SEMIPARAMETRIC INSTRUMENTAL VARIABLE ESTIMATION OF SIMULTANEOUS EQUATION SAMPLE SELECTION MODELS

by

Lung-Fei Lee

Discussion Paper No. 263, June 1991

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, MN 55455

Semiparametric Instrumental Variable Estimation of Simultaneous Equation Sample Selection Models

by
Lung-Fei Lee

Department of Economics
University of Minnesota
Minneapolis, Minnesota

May 1991

OUTLINE

The identification and estimation of a semiparametric simultaneous equation model with selectivity have been considered. The identification of structural parameters from reduced form parameters in the semiparametric model requires stronger conditions than the usual rank condition in the classical simultaneous equation model or the parametric simultaneous equation sample selection model. The necessary order condition for identification in the semiparametric model corresponds to the over-identification condition in the classical model. Semiparametric two-stage estimation methods which generalize the two-stage least squares method and the generalized two-stage least squares method for the parametric model are introduced. The semiparametric generalized least squares estimator is shown to be asymptotically efficient in a class of semiparametric instrumental variable estimators.

JEL classification number: 211

Keywords:
Semiparametric model, sample selection, simultaneity, index model, identification, instrumental variables, asymptotic efficiency.

Correspondence address:
Lung-Fei Lee, 1035 Management and Economics, Department of Economics, University of Minnesota, 271 19th Avenue S., Minneapolis, Minnesota 55455

Semiparametric Instrumental Variable Estimation of Simultaneous Equation Sample Selection Models

by Lung-Fei Lee*

1. Introduction

For the estimation of simultaneous equation sample selection models with parametric (normal) disturbances, several methods are available in the econometric literature, e.g., Lee, Maddala and Trost [1980], Lee [1981], Amemiya [1983] and Newey [1987]. The approach introduced in Lee, Maddala and Trost [1980] combines Heckman's two-stage and Theil's two-stage least squares procedures. Amemiya [1983] considered a class of estimators derived from modified minimum distance procedures. Relative efficiency of such procedures has been considered in Lee [1981], Amemiya [1983] and Newey [1987].

In this article, we will consider instrumental variable (IV) methods for the estimation of simultaneous equation sample selection models without parametric distributional assumptions. Semiparametric instrumental variable methods for the estimation of sample selection models have been considered in Powell [1987] (see also Robinson [1988]). In Powell [1987], since his interest is in general semiparametric instrumental variable methods, he has not focused attention on any specific simultaneous equation structures of the model. Many interesting issues, such as the rank identification condition, which are well-known for the classical simultaneous equation model, have not been addressed for the semiparametric simultaneous equation model. In this article, we are interested in the specific structure of simultaneous equation sample selection models. We investigate the problem of structural parameter identification, the role of identification conditions on semiparametric instrumental variable estimation, and the proper construction of instrumental variables from the system. We will also investigate the possible generalization of the (semiparametric) two-stage least squares estimation method and the construction of efficient semiparametric instrumental variable estimators.

* I appreciate having financial support from NSF under grant no. SES-9010516 for my research. The issues in this article are partially motivatcd by an empircial project under way with my colleagues Professors Mark R. Rosenzweig and Mark M. Pitt.

2. Semiparametric Simultaneous Equation Models with Selectivity and Instrumental Variable Estimation

In this article, our discussion will focus on the estimation of a single equation. The estimation of multiple equations can be easily generalized. Consider a single structural equation:

$$y^* = z^*\alpha_o + xJ\gamma_o + u_1, \tag{2.1}$$

where y^* is a latent endogenous variable, z^* is a G_1-dimensional row vector of latent endogenous variables not including y^*, x is a K-dimensional vector consisting of all exogenous variables in the system, and xJ, where J is a selection matrix, represents the subset of exogenous variables included in this structural equation. The reduced form equation of z^* is

$$z^* = x\Pi_2 + v_2, \tag{2.2}$$

where Π_2 is a $K \times G_1$ matrix and v_2 is a G_1 row vector of disturbances. The endogenous variables y^* and z^* are well-defined in the whole population but their sample observations y and z are subject to selection. The latent selection equation is

$$d^* = x\zeta_o + \epsilon, \tag{2.3}$$

where d^* is a latent variable. The values of y and z are observable if and only if $d^* > 0$. As in Ichimura [1987] and Powell [1987], we consider the index model framework where the joint distribution of (u_1, v_2, ϵ) conditional on x can be a function of the index $x\zeta_o$. Such a framework is slightly more general than the case where the disturbances are independent with x.

Conditional on $d_i^* > 0$ and x_i, (2.1) implies that

$$y_i = z_i\alpha_o + x_iJ\gamma_o + E(u_1|x_i\zeta_o,\ d_i^* > 0) + u_{di}, \tag{2.4}$$

where

$$u_{di} = u_{1i} - E(u_1|x_i\zeta_o,\ d_i^* > 0). \tag{2.5}$$

Let $K(\cdot)$ be a kernel function with a bandwidth parameter a_n (Silverman [1986] or Bierens [1985]). Let $w = (z, xJ)$ and $\beta_o = (\alpha_o, \gamma_o)$. For any possible value of (β, ζ) of (β_o, ζ_o), the conditional expectation function $E(y - w\beta | x_i\zeta, d^* > 0, \zeta)$ of $y - w\beta$, conditional on the random variables $x\zeta$ and $d^* > 0$, can be estimated by the following nonparametric regression function:

$$E_n(y - w\beta | x_i\zeta) = \frac{\sum_{j\neq i}^{n}(y_j - w_j\beta)K(\frac{x_i\zeta - x_j\zeta}{a_n})}{\sum_{j\neq i}^{n} K(\frac{x_i\zeta - x_j\zeta}{a_n})}, \tag{2.6}$$

where n is the sample size for the observations of (y, z, x) conditional on $d^* > 0$ (Ichimura [1987]). Given a $\sqrt{n}$-consistent estimate $\hat{\zeta}$ of ζ, Powell [1987] has proposed an instrumental variable method for the estimation of β from the following equation:

$$y_i - E_n(y|x_i\hat{\zeta}) = (w_i - E_n(w|x_i\hat{\zeta}))\beta + \hat{u}_{ni}, \tag{2.7}$$

where $E_n(s|x_i\zeta) = \sum_{j\neq i}^{n} s_j K(\frac{x_i\zeta - x_j\zeta}{a_n}) / \sum_{j\neq i}^{n} K(\frac{x_i\zeta - x_j\zeta}{a_n})$ for any random variable s (see also Robinson [1988]). Instrumental variable methods require the construction of instrumental variables for the transformed variables $w_i - E_n(w|x\hat{\zeta})$ (Powell [1987]). A simple instrumental variable estimator with instrumental variables p can be

$$\left(\sum_{i=1}^{n} p_i'(w_i - E_n(w|x\hat{\zeta})\right)^{-1} \sum_{i=1}^{n} p_i'(y_i - E_n(y|x\hat{\zeta})). \tag{2.8}$$

However, due to the technical difficulty of handling the denominator in the nonparametric regression function in (2.6), some modifications are needed to overcome this difficulty. Various ways have been introduced in the literature. Powell [1987] uses the denominator in (2.6) as the weight in the summations of (2.8) so as to cancel the denominator of (2.6).[1] An alternative suggestion is to trim the tails of the distribution of x or the index $x\hat{\zeta}$ (Robinson [1988], Klein and Spady [1987] and Ichimura and Lee [1988]). Trimming will be applied to the index $x\hat{\zeta}$ when its values are greater than some upper quantile or less than some lower quantile. Suppose that $t_n(x_i\hat{\zeta}, \hat{\xi})$ is a trimming indicator with value 0 when $x_i\hat{\zeta}$ is deleted, where $\hat{\xi}$ is a vector of quantiles. A simple unweighted instrumental variable estimator is

$$\hat{\beta}_p = \left(\sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}) p_i'(w_i - E_n(w|x\hat{\zeta})\right)^{-1} \sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}) p_i'(y_i - E_n(y|x\hat{\zeta})). \tag{2.9}$$

The trimming procedure is preferred so as not to complicate the proper function of weighting.

