x)\big| \le \overline{\overline{c}}\|\widehat{f} - f\|.$$

Since $\overline{M}$ is compact and $\|\widehat{f} - f\| \xrightarrow{p} 0$, it follows that

$$\left[\int_{\overline{M}} \widehat{\omega}(x)\overline{K}_y\widehat{\omega}(x)'\widehat{f}(y,x)\,dx\right] \xrightarrow{p} \left[\int_{\overline{M}} \omega(x)\overline{K}_y\omega(x)'f(y,x)\,dx\right].$$

Hence, $\widehat{V}_\beta(y) \xrightarrow{p} V_\beta(y)$. *Q.E.D.*

PROOF OF THEOREM 4.3: Let $\mathcal{F}$ denote the set of functions g that satisfy Assumption 4.7 with $s = 0$. Let $\|g\|$ denote the maximum of the supremum of the values and derivatives up to the third order of g over the compact set defined as the closure of the neighborhood defined in Assumption 4.8. We first analyze the functional that, for any g, assigns the value of x at which $\partial^2 \log g_{Y|X=x}(y)/\partial x^2 = 0$. Define the functional $\Upsilon(g,x)$ by $\Upsilon(g,x) = \partial^2 \log g_{Y|X=x}(y)/\partial x^2$. We will show that there exists a functional $\kappa(g)$ on a neighborhood of f which is defined implicitly by $\Upsilon(g,\kappa(g)) = 0$ and satisfies a Taylor expansion of the form $\kappa(f+h) = \kappa(f) + D\kappa(f;h) + R\kappa(f;h)$ with $|R\kappa(f;h)|$ of the order $\|h\|^2$. We then use this to analyze the functional defining our estimator. We will denote $g_{Y,X}(y,x)$ by $g(x)$ and $g_X(x)$ by $\widetilde{g}(x)$, with similar notation for other functions in $\mathcal{F}$. By Assumption 4.8, the density functions $f_{Y,X}$ and f_X are uniformly bounded away from zero on the closure of any neighborhood defined in Assumption 4.8. It then follows by the definition of $\|\cdot\|$ that there exist $\nu_1 > 0$ small enough and $\nu_3 > 0$ such that, for all g in $\mathcal{F}$ with $\|g - f\| \le \nu_1$ and all h in $\mathcal{F}$ with $\|h - f\| \le \nu_1$, $g(x) + h(x) > \nu_3$ and $\widetilde{g}(x) + \widetilde{h}(x) > \nu_3$ for all x in a neighborhood $B(x^*, \delta)$ for some $\delta > 0$ small enough. For any such g, h and x, δ

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

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

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

$$
\begin{aligned}
&\Upsilon(g, x+\delta)-\Upsilon(g,x)\\
&\quad=\frac{\dfrac{\partial^2 g(x+\delta)}{\partial x^2}}{g(x+\delta)}-\frac{\left[\dfrac{\partial g(x+\delta)}{\partial x}\right]^2}{[g(x+\delta)]^2}-\frac{\dfrac{\partial^2\widetilde{g}(x+\delta)}{\partial x^2}}{\widetilde{g}(x+\delta)}+\frac{\left[\dfrac{\partial\widetilde{g}(x+\delta)}{\partial x}\right]^2}{[\widetilde{g}(x+\delta)]^2}\\
&\qquad-\left(\frac{\dfrac{\partial^2 g(x)}{\partial x^2}}{g(x)}-\frac{\left[\dfrac{\partial g(x)}{\partial x}\right]^2}{[g(x)]^2}-\frac{\dfrac{\partial^2\widetilde{g}(x)}{\partial x^2}}{\widetilde{g}(x)}+\frac{\left[\dfrac{\partial\widetilde{g}(x)}{\partial x}\right]^2}{[\widetilde{g}(x)]^2}\right).
\end{aligned}
$$

Define

$$
\begin{aligned}
&D_g\Upsilon(g,x;h)\\
&\quad=\left(\frac{\partial^2 h(x)}{\partial x^2}g(x)^2-\frac{\partial^2 g(x)}{\partial x^2}h(x)g(x)\right.\\
&\qquad\left.-2\frac{\partial h(x)}{\partial x}\frac{\partial g(x)}{\partial x}g(x)+2\left(\frac{\partial g(x)}{\partial x}\right)^2 h(x)\right)\Big/\big[g(x)\big]^3\\
&\qquad-\left(\frac{\partial^2\widetilde{h}(x)}{\partial x^2}\widetilde{g}(x)^2-\frac{\partial^2\widetilde{g}(x)}{\partial x^2}\widetilde{h}(x)\widetilde{g}(x)\right.\\
&\qquad\left.-2\frac{\partial\widetilde{h}(x)}{\partial x}\frac{\partial\widetilde{g}(x)}{\partial x}\widetilde{g}(x)+2\left(\frac{\partial\widetilde{g}(x)}{\partial x}\right)^2\widetilde{h}(x)\right)\Big/\big[\widetilde{g}(x)\big]^3 \quad \text{and}\\
&D_x\Upsilon(g,x;\delta)=\frac{\partial^3\log g_{Y|X=x}(y)}{\partial x^3}\delta,\\
&R_g\Upsilon(g,x;h)=\Upsilon(g+h,x)-\Upsilon(g,x)-D_g\Upsilon(g,x;h) \quad \text{and}\\
&R_x\Upsilon(g,x;\delta)=\Upsilon(g,x+\delta)-\Upsilon(g,x)-D_x\Upsilon(g,x;\delta).
\end{aligned}
$$

The definition of $\mathcal{F}$ and Assumptions 4.7 and 4.8 imply that there exists $a<\infty$ such that, for all (g,x) in a neighborhood of (f,x^*),

$$
\begin{aligned}
&\left\|D_x\Upsilon(g,x;\delta)\right\|\le a|\delta|; \quad \left\|R_x\Upsilon(g,x;\delta)\right\|\le a|\delta|^2;\\
&\left\|D_g\Upsilon(g,x;h)\right\|\le a\|h\|; \quad \text{and} \quad \left\|R_g\Upsilon(g,x;h)\right\|\le a\|h\|^2.
\end{aligned}
$$

Moreover, it can be verified that on a neighborhood of (f, x^*), $D_x Y(g, x; \delta)$ and $D_g Y(g, x; h)$ are also Fréchet differentiable on (g, x) and their derivatives are continuous on (g, x). Assumption 4.8 and the definition of $\mathcal{F}$ imply that, for all (g, x) in a neighborhood of (f, x^*), $D_x Y(g, x; \delta)$ is invertible. It then follows by the Implicit Function theorem on Banach spaces that there exists a unique functional κ such that, for all g in a ν_4 neighborhood of f ($\nu_4 < \nu_1$),

$$Y\big(g, \kappa(g)\big) = 0.$$

The Fréchet derivative at g is given by

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

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

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

$$\begin{aligned}\Psi\big(g, \kappa(g)\big) = -\bigg[&\frac{\partial^2 g_{Y,X}(y, \kappa(g))}{\partial y_2\, \partial x} g_{Y,X}\big(y, \kappa(g)\big) \\ &- \frac{\partial g_{Y,X}(y, \kappa(g))}{\partial y_2}\frac{\partial g_{Y,X}(y, \kappa(g))}{\partial x}\bigg] \\ \bigg/\bigg[&\frac{\partial^2 g_{Y,X}(y, \kappa(g))}{\partial y_1\, \partial x} g_{Y,X}\big(y, \kappa(g)\big) \\ &- \frac{\partial g_{Y,X}(y, \kappa(g))}{\partial y_1}\frac{\partial g_{Y,X}(y, \kappa(g))}{\partial x}\bigg].\end{aligned}$$

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

$$\begin{aligned}&D_g\Psi\big(g, x^*; h\big) \\ &\quad = -\bigg[\frac{\partial^2 h(x^*)}{\partial y_2\, \partial x} g\big(x^*\big) + \frac{\partial^2 g(x^*)}{\partial y_2\, \partial x} h\big(x^*\big) \\ &\qquad - \frac{\partial h(x^*)}{\partial y_2}\frac{\partial g(x^*)}{\partial x} - \frac{\partial g(x^*)}{\partial y_2}\frac{\partial h(x^*)}{\partial x}\bigg] \\ &\qquad \bigg/\bigg[\frac{\partial^2 g(x^*)}{\partial y_1\, \partial x} g\big(x^*\big) - \frac{\partial g(x^*)}{\partial y_1}\frac{\partial g(x^*)}{\partial x}\bigg]\end{aligned}$$

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

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

$$\left.-\frac{\partial h(x^*)}{\partial y_1}\frac{\partial g(x^*)}{\partial x}-\frac{\partial g(x^*)}{\partial y_1}\frac{\partial h(x^*)}{\partial x}\right]$$

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

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

Then,

$$D\Psi(f,\kappa(f);h)=D_f\Psi(f,x^*;h)+D_x\Psi(f,x^*;D\kappa(f;h))\quad\text{and}$$

$$R\Psi(f,\kappa(f);h)=\Psi(f+h,\kappa(f+h))-\Psi(f,\kappa(f))$$

$$-D\Psi(f,\kappa(f);h).$$

The properties we derived on $D\kappa$ and $R\kappa$ and Assumptions 4.7–4.9 imply that, for some $b<\infty$, $|D\Psi(f,\kappa(f);h)|\le b\|h\|$ and $|R\Psi(f,\kappa(f);h)|\le b\|h\|^2$. By applying Lemma 5.3 in Newey (1994) repeatedly, as we did in Lemma A.5, it follows by Assumptions 4.10 and 4.11 that when $h=\widehat{f}-f$,

$$\sqrt{N\sigma_N^7}D\Psi(f,\kappa(f);h)$$

$$=\sqrt{N\sigma_N^7}\frac{-[f(x^*)]}{\left[\frac{\partial^2 f(x^*)}{\partial y_1\,\partial x}f(x^*)-\frac{\partial f(x^*)}{\partial y_1}\frac{\partial f(x^*)}{\partial x}\right]}\frac{\partial^2 h(x^*)}{\partial y_2\,\partial x}$$

$$+\sqrt{N\sigma_N^7}\frac{\left[\frac{\partial^2 f(x^*)}{\partial y_2\,\partial x}f(x^*)-\frac{\partial f(x^*)}{\partial y_2}\frac{\partial f(x^*)}{\partial x}\right][f(x^*)]}{\left[\frac{\partial^2 f(x^*)}{\partial y_1\,\partial x}f(x^*)-\frac{\partial f(x^*)}{\partial y_1}\frac{\partial f(x^*)}{\partial x}\right]^2}\frac{\partial^2 h(x^*)}{\partial y_1\,\partial x}$$

$$+\sqrt{N\sigma_N^7}\frac{\partial\left(\frac{-\partial^2\log f_{Y|X=x^*}(y)/\partial y_2\,\partial x}{\partial^2\log f_{Y|X=x^*}(y)/\partial y_1\,\partial x}\right)}{\partial x}$$

$$\times\left(\frac{\partial^3\log f_{Y|X=x^*}(y)}{\partial x^3}\right)^{-1}\left[\frac{-1}{f(x^*)}\right]\frac{\partial^2 h(x^*)}{\partial x^2}+o_p(1).$$