[1] The weighting scheme in Powell [1987] has nothing to do with the variance of the disturbance $\hat{u}_{ni}$ in (2.7).

As in the classical simultaneous equation model, the consistency of IV estimators depends on proper instruments constructed from the list of exogenous variables x in the system. Consistency of the IV estimators is possible only if the structural equation is identifiable. In subsequent sections, we will first address the identification problem of this system. Problems on how to select proper instrumental variables and the construction of efficient IV estimation of (2.7) will then be considered.

The estimation method can be generalized to cover more general cases where the selection mechanism is determined by several inequality conditions, for example, models with polytomous or sequential choices. For the general case, d^* in (2.3) is a finite dimensional (row) vector of latent equations. The samples of y and z are observed if and only if $d^* > 0$. The implied regression equation (2.4) becomes a model with multiple indices (Stoker [1986] and Ichimura and Lee [1988]). $x\zeta$ will represent a vector of indices with ζ being a matrix. The semiparametric estimation method above can be generalized. $K(\cdot)$ will now be a higher dimensional kernel with the dimension of $x\zeta$, and the trimming will be applied to all the indices in $x\zeta$. The bandwidth a_n needs to be wider in the nonparametric regression estimation in (2.6). The detailed analysis in the appendix is applicable to the general model.

3. Identification

Let

$$y^* = x\pi_1 + v_1 \tag{3.1}$$

be the reduced form equation for y^*. As in the classical simultaneous equation model, the identification of structural parameters is directly related to the reduced form parameters. Within the index model framework, identification of Equations (2.2), (2.3) and (3.1) has been considered in Ichimura [1987], Chamberlain [1986], Powell [1987], and Ichimura and Lee [1988]. Conditional on $d^* > 0$ and x,

$$E(y|x, d^* > 0) = x\pi_1 + E(v_1|x\zeta_o,\ x\zeta_o > -\epsilon, \zeta_o), \tag{3.2}$$

and

$$E(z|x, d^* > 0) = x\Pi_2 + E(v_2|x\zeta_o,\ x\zeta_o > -\epsilon, \zeta_o). \tag{3.3}$$

As shown in Ichimura [1987] for the single index model, ζ_o in the selection equation (2.3) can best be identified up to an unknown scale. When the regressors in x are all qualitative variables, ζ_o can not even be identified up to a scale. Identification requires the presence of a relevant continuous exogenous variable of which the coefficient has a known sign. A convenient normalization (Ichimura [1987]) is to set such a coefficient to be unity. Contrary to the classical simultaneous equation model, the reduced form parameter vectors π_1 and Π_2 in (3.1) and (2.2) are not identifiable. This is so because $x\pi_1$ and $x\zeta_o$ contain the same set of variables x and they can not be distinguished from each other in (3.2). Similarly, this is so for $x\Pi_2$ and $x\zeta_o$. This identification problem has been studied in Ichimura and Lee [1988] and Powell [1987] in the analysis of index models with nonparametric regression functions. The same conclusion has been derived in Chamberlain [1986] from the nonparametric likelihood function of the model. Even though π_1 and Π_2 are not identifiable, some transformations of them can be identified.

With the normalization suggested by Ichimura [1987], let $x\zeta_o = x_1 + x_2\delta_o$, where x_1 is a continuous exogenous variable. Conformably, $x\pi_1 = x_1\pi_{11} + x_2\pi_{12}$, $x\Pi_2 = x_1\pi'_{21} + x_2\Pi_{22}$, and $x\gamma_o = x_1\gamma_{o,1} + x_2\gamma_{o,2}$. The reduced form equation (3.1) can be rewritten into

$$y^* = x_2\pi_1^* + v_1^*, \tag{3.4}$$

where $\pi_1^* = \pi_{12} - \delta_o \pi_{11}$, and $v_1^* = v_1 + (x_1 + x_2\delta_o)\pi_{11}$. Similarly,

$$z^* = x_2\Pi_2^* + v_2^*, \tag{3.5}$$

where $\Pi_2^* = \Pi_{22} - \delta_o \pi_{21}'$, and $v_2^* = v_2 + (x_1 + x_2\delta_o)\pi_{21}'$. It follows that

$$E(y|x, d^* > 0) = x_2\pi_1^* + E(v_1^*|x\zeta_o, d^* > 0, \delta_o), \tag{3.6}$$

and

$$E(z|x, d^* > 0) = x_2\Pi_2^* + E(v_2^*|x\zeta_o, d^* > 0, \delta_o). \tag{3.7}$$

The index $x\zeta_o$ is distinguishable from $x_2\pi_1^*$ and $x_2\Pi_2^*$, because x_1 appears only in $x\zeta_o$ but neither in $x_2\pi_1^*$ nor in $x_2\Pi_2^*$. The transformed parameters π_1^* and Π_2^* and δ_o are identifiable as long as $E(v_j^*|x\zeta_o, d^* > 0, \delta_o)$ for $j = 1, 2$ are not linear in $x\zeta_o$ (Ichimura and Lee [1988]).

The structural parameters α and γ are related to the reduced form parameters π_1^*, Π_2^* and δ. Substituting (3.5) into the structural equation y^* (before imposing any explicit exclusion restrictions), we have

$$\begin{aligned} y^* &= z^*\alpha_o + x\bar{\gamma}_o + u_1 \\ &= (x_2\Pi_2^* + v_2^*)\alpha_o + x_1\gamma_{o,1} + x_2\gamma_{o,2} + u_1 \\ &= x_2(\Pi_2^*\alpha_o - \gamma_{o,1}\delta_o + \gamma_{o,2}) + u_1^*, \end{aligned} \tag{3.8}$$

where $u_1^* = u_1 + v_2^*\alpha_o + \gamma_{o,1}(x_1 + x_2\delta_o)$. Comparing (3.8) with (3.4),

$$\pi_1^* = \Pi_2^*\alpha_o - \gamma_{o,1}\delta_o + \gamma_{o,2}. \tag{3.9}$$

From (3.9), we see that the identification of the structural parameters α_o, $\gamma_{o,1}$ and $\gamma_{o,2}$ requires restrictions on the structural equation (2.1). Consider the case of exclusion restrictions where some of the exogenous variables do not enter into (2.1). It is convenient to consider separately the two cases of (1) x_1 appearing in (2.1), and (2) x_1 not appearing in (2.1). Consider first the case that x_1 is excluded from (2.1), which is equivalent to saying that $\gamma_{o,1} = 0$. Without loss of generality, suppose that the first k_1 exogenous variables in x_2 are included but the remaining $K - 1 - k_1$ variables in x_2 are excluded from (2.1), i.e., $\gamma_{o,2} = (\gamma_{o,21}', 0)'$ where $\gamma_{o,21}$ is of dimension k_1. Conformably, let $\pi_1^* = (\pi_{11}^{*\prime}, \pi_{12}^{*\prime})'$, $\Pi_2^* = (\Pi_{21}^{*\prime}, \Pi_{22}^{*\prime})'$ and $\delta_o = (\delta_{o,1}', \delta_{o,2}')'$. Since $\gamma_{o,1} = 0$, (3.9) is equivalent to

$$\pi_{11}^* = \Pi_{21}^*\alpha_o + \gamma_{o,21} \tag{3.10}$$

and

$$\pi_{12}^{*} = \Pi_{22}^{*}\alpha_o. \tag{3.11}$$

It follows from (3.10) and (3.11) that a necessary and sufficient condition for the identification of α_o and $\gamma_{o,21}$ is the rank condition that $rank\Pi_{22}^{*} = G_1$. The necessary order condition is $K - 1 - k_1 \geq G_1$, i.e., the number of excluded variables in x_2 from the structural equation is greater than or equal to the number of endogenous variables on the right hand side of (2.1). Consider next the case that x_1 is included in the structural equation. As in the previous case, suppose only the first k_1 exogenous variables in x_2 are included in (2.1). Equation (3.9) is now eqivalent to

$$\pi_{11}^{*} = \Pi_{21}^{*}\alpha_o - \gamma_{o,1}\delta_{o,1} + \gamma_{o,21} \tag{3.12}$$

and

$$\pi_{12}^{*} = \Pi_{22}^{*}\alpha_o - \gamma_{o,1}\delta_{o,2}. \tag{3.13}$$

The rank identification for this case is $rank[\Pi_{22}^{*},\ -\delta_{o,2}] = G_1 + 1$, and the necessary order condition is $K - 1 - k_1 \geq G_1 + 1$. In any event, the identification condition is slightly stronger than the identification condition for the parametric simultaneous equation model. The exact identification of (2.1) for the parametric model becomes underidentification for the semiparametric model. The order identification condition for the semiparametric model corresponds to the overidentification condition in the parametric simultaneous equation model. In summary, the rank identification of the structural parameters in $y = z\alpha_o + xJ\gamma_o + u_1$ from (3.9) is that $[\Pi_2^{*},\ (-\delta_o,\ I)J]$ has full (column) rank.

The identification condition can be extended to the general model where $x\zeta$ represents a vector of indices. To distinguish the indices in $x\zeta$ for identification, each index is required to contain a relevant continuous exogenous variable which does not appear in the other indices (Ichimura and Lee [1988]). Suppose $x\zeta$ contains m indices. With normalization, $x\zeta = (x_{11} + x_2\delta_{(1)}, \cdots, x_{1m} + x_2\delta_{(m)})$, where $x_{11}, \cdots, x_{1m}$ are m distinct continuous variables not contained in the subvector x_2. For this general model, $\delta = [\delta_{(1)}, \cdots, \delta_{(m)}]$ is now a matrix. The rank identification condition is that $[\Pi_2^{*},\ (-\delta_o,\ I)J]$ has full (column) rank.

4. Semiparametric Two-Stage Least Squares Estimation

For the classical simultaneous equation model, the popular IV method is the two-stage least squares method (2LS). A generalization of 2LS to the estimation of parametric simultaneous equation models with selectivity has been introduced in Lee, Maddala and Trost [1980]. The endogenous variables are regressed on all the exogenous variables and a sample selection bias term (inverse Mill's ratio) in the first stage. The regression predictors are then used as the instrumental variables in the second stage estimation of the selection bias corrected structural equation.

For the estimation of the semiparametric model, define the following matrices:

$$\hat{X}_2 = \begin{pmatrix} t_n(x_1\hat{\zeta},\hat{\xi}_n)(x_{21} - E_n(x_2|x_1\hat{\zeta})) \\ \vdots \\ t_n(x_n\hat{\zeta},\hat{\xi}_n)(x_{2n} - E_n(x_2|x_n\hat{\zeta})) \end{pmatrix}, \qquad \hat{W} = \begin{pmatrix} t_n(x_1\hat{\zeta},\hat{\xi}_n)(w_1 - E_n(w|x_1\hat{\zeta})) \\ \vdots \\ t_n(x_n\hat{\zeta},\hat{\xi}_n)(w_n - E_n(w|x_1\hat{\zeta})) \end{pmatrix},$$

and

$$\hat{Y} = \begin{pmatrix} t_n(x_1\hat{\zeta},\hat{\xi}_n)(y_1 - E_n(y|x_1\hat{\zeta})) \\ \vdots \\ t_n(x_n\hat{\zeta},\hat{\xi}_n)(y_n - E_n(y|x_n\hat{\zeta})) \end{pmatrix},$$

where $t_n(x\hat{\zeta},\hat{\xi}_n)$ is a smooth quantile trimming function of $x\hat{\zeta}$ introduced in Appendices A.2 and A.3. A semiparametric two-stage least squares estimator (S2LS) is

$$\hat{\beta}_{S2LS} = [\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{W}]^{-1}\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{Y}. \tag{4.1}$$

This estimator can be interpreted as being derived from a two-stage estimator procedure. In the first stage estimation the reduced form equations for z in (3.7) are estimated by a semiparametric least squares method, and the predicted values for $z - E_n(z|x\hat{\zeta})$ are used as instrumental variables for $z - E_n(z|x\hat{\zeta})$. In addition to z, if x_1 is included in w, an auxiliary reduced form equation for x_1 is also estimated by a semiparametric least squares method, and the predicted value for $x_1 - E_n(x_1|x\hat{\zeta})$ will be used as an instrumental variable for $x_1 - E_n(x_1|x\hat{\zeta})$ for the estimation of (2.7). This two-stage estimator has also a two-stage semiparametric least squares interpretation. To see these, define an auxiliary equation for x_1:

$$x_1 = -x_2\delta_o + v_{1,*}, \tag{4.2}$$

where $v_{1,*} = x_1 + x_2\delta_o$. Since $w^* = (z^*, xJ)$, (3.5) and (4.2) imply that

$$w^* = x_2[\Pi_2^*,\ (-\delta_o,\ I)J] + v^*, \tag{4.2$'$}$$

and conditional on $d^* > 0$ and x,

$$w = x_2[\Pi_2^*,\ (-\delta_o,\ I)J] + E(v^*|x, d^* > 0) + v_d,$$

where $v_d = v^* - E(v^*|x, d^* > 0)$. Similarly, as in (2.7), since the distribution of v^* is a function of $x\zeta_o$,

$$w_i - E_n(w|x_i\hat{\zeta}) = (x_{2i} - E_n(x_2|x_i\hat{\zeta}))\Pi_w + \hat{v}_{ni}, \tag{4.3}$$

where $\Pi_w = [\Pi_2^*,\ (-\delta_o,\ I)J]$. A semiparametric least squares (SLS) estimator of (4.3) with trimming function t_n is

$$\hat{\Pi}_w = (\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{W}. \tag{4.4}$$