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

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

By Lemma 5.3 in Newey (1994), with $l = 2$, $k_1 = 3$, $k_2 = 0$, it follows that

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

where $\widetilde{V}$ is as defined prior to the statement of Theorem 4.3. Assumptions 4.7, 4.10, and 4.11 together with Lemma B.3 in Newey (1994) imply that $\|\widehat{f} - f\| = O_p(\sqrt{\ln(N)/(N\sigma^9)} + \sigma^s)$. By Assumption 4.11, $\sqrt{N\sigma^7}(\sqrt{\ln(N)/(N\sigma^9)} + \sigma^s)^2 \to 0$. Hence, since $|R\Psi(f, \kappa(f); \widehat{f} - f)| \le b\|\widehat{f} - f\|^2$, it follows that $\sqrt{N\sigma^7} R\Psi(f, \kappa(f); \widehat{f} - f) = o_p(1)$. Hence,

$$
\begin{aligned}
&\sqrt{N\sigma_N^7}\big[\partial \widehat{m^1(y_2, \varepsilon_1)}/\partial y_2 - \partial m^1(y_2, \varepsilon_1)/\partial y_2\big] \\
&= \sqrt{N\sigma^7}\big(\Psi\big(\widehat{f}, \kappa(\widehat{f})\big) - \Psi\big(f, \kappa(f)\big)\big) \\
&= \sqrt{N\sigma^7} D\Psi\big(f, \kappa(f); \widehat{f} - f\big) + \sqrt{N\sigma^7} R\Psi\big(f, \kappa(f); \widehat{f} - f\big) \\
&= \sqrt{N\sigma^7} D\Psi\big(f, \kappa(f); \widehat{f} - f\big) + o_p(1) \\
&\to N(0, \widetilde{V}) \quad \text{in distribution.}
\end{aligned}
$$

To show that $\widehat{\widetilde{V}}$ converges to $\widetilde{V}$ in probability, we first note that, as shown above,

$$
\begin{aligned}
\big|\widehat{x}^* - x^*\big| = \big|\kappa(\widehat{f}) - \kappa(f)\big| &\le \big|D\kappa(f; \widehat{f} - f)\big| + \big|R\kappa(f; \widehat{f} - f)\big| \\
&\le c\|\widehat{f} - f\| + c\|\widehat{f} - f\|^2 \to 0 \quad \text{in probability.}
\end{aligned}
$$

Define the functionals

$$\Omega_1(g) = \frac{\left[\dfrac{\partial^2 g(\kappa(g))}{\partial y_2\,\partial x} g(\kappa(g)) - \dfrac{\partial g(\kappa(g))}{\partial y_2}\dfrac{\partial g(\kappa(g))}{\partial x}\right][g(\kappa(g))]}{\left[\dfrac{\partial^2 g(\kappa(g))}{\partial y_1\,\partial x} g(\kappa(g)) - \dfrac{\partial(\kappa(g))}{\partial y_1}\dfrac{\partial g(\kappa(g))}{\partial x}\right]^2},$$

$$\Omega_2(g) = \frac{-[g(\kappa(g))]}{\left[\dfrac{\partial^2 g(\kappa(g))}{\partial y_1\,\partial x} g(\kappa(g)) - \dfrac{\partial g(\kappa(g))}{\partial y_1}\dfrac{\partial g(\kappa(g))}{\partial x}\right]}, \quad \text{and}$$

$$\Omega_3(g) = \frac{\partial\left(\dfrac{-\partial^2 \log g_{Y|X=\kappa(g)}(y)/\partial y_2\,\partial x}{\partial^2 \log g_{Y|X=\kappa(g)}(y)/\partial y_1\,\partial x}\right)}{\partial x} \times \left(\frac{\partial^3 \log g_{Y|X=\kappa(g)}(y)}{\partial x^3}\right)^{-1}\left[\frac{-1}{g(\kappa(g))}\right].$$

Following steps analogous to those in the proof of Lemma A.1 in the Supplemental Material, it can be shown that there exist a finite $\overline{c} > 0$ and functionals $D\Omega_1$, $R\Omega_1$, $D\Omega_2$, $R\Omega_2$, $D\Omega_3$, and $R\Omega_3$ such that, for all $\widehat{f}$ in a neighborhood of f, and for $j = 1, 2, 3$,

$$\begin{aligned}\left|\Omega_j(\widehat{f}) - \Omega_j(f)\right| &= \left|D\Omega_j(f; \widehat{f} - f) + R\Omega_j(f; \widehat{f} - f)\right| \\ &\le \left|D\Omega_j(f; \widehat{f} - f)\right| + \left|R\Omega_j(f; \widehat{f} - f)\right| \\ &\le \overline{c}\|\widehat{f} - f\| + \overline{c}\|\widehat{f} - f\|^2.\end{aligned}$$

Hence, since $\|\widehat{f} - f\| \xrightarrow{p} 0$, it follows that $|\Omega_j(\widehat{f}) - \Omega_j(f)| \xrightarrow{p} 0$. By the definition of $\widetilde{V}$, this implies that $\widehat{\widetilde{V}} \to \widetilde{V}$ in probability. *Q.E.D.*

PROOF OF THEOREM 4.4: The proof is similar to the proof of Theorem 4.3. Let $\mathcal{F}$ denote the set of functions g that satisfy Assumption 4.7′ with $s = 0$. Let $\|g\|$ denote the maximum of the supremum of the values and derivatives up to the second 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 different values x_1 and x_2, at which $\partial \log g_{Y|X=x_1}(y)/\partial x = 0$ and $\partial \log g_{Y|X=x_2}(y)/\partial x = 0$. When $g = f$, $x_1 = x^*$ and $x_2 = x^{**}$ in the definition of the estimator for $\partial m^1(y_2, \varepsilon_1)/\partial y_2$. As in the proof of Theorem 4.3, we will denote $g_{Y,X}(y, x)$ by $g(x)$ and $g_X(x)$ by $\widetilde{g}(x)$, with similar notation for other functions in $\mathcal{F}$. Since $x_1 \neq x_2$, the asymptotic covariance of our kernel estimators for the values of x_1 and x_2 is zero. Define the functional $\Upsilon(g, x_1, x_2) = (\partial \log g_{Y|X=x_1}(y)/\partial x, \partial \log g_{Y|X=x_2}(y)/\partial x)'$. We first show that there exists a functional $\kappa(g) = (\kappa^1(g), \kappa^2(g))$ satisfying $\kappa(f) = (x^*, x^{**})$ which is defined implicitly in a neighborhood of f by

$$\Upsilon\big(g, \kappa^1(g), \kappa^2(g)\big) = 0.$$

By Assumption 4.8′, the density functions $f_{Y,X}$ and f_X are uniformly bounded away from zero on the closure of the neighborhood defined in Assumption 4.8′. It then follows by the definition of $\|\cdot\|$ that there exist $\nu_1, \nu_2, \nu_3 > 0$ small enough such that, for all g in $\mathcal{F}$ with $\|g - f\| \leq \nu_1$, all h in $\mathcal{F}$ with $\|h\| \leq \nu_2$, all x_1 in a δ_1 neighborhood of x^*, and all x_2 in a δ_2 neighborhood x^{**}, $g(x_j) + h(x_j) > \nu_3$ and $\widetilde{g}(x_j) + \widetilde{h}(x_j) > \nu_3$ ($j = 1, 2$). Hence, by Assumptions 4.7′ and 4.8′,

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

Define

$$\begin{aligned}
&D_g \Upsilon(g, x_1, x_2; h) \\
&= \begin{pmatrix}
\dfrac{\frac{\partial h(x_1)}{\partial x}}{g(x_1)} - \dfrac{\frac{\partial g(x_1)}{\partial x} h(x_1)}{(g(x_1))^2} - \dfrac{\frac{\partial \widetilde{h}(x_1)}{\partial x}}{\widetilde{g}(x_1)} + \dfrac{\frac{\partial \widetilde{g}(x_1)}{\partial x} \widetilde{h}(x_1)}{(\widetilde{g}(x_1))^2} \\
\dfrac{\frac{\partial h(x_2)}{\partial x}}{g(x_2)} - \dfrac{\frac{\partial g(x_2)}{\partial x} h(x_2)}{(g(x_2))^2} - \dfrac{\frac{\partial \widetilde{h}(x_2)}{\partial x}}{\widetilde{g}(x_2)} + \dfrac{\frac{\partial \widetilde{g}(x_2)}{\partial x} \widetilde{h}(x_2)}{(\widetilde{g}(x_2))^2}
\end{pmatrix},
\end{aligned}$$

$$
\begin{aligned}
D_{x_1} \Upsilon(g, x_1, x_2; \delta_1) &= \frac{\partial^2 \log g_{Y|X=x_1}(y)}{\partial x^2} \delta_1, \\
D_{x_2} \Upsilon(g, x_1, x_2; \delta_2) &= \frac{\partial^2 \log g_{Y|X=x_2}(y)}{\partial x^2} \delta_2, \\
R_g \Upsilon(g, x_1, x_2) &= \Upsilon(g+h, x_1, x_2) - \Upsilon(g, x_1, x_2) \\
&\qquad - D_g \Upsilon(g, x_1, x_2; h), \\
R_{x_1} \Upsilon(g, x_1, x_2; \delta_1) &= \Upsilon(g, x_1+\delta_1, x_2) - \Upsilon(g, x_1, x_2) \\
&\qquad - D_{x_1} \Upsilon(g, x_1, x_2; \delta_1), \quad \text{and} \\
R_{x_2} \Upsilon(g, x_1, x_2; \delta_2) &= \Upsilon(g, x_1, x_2+\delta_2) - \Upsilon(g, x_1, x_2) \\
&\qquad - D_{x_2} \Upsilon(g, x_1, x_2; \delta_2).
\end{aligned}
$$