It follows from (4.1) and (4.4) that $\hat{\beta}_{S2LS}$ is an IV estimator with $\hat{X}_2\hat{\Pi}_w$ as the instrumental matrix for $\hat{W}$ in the estimation of (2.7). Define the residual matrix $\hat{V} = \hat{W} - \hat{X}_2\hat{\Pi}_w$. Since $\hat{V} = [I - \hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2']\hat{W}$, it follows that $\hat{V}'\hat{X}_2\hat{\Pi}_w = 0$, and $\hat{\beta}_{S2LS} = [(\hat{X}_2\hat{\Pi}_w)'(\hat{X}_2\hat{\Pi}_w)]^{-1}(\hat{X}_2\hat{\Pi}_w)'\hat{Y}$ which has a two-stage least squares interpretation. The regressor x_1 plays an interesting role in the estimation. This variable is exogenous in the equation system, however it behaves as if it were an endogenous variable in the estimation. It has been excluded from the list of regressors in the first stage SLS estimation. This feature is compatible with the order identification condition in section 3.

Asymptotic properties of the S2LS are derived in Appendix A.5 in detail. Consistency of the S2LS requires the rank identification condition. Since sample observations for (2.4) are available only after selection, all expectations will be taken as conditional expectations conditional on $d_i^* > 0$. To simplify notations, the conditional argument $d^* > 0$ will be suppressed. Thus $E(\cdot|x_i\zeta_o, \delta_o)$ stands for $E(\cdot|x_i\zeta_o, d_i^* > 0, \delta_o)$ in subsequent presentations and the appendix. As shown in (A.5.1) of the appendix,

$$\frac{1}{n}\hat{X}_2'\hat{X}_2 \xrightarrow{p} C, \tag{4.5}$$

where

$$C = E[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o, \delta_o))'(x_2 - E(x_2|x\zeta_o, \delta_o))], \tag{4.6}$$

and I_T is the indicator function of T, where $T = [\xi_p(\delta_o),\ \xi_{(1-p)}(\delta_o)]$ with $\xi_p(\delta_o)$ and $\xi_{(1-p)}(\delta_o)$ being respectively the pth and $(1-p)$th quantiles of $x\zeta_o$. On the other hand, from (A.5.2),

$$\frac{1}{n}\hat{W}'\hat{X}_2 \xrightarrow{p} \Pi_w' C. \tag{4.7}$$

Therefore, under the assumption that C is nonsingular,

$$\frac{1}{n}\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{W} \xrightarrow{p} \Pi_w' C \Pi_w. \tag{4.8}$$

The rank identification condition that Π_w has full column rank is necessary for the limiting matrix in (4.8) to be nonsingular. Equations (2.7) and (4.1) imply that

$$\hat{\beta}_{S2LS} - \beta_o = [\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{W}]^{-1}\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{U}_n, \tag{4.9}$$

where $\hat{U}_n = (\hat{u}_{n1}, \cdots, \hat{u}_{nn})'$. Since

$$\begin{aligned}\text{plim}_{n\to\infty}\frac{1}{n}\hat{X}_2'\hat{U}_n &= E(I_T(x\zeta_o)[z_1 - E(z_1|x\zeta_o,\delta_o)][u_1 - E(u_1|x\zeta_o,\delta_o)]) \\ &= 0,\end{aligned}$$

$\hat{\beta}_{S2LS}$ is a consistent estimator of β_o.

Given that $\sqrt{n}(\hat{\delta} - \delta_o) \xrightarrow{D} N(0, V_\delta)$, the asymptotic distribution of $\hat{\beta}_{S2LS}$ can be derived from (4.9). Let ξ_o denote $(\xi_p(\delta_o),\ \xi_{(1-p)}(\delta_o))$. As shown in (A.5.8) of the appendix,

$$\begin{aligned}\frac{1}{\sqrt{n}}\hat{X}_2'\hat{U}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)(x_{2i} - E(x_2|x_i\zeta_o,\delta_o))'(u_{1i} - E(u_1|x_i\zeta_o,\delta_o)) \\ &\quad - E\left(I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))'\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right)\sqrt{n}(\hat{\delta} - \delta_o) + o_p(1) \\ &\xrightarrow{D} N(0,\Delta),\end{aligned} \tag{4.10}$$

where

$$\Delta = E[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))'\omega(x\zeta_o,\delta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))] + DV_\delta D', \tag{4.11}$$

and

$$D = E[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))'\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}]. \tag{4.12}$$

It follows that

$$\sqrt{n}(\hat{\beta}_{S2LS} - \beta_o) \xrightarrow{D} N(0,\Omega), \tag{4.13}$$

where

$$\Omega = (\Pi_w' C\Pi_w)^{-1}\Pi_w'\Delta\Pi_w(\Pi_w' C\Pi_w)^{-1}. \tag{4.14}$$

5. Semiparametric Generalized Two-Stage Least Squares Estimation

The S2LS in (4.1) is simple, but it is not an efficient IV estimator for the estimation of (2.7) because it has not incorporated the complicated covariances structure of $\hat{u}_{ni}$ in estimation.

The disturbance $\hat{u}_{ni}$ in (2.7) can be decomposed into three components:

$$\begin{aligned}\hat{u}_{ni} &= u_{1i} - E_n(u_1|x_i\hat{\zeta}) \\ &= (u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) - (E_n(u_1|x_i\hat{\zeta}) - E_n(u_1|x_i\zeta_o)) - (E_n(u_1|x_i\zeta_o) - E(u_1|x_i\zeta_o, \delta_o)).\end{aligned} \tag{5.1}$$

The first component represents the disturbance u_{di} in the structural equation (2.4) after the correction of selection bias. The second component represents the disturbance introduced in $E_n(u_1|x_i\zeta_o)$ by replacing ζ_o by the estimate $\hat{\zeta}$. In the parametric two-stage estimation of the sample selection model with a discrete choice decision rule, these two components of the disturbance are asymptotically uncorrelated (Heckman [1979] and Lee, Maddala and Trost [1980]).[2] This is also the case for the semiparametric model in (2.1) and (2.3) as shown in the appendix. The last component represents the error introduced by the nonparametric estimate of the conditional expectation of u_{1i}. Even though the last component has a rather complicated structure, it does not influence the asymptotic distribution of $\hat{\beta}_{S2LS}$, due to an asymptotic orthogonality property of the index model (see the appendix for detail, in particular, Propositions 4 and 6 of Appendix A.4).