The definition of $\mathcal{F}$ and Assumptions 4.7′ and 4.8′ imply that, for some $a < \infty$,

$$
\begin{aligned}
&\left\| D_{x_j} \Upsilon(g, x_1, x_2; \delta_j) \right\| \leq a|\delta_j| \quad \text{and} \\
&\left\| R_{x_j} \Upsilon(g, x_1, x_2; \delta_j) \right\| \leq a|\delta_j|^2 \quad (\text{for } j = 1, 2), \\
&\left\| D_g \Upsilon(g, x_1, x_2; h) \right\| \leq a\|h\| \quad \text{and} \quad \left\| R_g \Upsilon(g, x_1, x_2; h) \right\| \leq a\|h\|^2.
\end{aligned}
$$

Hence, $D_x \Upsilon(g, x_1, x_2; \delta)$ is the Fréchet derivative of Υ with respect to x and $D_g \Upsilon(g, x_1, x_2; h)$ is the Fréchet derivative of Υ with respect to g. By their definitions and Assumptions 4.7′ and 4.8′, it follows that both Fréchet derivatives are themselves Fréchet differentiable and their derivatives are continuous and uniformly bounded on a neighborhood of (f, x^*, x^{**}). Moreover, each $D_{x_j} \Upsilon(g, x_1, x_2; \delta_j)$ $(j = 1, 2)$ has a continuous inverse on a neighborhood of (f, x^*, x^{**}). It then follows by the Implicit Function theorem on Banach spaces that there exist unique functionals κ^1 and κ^2 such that $\kappa^1(f) = x^*$, $\kappa^2(f) = x^{**}$, for all g in a neighborhood of f,

$$
\Upsilon\bigl(g, \kappa^1(g), \kappa^2(g)\bigr) = 0,
$$

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

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

Moreover, κ^1 and κ^2 satisfy a first order Taylor expansion around f with remainder term for $(\kappa^j(f+h) - \kappa^j(f))$ bounded by $c\|h\|^2$ for a constant $c < \infty$.

Define the functional $\Psi(g, x_1, x_2)$ by

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

Then, $\Psi(\widehat{f}, \kappa^1(\widehat{f}), \kappa^2(\widehat{f})) = \partial m^1 \widehat{(y_2, \varepsilon_1)}/\partial y_2$ and $\Psi(f, \kappa^1(f), \kappa^2(f)) = \partial m^1(y_2, \varepsilon_1)/\partial y_2$. 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 $|\delta_2|$ in a neighborhood of 0,

$$\begin{aligned}
&\Psi\big(f + h, x^*, x^{**}\big) - \Psi\big(f, x^*, x^{**}\big) \\
&\quad = \left(\left[\frac{\partial f(x^{**})}{\partial y_2} + \frac{\partial h(x^{**})}{\partial y_2} \right] \big[f\big(x^*\big) + h\big(x^*\big)\big] \right. \\
&\qquad \left. - \left[\frac{\partial f(x^{*})}{\partial y_2} + \frac{\partial h(x^{*})}{\partial y_2} \right] \big[f\big(x^{**}\big) + h\big(x^{**}\big)\big] \right) \\
&\qquad \Big/ \left(\left[\frac{\partial f(x^{*})}{\partial y_1} + \frac{\partial h(x^{*})}{\partial y_1} \right] \big[f\big(x^{**}\big) + h\big(x^{**}\big)\big] \right. \\
&\qquad \left. - \left[\frac{\partial f(x^{**})}{\partial y_1} + \frac{\partial h(x^{**})}{\partial y_1} \right] \big[f\big(x^*\big) + h\big(x^*\big)\big] \right) \\
&\qquad - \frac{\dfrac{\partial f(x^{**})}{\partial y_2} f(x^*) - \dfrac{\partial f(x^*)}{\partial y_2} f(x^{**})}{\dfrac{\partial f(x^*)}{\partial y_1} f(x^{**}) - \dfrac{\partial f(x^{**})}{\partial y_1} f(x^*)}, \\
&\Psi\big(f, x^* + \delta_1, x^{**}\big) - \Psi\big(f, x^*, x^{**}\big) \\
&\quad = \frac{\dfrac{\partial f(x^{**})}{\partial y_2} f(x^* + \delta_1) - \dfrac{\partial f(x^* + \delta_1)}{\partial y_2} f(x^{**})}{\dfrac{\partial f(x^* + \delta_1)}{\partial y_1} f(x^{**}) - \dfrac{\partial f(x^{**})}{\partial y_1} f(x^* + \delta_1)} \\
&\qquad - \frac{\dfrac{\partial f(x^{**})}{\partial y_2} f(x^*) - \dfrac{\partial f(x^*)}{\partial y_2} f(x^{**})}{\dfrac{\partial f(x^*)}{\partial y_1} f(x^{**}) - \dfrac{\partial f(x^{**})}{\partial y_1} f(x^*)}, \quad \text{and}
\end{aligned}$$

$$\Psi\big(f, x^*, x^{**} + \delta_2\big) - \Psi\big(f, x^*, x^{**}\big)$$
$$= \frac{\dfrac{\partial f(x^{**} + \delta_2)}{\partial y_2} f(x^*) - \dfrac{\partial f(x^*)}{\partial y_2} f(x^{**} + \delta_2)}{\dfrac{\partial f(x^*)}{\partial y_1} f(x^{**} + \delta_2) - \dfrac{\partial f(x^{**} + \delta_2)}{\partial y_1} f(x^*)}$$
$$- \frac{\dfrac{\partial f(x^{**})}{\partial y_2} f(x^*) - \dfrac{\partial f(x^*)}{\partial y_2} f(x^{**})}{\dfrac{\partial f(x^*)}{\partial y_1} f(x^{**}) - \dfrac{\partial f(x^{**})}{\partial y_1} f(x^*)}.$$

Define

$$D_f \Psi\big(f, x^*, x^{**}; h\big)$$
$$= \frac{\left[\dfrac{\partial h(x^{**})}{\partial y_2} f\big(x^*\big) - \dfrac{\partial h(x^*)}{\partial y_2} f\big(x^{**}\big) + \dfrac{\partial f(x^{**})}{\partial y_2} h\big(x^*\big) - \dfrac{\partial f(x^*)}{\partial y_2} h\big(x^{**}\big)\right]}{\left[\dfrac{\partial f(x^*)}{\partial y_1} f\big(x^{**}\big) - \dfrac{\partial f(x^{**})}{\partial y_1} f\big(x^*\big)\right]}$$
$$- \left[\frac{\partial f(x^{**})}{\partial y_2} f\big(x^*\big) - \frac{\partial f(x^*)}{\partial y_2} f\big(x^{**}\big)\right]$$
$$\times \frac{\left[\dfrac{\partial h(x^*)}{\partial y_1} f\big(x^{**}\big) - \dfrac{\partial h(x^{**})}{\partial y_1} f\big(x^*\big) + \dfrac{\partial f(x^*)}{\partial y_1} h\big(x^{**}\big) - \dfrac{\partial f(x^{**})}{\partial y_1} h\big(x^*\big)\right]}{\left[\dfrac{\partial f(x^*)}{\partial y_1} f\big(x^{**}\big) - \dfrac{\partial f(x^{**})}{\partial y_1} f\big(x^*\big)\right]^2},$$

$$R_f \Psi\big(f, x^*, x^{**}; h\big) = \Psi\big(f + h, x^*, x^{**}\big) - \Psi\big(f, x^*, x^{**}\big) - D_f \Psi\big(f, x^*, x^{**}; h\big),$$

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

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

$$D\Psi\big(f, x^*, x^{**}; h\big) = D_f \Psi\big(f, x^*, x^{**}; h\big) + D_{x_1} \Psi\big(f, x^*, x^{**}; D\kappa^1(f; h)\big) + D_{x_2} \Psi\big(f, x^*, x^{**}; D\kappa^2(f; h)\big), \quad \text{and}$$

$$R\Psi\big(f, x^*, x^{**}; h\big) = \Psi\big(f + h, \kappa^1(f + h), \kappa^2(f + h)\big) - \Psi\big(f, x^*, x^{**}\big) - D\Psi\big(f, x^*, x^{**}; h\big).$$

Assumptions 4.7′ and 4.8′ imply that

$$\left|D\Psi\left(f, x^*, x^{**}; h\right)\right| \leq a\|h\| \quad \text{and} \quad \left|R\Psi\left(f, x^*, x^{**}; h\right)\right| \leq a\|h\|^2.$$

Assumptions 4.10′ and 4.11′ imply, following steps as in the proof of Lemma A.5 in the Supplemental Material, that for $\widehat{f} - f = h$,

$$\begin{aligned}
&\sqrt{N\sigma_N^5} D\Psi\left(f, x^*, x^{**}; h\right) \\
&= \sqrt{N\sigma_N^5} \frac{\left[\dfrac{\partial h(x^{**})}{\partial y_2} f(x^*) - \dfrac{\partial h(x^*)}{\partial y_2} f(x^{**})\right]}{\left[\dfrac{\partial f(x^*)}{\partial y_1} f(x^{**}) - \dfrac{\partial f(x^{**})}{\partial y_1} f(x^*)\right]} \\
&\quad - \sqrt{N\sigma_N^5} \\
&\quad \times \frac{\left[\dfrac{\partial f(x^{**})}{\partial y_2} f(x^*) - \dfrac{\partial f(x^*)}{\partial y_2} f(x^{**})\right]\left[\dfrac{\partial h(x^*)}{\partial y_1} f(x^{**}) - \dfrac{\partial h(x^{**})}{\partial y_1} f(x^*)\right]}{\left[\dfrac{\partial f(x^*)}{\partial y_1} f(x^{**}) - \dfrac{\partial f(x^{**})}{\partial y_1} f(x^*)\right]^2} \\
&\quad + \sqrt{N\sigma_N^5} \\
&\quad \times \frac{\dfrac{\partial}{\partial x_1}\left(\dfrac{\partial \log f_{Y|X=x^{**}}(y)/\partial y_2 - \partial \log f_{Y|X=x_1}(y)/\partial y_2}{\partial \log f_{Y|X=x_1}(y)/\partial y_1 - \partial \log f_{Y|X=x^{**}}(y)/\partial y_1}\right)\Bigg|_{x_1=x^*}}{\left(\dfrac{\partial^2 \log f_{Y|X=x^*}(y)}{\partial x^2}\right) f(x^*)} \left(\frac{-\partial h(x^*)}{\partial x}\right) \\
&\quad + \sqrt{N\sigma_N^5} \\
&\quad \times \frac{\dfrac{\partial}{\partial x_2}\left(\dfrac{\partial \log f_{Y|X=x_2}(y)/\partial y_2 - \partial \log f_{Y|X=x^*}(y)/\partial y_2}{\partial \log f_{Y|X=x^*}(y)/\partial y_1 - \partial \log f_{Y|X=x_2}(y)/\partial y_1}\right)\Bigg|_{x_2=x^{**}}}{\left(\dfrac{\partial^2 \log f_{Y|X=x^{**}}(y)}{\partial x^2}\right) f(x^{**})} \left(\frac{-\partial h(x^{**})}{\partial x}\right) \\
&\quad + o_p(1).
\end{aligned}$$