The variance of u_{di} in (2.4) is a function of $x_i\zeta_o$. It can be estimated nonparametrically. Let H be a kernel function with a bandwidth b_n.[3] Define

$$\omega_{ni}(\hat{\beta}, \hat{\delta}) = \frac{\sum_{j\neq i}^{n}(y_j - w_j\hat{\beta})^2 H(\frac{x_i\hat{\zeta} - x_j\hat{\zeta}}{b_n})}{\sum_{j\neq i}^{n} H(\frac{x_i\hat{\zeta} - x_j\hat{\zeta}}{b_n})} - \left[\frac{\sum_{j\neq i}^{n}(y_j - w_j\hat{\beta}) H(\frac{x_i\hat{\zeta} - x_j\hat{\zeta}}{b_n})}{\sum_{j\neq i}^{n} H(\frac{x_i\hat{\zeta} - x_j\hat{\zeta}}{b_n})}\right]^2, \tag{5.2}$$

where $\hat{\beta}$ is an initial consistent estimate of β (e.g., $\hat{\beta}_{S2LS}$ in (4.1)). $\hat{\omega}_n(\hat{\beta}, \hat{\delta})$ provides a nonparametric estimate of the variance $\omega(x_i\zeta_o, \delta_o)$ of u_{di}^2. The above arguments suggest the following covariance matrix:

$$\hat{\Sigma} = \hat{\Lambda}_n + \frac{\partial E_n(\hat{\beta}, \hat{\zeta})}{\partial \delta'} \hat{V}_{n,\delta} \frac{\partial E_n'(\hat{\beta}, \hat{\zeta})}{\partial \delta}, \tag{5.3}$$

[2] For the sample selection model with a tobit type decision rule, if the first-stage estimate is a tobit MLE, the two components can be correlated (see Lee et al [1980]).

[3] The kernel K with a_n in (2.6) can be used. However, it is desirable to use a separate kernel function so as to avoid unnecessary stronger requirements on the rate of convergence for bandwidth parameters.

where $\hat{\Lambda}_n$ is a diagonal matrix with diagonal elements $\hat{\omega}_{ni}(\hat{\beta}, \hat{\delta})$, $i = 1, \cdots, n$; $\hat{V}_{n,\delta}$ is a consistent estimate of the asymptotic covariance matrix V_δ of $\hat{\delta}$; and

$$\frac{\partial E_n'(\hat{\beta}, \hat{\zeta})}{\partial \delta} = \left[t_n(x_1\hat{\zeta}, \hat{\xi}_n) \frac{\partial E_n(y - w\hat{\beta}|x_1\hat{\zeta})}{\partial \delta}, \cdots, t_n(x_n\hat{\zeta}, \hat{\xi}_n) \frac{\partial E_n(y - w\hat{\beta}|x_n\hat{\zeta})}{\partial \delta} \right]. \tag{5.4}$$

By a formula of inversion of a partitioned matrix, the inverse of $\hat{V}_{n,\delta}$ involves only inversion of matrices of the dimension of δ:

$$\hat{\Sigma}^{-1} = \hat{\Lambda}_n^{-1} - \hat{\Lambda}_n^{-1} \frac{\partial E_n(\hat{\beta}, \hat{\zeta})}{\partial \delta'} (\hat{V}_{n,\delta}^{-1} + \frac{\partial E_n'(\hat{\beta}, \hat{\zeta})}{\partial \delta} \hat{\Lambda}_n^{-1} \frac{\partial E_n(\hat{\beta}, \hat{\zeta})}{\partial \delta'})^{-1} \frac{\partial E_n'(\hat{\beta}, \hat{\zeta})}{\partial \delta} \hat{\Lambda}_n^{-1}. \tag{5.5}$$

For the parametric simultaneous equation sample selection model, several generalized two-stage least squares estimators (G2LS) have been introduced. For the semiparametric model, the following estimator is a semiparametric generalized two-stage least squares estimator (SG2LS):

$$\hat{\beta}_{SG} = [\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{W}]^{-1}\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{Y}. \tag{5.6}$$

An alternative SG2LS estimator is

$$\tilde{\beta}_{SG} = [\hat{W}'\hat{\Sigma}^{-1}\hat{X}_2(\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{W}]^{-1}\hat{W}'\hat{\Sigma}^{-1}\hat{X}_2(\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{Y}. \tag{5.7}$$

It can be shown that these two estimators have the same asymptotic distribution.[4] The computation of $\hat{\beta}_{SG}$ is simpler, but the data transformations in $\tilde{\beta}_{SG}$ are intuitively appealing.

Substituting (2.7) into (5.6) and (5.7),

$$\hat{\beta}_{SG} - \beta_o = [\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{W}]^{-1}\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{U}_n, \tag{5.8}$$

and

$$\tilde{\beta}_{SG} - \beta_o = [\hat{W}'\hat{\Sigma}^{-1}\hat{X}_2(\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{W}]^{-1}\hat{W}'\hat{\Sigma}^{-1}\hat{X}_2(\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{U}_n. \tag{5.9}$$

With (5.5), as shown in (A.5.3)-(A.5.7) and (A.5.9),

$$\frac{1}{n}\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2 \xrightarrow{p} \Gamma, \tag{5.10}$$

[4] For the parametric sample selection model, the asymptotic equivalency between two such similar estimators has been shown in Lee [1981] and Amemiya [1983].

$$\frac{1}{n}\hat{X}_2'\hat{\Sigma}^{-1}\hat{W} \xrightarrow{P} \Gamma\Pi_w, \tag{5.11}$$

and

$$\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\Sigma}^{-1}\hat{U}_n \xrightarrow{D} N(0,\ \Gamma), \tag{5.12}$$

where

$$\begin{aligned}\Gamma =& E[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))\omega^{-1}(x\zeta_o,\delta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))] \\ &- E[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))\omega^{-1}(x\zeta_o,\delta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}] \\ &\cdot \left(V_\delta^{-1} + E\left[I_T(x\zeta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta}\omega^{-1}(x\zeta_o,\delta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right]\right)^{-1} \\ &\cdot E[I_T(x\zeta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta}\omega^{-1}(x\zeta_o,\delta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))].\end{aligned} \tag{5.13}$$

Equations (5.10)-(5.12) imply that

$$\sqrt{n}(\tilde{\beta}_{SG} - \beta_o) \xrightarrow{D} N(0,\ (\Pi_w'\Gamma\Pi_w)^{-1}). \tag{5.14}$$

From (4.5), (4.7), (5.10) and (5.11), both $\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}$ and $\hat{W}'\hat{\Sigma}^{-1}\hat{X}_2(\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2)^{-1}$ converge in probability to Π_w'. Hence $\sqrt{n}(\hat{\beta}_{SG} - \beta_o)$ has the same limiting distribution of $\sqrt{n}(\tilde{\beta}_{SG} - \beta_o)$. The asymptotic covariance matrix of $\hat{\beta}_{SG}$ (or $\tilde{\beta}_{SG}$) can be consistently estimated by

$$\hat{\Omega}_{G,n} = [\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{W}]^{-1} \tag{5.15}$$

or

$$\tilde{\Omega}_{G,n} = [\hat{W}'\hat{\Sigma}^{-1}\hat{X}_2(\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{W}]^{-1}. \tag{5.16}$$

Let $\hat{Q} = \hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}'_2\hat{W}$. By the generalized Cauchy-Schwartz inequality,

$$\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2 \geq \hat{X}_2'\hat{Q}(\hat{Q}'\hat{\Sigma}\hat{Q})^{-1}\hat{Q}'\hat{X}_2.$$

Hence

$$\begin{aligned}\Pi_w'\Gamma\Pi_w &= \Pi_w'\text{plim}\frac{1}{n}\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2\Pi_w \\ &\geq \Pi_w'\text{plim}\frac{1}{n}\hat{X}_2'\hat{Q}(\hat{Q}'\hat{\Sigma}\hat{Q})^{-1}\hat{Q}'\hat{X}_2\Pi_w \\ &= \Pi_w'\text{plim}\frac{1}{n}\hat{X}_2'\hat{W}(\Pi_w'\hat{X}_2'\hat{\Sigma}\hat{X}_2\Pi_w)^{-1}\hat{W}'\hat{X}_2\Pi_w \\ &= \Omega^{-1}\end{aligned} \tag{5.17}$$

by (4.5), (4.7) and (4.14). Hence $\tilde{\beta}_{SG}$ is asymptotically efficient relative to $\hat{\beta}_{S2LS}$.