By Lemma 5.2 in Newey (1994), it follows by Assumptions 4.7′, 4.8′, 4.10′, and 4.11′ that $\sqrt{N\sigma_N^5} D\Psi(f, x^*, x^{**}; h) \to N(0, \overline{V})$, where $\overline{V}$ is as defined prior to the statement of Theorem 4.4. Since $\sqrt{N\sigma_N^5}|R\Psi(f, x^*, x^{**}; \widehat{f} - f)| \leq \sqrt{N\sigma_N^5} a\|\widehat{f} - f\|^2 = o_p(1)$, by Assumption 4.10′ and Lemma B.2 in Newey

(1994), we can conclude that

$$\sqrt{N\sigma_N^5}\left(\frac{\partial \widehat{m^1(y_2,\varepsilon_1)}}{\partial y_2}-\frac{\partial m^1(y_2,\varepsilon_1)}{\partial y_2}\right)$$
$$=\sqrt{N\sigma_N^5}\big(\Psi\big(\widehat{f},\kappa^1(\widehat{f}),\kappa^2(\widehat{f})\big)-\Psi\big(f,x^*,x^{**}\big)\big)\xrightarrow{d}N(0,\overline{V}).$$

Following steps analogous to those used in the proof of Theorem 4.3, as in the proof of Lemma A.1, it can be shown that the estimators for each ω_j^i defined before the statement of Theorem 4.4 converge in probability to ω_j^i. Similarly, $\widehat{f}_{Y,X}(y,\widehat{x}^*)\xrightarrow{p}f_{Y,X}(y,x^*)$, and $\widehat{f}_{Y,X}(y,\widehat{x}^{**})\xrightarrow{p}f_{Y,X}(y,x^{**})$. Hence, $\widehat{\overline{V}}\xrightarrow{p}\overline{V}$. *Q.E.D.*

REFERENCES

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models With Conditional Moments Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843. [2]

ALTONJI, J. G., AND R. L. MATZKIN (2001): "Panel Data Estimators for Nonseparable Models With Endogenous Regressors," Working Paper T0267, NBER. [3]

——— (2005): "Cross Section and Panel Data Estimators for Nonseparable Models With Endogenous Regressors," *Econometrica*, 73 (3), 1053–1102. [3]

ALTONJI, J. G., H. ICHIMURA, AND T. OTSU (2012): "Estimating Derivatives in Nonseparable Models With Limited Dependent Variables," *Econometrica*, 80 (4), 1701–1719. [3]

BERRY, S. T., AND P. A. HAILE (2009): "Identification in Differentiated Products Markets Using Market Level Data," Working Paper, Yale University. [4]

——— (2011): "Identification in a Class of Nonparametric Simultaneous Equations Models," Discussion Paper 1787, Cowles Foundation. [4,12]

——— (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82 (5), 1749–1797. [4]

BHATTACHARYA, P. K. (1963): "On an Analog of Regression Analysis," *The Annals of Mathematical Statistics*, 34 (4), 1459–1473. [2]

BLUNDELL, R., AND R. L. MATZKIN (2014): "Control Functions in Nonseparable Simultaneous Equations Models," *Quantitative Economics*, 5 (2), 271–295. [3]

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress*, Vol. II, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky. Cambridge: Cambridge University Press. [3]

BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75 (6), 1613–1669. [2]

BOWDEN, R. (1973): "The Theory of Parametric Identification," *Econometrica*, 41, 1069–1074. [2]

BROWN, D. J., AND R. L. MATZKIN (1998): "Estimation of Nonparametric Functions in Simultaneous Equations Models, With an Application to Consumer Demand," Discussion Paper 1175, Cowles Foundation. [4]

BROWN, D., M. WEGKAMP, AND R. DEB (2007): "Tests of Independence in Separable Econometric Models: Theory and Application," Discussion Paper 1395RR, Cowles Foundation. [12]

CHEN, X., AND D. POUZO (2012): "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals," *Econometrica*, 80 (1), 277–321. [2]

CHEN, X., V. CHERNOZHUKOV, S. LEE, AND W. NEWEY (2014): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, 82 (2), 785–809. [2]

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261. [2]

CHERNOZHUKOV, V., G. IMBENS, AND W. NEWEY (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4–14. [2]

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71 (5), 1405–1441. [3]

CHIAPPORI, P.-A., AND I. KOMUNJER (2009): "On the Nonparametric Identification of Multiple Choice Models," Discussion Paper, UCSD. [4]

DAROLLES, S., Y. FAN, J. P. FLORENS, AND E. RENAULT (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79 (5), 1541–1565. [2]

DAROLLES, S., J. P. FLORENS, AND E. RENAULT (2002): "Nonparametric Instrumental Regression," Report, IDEI Toulouse. [2]

FISHER, F. M. (1959): "Generalization of the Rank and Order Conditions for Identifiability," *Econometrica*, 27 (3), 431–447. [2]

——— (1961): "Identifiability Criteria in Nonlinear Systems," *Econometrica*, 29, 574–590. [2]

——— (1965): "Identifiability Criteria in Nonlinear Systems: A Further Note," *Econometrica*, 33, 197–205. [2]