The SG2LS estimators $\hat{\beta}_{SG}$ and $\tilde{\beta}_{SG}$ are not only asymptotically efficient relative to the S2LS estimator $\hat{\beta}_{S2LS}$. They have also an asymptotically optimal property. They are asymptotically efficient IV estimators for the estimation of Equation (2.7). Let p_i be an instrumental variable for w_i and $\hat{\beta}_p$ be the IV estimator:

$$\hat{\beta}_p = (P'\hat{W})^{-1}P'\hat{Y}, \tag{5.18}$$

where

$$P = [t_n(x_1\hat{\zeta}, \hat{\xi}_n)p_1', \cdots, t_n(x_n\hat{\zeta}, \hat{\xi}_n)p_n']'. \tag{5.19}$$

Equation (2.7) implies that

$$\hat{\beta}_p - \beta_o = (P'\hat{W})^{-1}P'\hat{U}_n. \tag{5.20}$$

For this case, the component $E_n(u_1|x_i\zeta_o) - E(u_1|x_i\zeta_o, \delta_o)$ in $\hat{u}_{ni}$ has a significant impact on the asymptotic distribution of $\hat{\beta}_p$. From (A.5.11) of the appendix,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o)p_i'(u_{1i} - E_n(u_1|x_i\zeta_o)) \overset{D}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o)(p_i - E_n(p|x_i\zeta_o))(u_{1i} - E_n(u_1|x_i\zeta_o)), \tag{5.21}$$

where $\overset{D}{=}$ means that both sides have the same limiting distribution. Similar to (4.7),

$$\begin{aligned}\frac{1}{n}P'\hat{W} &\xrightarrow{p} E(I_T(x\zeta_o)p'(x_2 - E(x_2|x\zeta_o, \delta_o))\Pi_w \\ &= E(I_T(x\zeta_o)(p - E(p|x\zeta_o, \delta_o))'(x_2 - E(x_2|x\zeta_o, \delta_o))\Pi_w.\end{aligned} \tag{5.22}$$

The latter equality in (5.22) holds because $I_T(x\zeta_o)$ is a function of $x\zeta_o$ alone. From (A.5.10) and (A.5.11) in Appendix A.5,

$$\frac{1}{\sqrt{n}}P'\hat{U}_n \xrightarrow{D} N(0, \ \Delta_p), \tag{5.23}$$

where

$$\begin{aligned}\Delta_p =& E[I_T(x\zeta_o)(p - E(p|x\zeta_o, \delta_o))'\omega(x\zeta_o, \delta_o)(p - E(p|x\zeta_o, \delta_o))] \\ &+ E[I_T(x\zeta_o)(p - E(p|x\zeta_o, \delta_o))'(x_2 - E(x_2|x\zeta_o, \delta_o))\frac{\partial E(u_1|x\zeta_o, \delta_o)}{\partial t}] \\ &\cdot V_\delta E[I_T(x\zeta_o)(p - E(p|x\zeta_o, \delta_o))'(x_2 - E(x_2|x\zeta_o, \delta_o)]'\frac{\partial E(u_1|x\zeta_o, \delta_o)}{\partial t}.\end{aligned} \tag{5.24}$$

It follows that

$$\sqrt{n}(\hat{\beta}_p - \beta) \xrightarrow{D} N(0, \ \Omega_p), \tag{5.25}$$

where

$$\begin{aligned}\Omega_p = & \{E(I_T(x\zeta_o)(p - E(p|x\zeta_o, \delta_o))'(x_2 - E(x_2|x\zeta_o, \delta_o))\Pi_w\}^{-1} \\ &\cdot \Delta_p\{E(I_T(x\zeta_o)(p - E(p|x\zeta_o, \delta_o))(x_2 - E(x_2|x\zeta_o, \delta_o))'\Pi_w\}^{-1'}.\end{aligned} \tag{5.26}$$

From (5.22), (5.25) and (5.26), we note that $\hat{\beta}_p$ has the same limiting distribution of the following IV estimator:

$$\tilde{\beta}_p = (\hat{P}'\hat{W})^{-1}\hat{P}'\hat{Y}, \tag{5.27}$$

where

$$\hat{P} = \begin{pmatrix} t_n(x_1\zeta_o, \xi_o)(p_1 - E_n(p|x_1\hat{\zeta}, \hat{\delta})) \\ \vdots \\ t_n(x_n\zeta_o, \xi_o)(p_n - E_n(p|x_n\hat{\zeta}, \hat{\delta})) \end{pmatrix}. \tag{5.28}$$

This indicates that for the estimation of the semiparametric simultaneous equation model, the residuals $p_i - E_n(p|x_i\zeta_o, \delta_o)$ and $p - E(p|x\zeta_o, \delta_o)$ play the crucial role rather than the variable p itself. The asymptotic covariance matrix of $\tilde{\beta}_p$ (or $\hat{\beta}_p$) can be consistently estimated by

$$\hat{\Omega}_{p,n} = [\hat{P}'\hat{W}]^{-1}\hat{P}'\hat{\Sigma}\hat{P}[\hat{W}'\hat{P}]^{-1}. \tag{5.29}$$

The asymptotic efficiency of $\tilde{\beta}_{SG}$ relative to $\tilde{\beta}_p$ follows from the following inequality:

$$\begin{aligned} \Pi_w'\Gamma\Pi_w &= \Pi_w' \text{plim}\frac{1}{n}\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2\Pi_w \\ &\geq \Pi_w' \text{plim}\frac{1}{n}\hat{X}_2'\hat{P}(\hat{P}'\hat{\Sigma}\hat{P})^{-1}\hat{P}'\hat{X}_2\Pi_w \\ &= \text{plim}\frac{1}{n}\hat{W}'\hat{P}(\hat{P}'\hat{\Sigma}\hat{P})^{-1}\hat{P}'\hat{W} \\ &= \Omega_p^{-1}. \end{aligned}$$

6. Conclusion

In this article, we have considered the identification and estimation of the sample selection simultaneous equation model without a parametric distributional assumption. Based only on index restrictions, the identification of structural equation parameters from reduced form parameters requires stronger exclusion restrictions than the identification of structural equations in the classical simultaneous equation model. The identification in this semiparametric model requires the underlying structural equation to be overidentified in the classical sense. Exact identification in the classical model becomes underidentification for the semiparametric model. Estimation of the structural equation by instrumental variable methods has been considered. Some two-stage estimation procedures which generalize the estimation procedures for the parametric model and the classical two-stage least squares method are introduced. Consistency and asymptotic normality of the estimators are proved. An asymptotically efficient instrumental variable method has also been derived. Some interesting features of the instrumental variable estimation in this model have been discovered. Residuals of the variables in the model which are derived from the projection of variables to the selection equation index play the crucial role for asymptotic properties of the estimators. Exogenous variables in the system used in normalization for the selection equation indices behave as if they were endogenous variables in two-stage estimation procedures. The latter feature is compatible with the identification condition for the model.

In this article, we have considered only single equation estimation methods. These methods can be generalized to the estimation of system equations by some semiparametric three-stage procedures. For the semiparametric sample selection model (without simultaneity), an asymptotically efficient estimator, in the sense that its asymptotic covariance matrix attains the semiparametric lower bound in Chamberlain [1986], has been derived in Lee [1990]. Such an estimator is a semiparametric maximum likelihood estimator. With the reduced form equations estimated by such a method, the structural parameters may then be estimated by Amemiya's minimum distance procedure (Amemiya [1978] and [1983]). One might conjecture that such a structural estimator could be an asymptotically efficient semiparametric estimator. Any any rate, such an estimation method is not a simple instrumental variable method.

Appendix

A.1 Model Assumptions

Assumption 1:

(1) The samples s_i, where $s = (y, z, x)$, $i = 1, \cdots, n$ are i.i.d. x is the vector consisting of all exogenous variables in the equation system. The moments of order $3 \cdot r$, where $r \geq 2$, of s exist.

(2) The parameter space Θ of δ is a compact subset of a finite dimensional Euclidean space, and δ_o is in the interior of Θ.

(3) x_1 is an m-dimensional vector of continuous variables.

(4) $\hat{\delta}$ is a $\sqrt{n}$-consistent estimator, $\sqrt{n}(\hat{\delta} - \delta_o)$ is asymptotically normal, $N(0, V_\delta)$, where V_δ is a positive definite matrix, and is aymptotically uncorrelated with $u_{1i} - E(u_1 | x_i \zeta_o, \delta_o)$ for all i.

(5) The matrix Π_w in (4.3) has full column rank.

(6) The matrices C in (4.6) and Γ in (5.13) are nonsingular.

Assumption 2:

(1) $K(v)$ on R^m is a kernel function with a bandwidth parameter a_n defined on a bounded support D,[5] i.e., $\int_D K(v)dv = 1$, and $\lim_{n\to\infty} a_n = 0$.

(2) $K(v)$ is twice differentiable and its second order derivatives satisfy a Lipschitz condition of order 1.[6]

(3) $K(v)$ is a higher order kernel function with zero moments up to the order s^*, $s^* = m + 2$, i.e.,

$$\int_D v_1^{i_1} \cdots v_m^{i_m} K(v)dv = 0$$

for all $0 \leq i_l, l = 1, \cdots, m$, and $1 \leq i_1 + \cdots + i_m < s^*$.

(4) $\{a_n\}$ is chosen with a rate such that $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{(1+6/r)(m+2)+2} = \infty$, and $\lim_{n\to\infty} n a_n^{2(m+2)} = \infty$, but $\lim_{n\to\infty} n a_n^{4s^*} = 0$.

[5] The boundedness of D is inessential. Relaxing this assumption will make our analysis relatively more complicated. In practice, kernel functions with bounded support are simpler to compute.

[6] A function $h(x)$ is said to satisfy a Lipschitz condition if there exists a constant c such that $\| h(x_1) - h(x_2) \| \leq c \| x_1 - x_2 \|$ for all x_1 and x_2.

Assumption 3:

(1) The density function $p(t|\delta)$ of $t = x\zeta$ in R^m is positive everywhere for each $\delta \in \Theta$. It is differentiable everywhere with t to the order s^*, and these derivatives are continuous at (t, δ) everywhere.

(2) $E(s \otimes \bar{x}|t, \delta)p(t|\theta)$, where $s = (y, z, x)$ and $\bar{x} = (1, x)$, is differentiable everywhere with t to the order $s^* + 1$, and these derivatives are continuous at (t, δ) everywhere.[7]

(3) $E(\|s \otimes \bar{x} \otimes \bar{x}\|^2|t, \delta)p(t|\delta)$, where $s = (1, y, z, x)$ and $\bar{x} = (1, x)$, is continuous at (t, δ) everywhere.

(4) $E(s \otimes \bar{x} \otimes \bar{x}|t, \delta)p(t|\delta)$, where $s = (1, y, z)$ and $\bar{x} = (1, x)$, is twice differentiable with t and its second order derivatives are continuous at (t, δ) everywhere.

Assumption 4:

(1) $H(v)$ on R^m is a kernel function with a bandwidth parameter b_n defined on a bounded support.

(2) $H(v)$ is differentiable and its derivative satisfies a Lipschitz condition of order 1.

(3) $H(v)$ is a kernel function with zero moments up to the order h^*.

(4) $\{b_n\}$ is chosen such that $\lim_{n\to\infty} \frac{n}{\ln n} b_n^{(1+6/r)(m+1)+1} = \infty$, and $\lim_{n\to\infty} n b_n^{2m} = \infty$, but $\lim_{n\to\infty} n b_n^{4h^*} = 0$.

Assumption 5:

(1) $E(s \otimes s|t, \delta)p(t|\theta)$, where $s = (1, y, z, x)$, of $t = x\zeta$ is differentiable everywhere to the order h^*, and these derivatives are continuous at (t, δ) everywhere.

(2) $E(\|s \otimes s\|^2|t, \delta)p(t|\delta)$, where $s = (1, y, z, x)$, is continuous at (t, δ) everywhere.

(3) $E(s \otimes s \otimes \bar{x}|t, \delta)p(t|\theta)$, where $s = (1, y, z, x)$ and $\bar{x} = (1, x)$, is differentiable everywhere with t and this derivative is continuous at (t, δ) everywhere.

(4) $E(\|s \otimes s \otimes \bar{x}\|^2|t, \delta)p(t|\delta)$, where $s = (1, y, z, x)$ and $\bar{x} = (1, x)$, is continuous at (t, δ) everywhere.

Assumption 1(4) for the first stage estimator will be satisfied with parametric or semiparametric estimators of discrete choice models such as the probit or logit estimators (under correct distributional assumptions), the Ichimura single index estimator (Ichimura [1987]), and the semiparametric maximum likelihood estimator of Klein and Spady [1987] for binary choice models, and the multiple index estimator in Ichimura and

[7] $\otimes$ denotes the Kronecker product.