——— (1966): *The Identification Problem in Econometrics*. New York: McGraw-Hill. [2]

HAAVELMO, T. M. (1943): "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11 (1), 1–12. [2]

——— (1944): "The Probability Approach in Econometrics," *Econometrica*, 12 (Supplement), 1–115. [2]

HALL, P., AND J. L. HOROWITZ (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, 33 (6), 2904–2929. [2]

HAUSMAN, J. A. (1983): "Specification and Estimation of Simultaneous Equation Models," in *Handbook of Econometrics*, Vol. 1, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland. [2]

HOROWITZ, J. L., AND S. LEE (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica*, 75 (4), 1191–1208. [2]

HSIAO, C. (1983): "Identification," in *Handbook of Econometrics*, Vol. 1, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland. [2]

HURWICZ, L. (1950): "Generalization of the Concept of Identification," in *Statistical Inference in Dynamic Economic Models*. Cowles Commission Monograph, Vol. 10. New York: Wiley. [2]

IMBENS, G. W., AND W. K. NEWEY (2003): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Working Paper, MIT. [3]

——— (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77 (5), 1481–1512. [3]

KOENKER, R. (2005): *Quantile Regression*. Econometric Society Monograph, Vol. 38. Cambridge: Cambridge University Press. [2]

KOOPMANS, T. C. (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17 (2), 125–144. [2]

KOOPMANS, T. C., AND O. REIERSOL (1950): "The Identification of Structural Characteristics," *The Annals of Mathematical Statistics*, 21, 165–181. [2]

KOOPMANS, T. C., A. RUBIN, AND R. B. LEIPNIK (1950): "Measuring the Equation System of Dynamic Economics," in *Statistical Inference in Dynamic Economic Models*. Cowles Commission Monograph, Vol. 10. New York: Wiley. [2]

MANSKI, C. F. (1983): "Closest Empirical Distribution Estimation," *Econometrica*, 51 (2), 305–319. [4]

MATZKIN, R. L. (1992): "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, 60, 239–270. [7]

——— (1994): “Restrictions of Economic Theory in Nonparametric Methods,” in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engel and D. L. McFadden. Amsterdam: Elsevier, Chapter 42. [7]

——— (1999): “Nonparametric Estimation of Nonadditive Random Functions,” Working Paper. [3,5,7]

——— (2003): “Nonparametric Estimation of Nonadditive Random Functions,” *Econometrica*, 71, 1339–1375. [3,5,7]

——— (2007a): “Nonparametric Identification,” in *Handbook of Econometrics*, Vol. 6B, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: Elsevier, Chapter 73, 5307–5368. [5]

——— (2007b): “Heterogeneous Choice,” in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge: Cambridge University Press. [4,5,12]

——— (2008): “Identification in Nonparametric Simultaneous Equations,” *Econometrica*, 76 (5), 945–978. [4,8,9,12,21,41]

——— (2010): “Estimation of Nonparametric Parametric Models With Simultaneity,” Working Paper, UCLA. [12]

——— (2012): “Identification in Nonparametric Limited Dependent Variable Models With Simultaneity and Unobserved Heterogeneity,” *Journal of Econometrics*, 166, 106–115. [4]

——— (2013): “Nonparametric Identification in Structural Economic Models,” *Annual Review of Economics*, 5, 457–486. [4]

——— (2015): “Supplement to ‘Estimation of Nonparametric Models With Simultaneity’,” *Econometrica Supplemental Material*, 83, http://dx.doi.org/10.3982/ECTA9348. [45]

NADARAYA, E. (1964): “On Estimating Regression,” *Theory of Probability and Its Applications*, 9, 141–142. [2]

NEWEY, W. K. (1994): “Kernel Estimation of Partial Means and a General Variance Estimator,” *Econometric Theory*, 10, 233–253. [46,52,56,57,63,64]

NEWEY, W., AND J. POWELL (1989): “Instrumental Variables Estimation of Nonparametric Models,” Working Paper. [2]

——— (2003): “Instrumental Variables Estimation of Nonparametric Models,” *Econometrica*, 71 (5), 1565–1578. [2]

POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): “Semiparametric Estimation of Index Coefficients,” *Econometrica*, 57, 1403–1430. [3]

ROTHENBERG, T. J. (1971): “Identification in Parametric Models,” *Econometrica*, 39, 577–591. [2]

STOKER, T. M. (1986): “Consistent Estimation of Scaled Coefficients,” *Econometrica*, 54, 1461–1481. [3]

STONE, C. J. (1977): “Consistent Nonparametric Regression,” *The Annals of Statistics*, 5 (4), 595–620. [2]

WALD, A. (1950): “Note on Identification of Economic Relations,” in *Statistical Inference in Dynamic Economic Models*. Cowles Commission Monograph, Vol. 10. New York: Wiley. [2]

WATSON, G. S. (1964): “Smooth Regression Analysis,” *Sankhyā: The Indian Journal of Statistics*, 26 (4), 359–372. [2]

WEGGE, L. L. (1965): “Identifiability Criteria for a System of Equations as a Whole,” *Australian Journal of Statistics*, 7, 67–77. [2]

Dept. of Economics, University of California—Los Angeles, 8283 Bunche Hall, Los Angeles, CA 90095, U.S.A.; matzkin@econ.ucla.edu.

Manuscript received June, 2010; final revision received September, 2013.