Lee [1987] for polytomous choice models. In general, for a choice model with L alternatives, if

$$\sqrt{n}(\hat{\delta} - \delta_o) = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} f(x_i, I_{1,i}, \cdots, I_{L,i}) + o_p(1),$$

where I_l is a dichotomous choice indicator for the l alternative and T is the sample size of the choice model, is asymptotically normal, Assumption 1(4) will be satisfied.[8] This is so, because

$$\begin{aligned} & E^*\{\sum_{j=1}^{T} f(x_j, I_{1,j}, \cdots, I_{L,j})(u_{1i} - E(u_1|x_i\zeta_o, \delta_o))\} \\ &= E^*\{f(x_i, I_{1,i}, \cdots, I_{L,i})(u_{1i} - E(u_1|x_i\zeta_o, \delta_o))\} \\ &= E^*\{E[u_{1i} - E(u_1|x_i\zeta_o, \delta_o)|x_i, I_{1i} = 1]f(x_i, I_{1,i}, \cdots, I_{L,i})\} \\ &= 0, \end{aligned}$$

where E^* denotes the unconditional expectation taken with the whole population. Assumption 1(5) is the rank identification condition, and Assumption 1(6) is for the limiting distributions of the S2LS and SG2LS estimators to be well-defined with the $\sqrt{n}$ rate of convergence.

The kernel function K with a bounded support in Assumption 2 has implicitly the following properties: $\int_D |K(v)|dv$ and $\int_D ||v||^{s^*}|K(v)|dv$ are finite; $K(v)$ and its first order derivatives are bounded; and $K(v)$ and $\frac{\partial K(v)}{\partial v}$ go to zero at their boundary.

The conditions in Assumptions 3 and 5 can be justified by some basic regularity conditions on the distributions of the variables in the models. However, the above assumptions are more direct. As an illustration, let $f(x_1|x_2)$ be the density function of x_1 conditional on x_2. Since $t = x_1 + x_2\delta$, where $t = x\zeta$, $x_1 = (x_{11}, \cdots, x_{1m})$, and $\delta = (\delta_{(1)}, \cdots, \delta_{(m)})$,

$$p(t|\delta) = \int f(t - x_2\delta|x_2)d\nu(x_2), \tag{A.1.1}$$

where $\nu(x_2)$ is the distribution measure of x_2, under the assumption that the support of x_1 is R^m. If $f(x_1|x_2)$ were continuous and bounded, the bounded convergence theorem will imply that $p(t|\delta)$ is a continuous function. The continuity and boundedness properties in Assumptions 3 and 5 are used to guarantee the stochastic convergence and to control asymptotic biases of nonparametric kernel estimates. See Lemmas 1 to 4 in Appendix A.4. Assumptions 4 and 5 are needed only for the nonparametric estimates ω_{ni} of the variance of u_{1i}. For the asymptotic properties of the S2LS estimator, these two assumptions are not needed.

[8] T refers to the sample size of the choice model where $T > n$. Implicitly, we assume that the fraction $\frac{n}{T}$ will converge asymptotically to a fixed positive number for stability.

A.2 TRIMMING INDEX

To trim the tails of the index $x\hat{\zeta}$, we can use some quantile statistics of $x\hat{\zeta}$. Without loss of generality, consider a single index. Let $0 < p < \frac{1}{2}$ be a specified order of quantile. The first-stage estimate $\hat{\delta}$ in $x\hat{\zeta} = x_1 + x_2\hat{\delta}$ can be discretized by LeCam's device. Let $\{\hat{\delta}_n\}$ be the sequence of the first-stage estimate of δ_o. Let $\| \delta \| = \max_{l=1,\cdots,k} |\delta_l|$, where $\delta = (\delta_1, \cdots, \delta_k)'$ is the norm of δ in the k-dimensional Euclidean space R^k. Let $R^k_n = \{\frac{1}{\sqrt{n}}(i_1, \cdots, i_k)|i_1, \cdots, i_k$ are integers$\}$ and let $\bar{\delta}_n$ be a point in R^k_n closest to $\hat{\delta}_n$ under $\| \cdot \|$. The $\{\bar{\delta}_n\}$ is a discretized sequence of estimates of δ_o. Let $\hat{\xi}_{np}$ and $\hat{\xi}_{n(1-p)}$ be respectively the pth sample quantile and the $(1-p)$th sample quantile of the observations of $x_i\bar{\zeta}$, $i = 1, \cdots, n$, where $x\bar{\zeta}_n = x_1 + x_2\bar{\delta}_n$. Observations of $x_i\hat{\zeta}$ will be trimmed whenever their values lie outside T_n, where $T_n = [\hat{\xi}_{np},\ \hat{\xi}_{n(1-p)}]$.

The discretization device provides some technical simplification for our asymptotic analysis. $\hat{\xi}_{np}$ and $\hat{\xi}_{n(1-p)}$ can be shown to be $\sqrt{n}$-consistent. Since $\hat{\delta}_n$ is $\sqrt{n}$-consistent and $\| \hat{\delta}_n - \bar{\delta}_n \| \leq \frac{1}{\sqrt{n}}$,

$$\begin{aligned}\sqrt{n}(\bar{\delta}_n - \delta_o) &= \sqrt{n}(\bar{\delta}_n - \hat{\delta}_n) + \sqrt{n}(\hat{\delta}_n - \delta_o) \\ &= O_p(1),\end{aligned}$$

i.e., $\bar{\delta}_n$ is also $\sqrt{n}$-consistent. For any fixed parameter vector δ, the samples $x_i\zeta$ are i.i.d. Let $\xi_p(\delta)$ be the pth quantile of $x\zeta$ and $\hat{\xi}_{np}(\delta)$ be the corresponding pth sample quantile of $x_i\zeta$, $i = 1, \cdots, n$. From Theorem 2.2.1 in Serfling [1980], we know that for any $\epsilon > 0$,

$$P(|\hat{\xi}_{np}(\delta) - \xi_p(\delta)| > \epsilon) \leq 2\exp(-n\delta_\epsilon^2), \qquad n \geq 1, \tag{A.2.1}$$

where $\delta_\epsilon = \min[F_\delta(\xi_p(\delta)+\epsilon)-p,\ p-F_\delta(\xi_p(\delta)-\epsilon)]$ and F_δ denotes the distribution of $x\zeta$.[9] Under Assumption 3(1), the density function $p(t|\delta)$ of $x\zeta$ is positive everywhere, and hence is bounded away from zero on any compact neighborhood $N(\xi_p(\delta_o)) \times N(\delta_o)$ of $(\xi_p(\delta_o),\ \delta_o)$. With (A.2.1), we can show that for any sequence $\{\delta_n\}$ such that $\lim_{n\to\infty} \delta_n = \delta_o$,

$$\sqrt{n}|\hat{\xi}_{np}(\delta_n) - \xi_p(\delta_n)| = O_p(1). \tag{A.2.2}$$

Let $\epsilon_n = \frac{c}{\sqrt{n}}$, where c is an arbitrary constant. By the mean value theorem, $F_{\delta_n}(\xi_p(\delta_n) + \epsilon_n) - p = p(\xi_p(\delta_n) + \lambda_n\epsilon_n|\delta_n)\epsilon_n$ for some λ_n, $0 \leq \lambda_n \leq 1$. By the continuity of $\xi_p(\delta)$ at $\delta = \delta_o$, $(\xi_p(\delta_n) + \lambda_n\epsilon_n,\ \delta_n) \in N(\xi_p(\delta_o)) \times N(\delta_o)$ for sufficiently large n. It follows that for large n, $F_{\delta_n}(\xi_p(\delta_n) + \epsilon_n) - p \geq b\epsilon_n$, where

[9] This exponential bound for the sample quantile follows from Hoeffding's inequality.

$b = \inf_{(t,\delta) \in N(\xi_p(\delta_o)) \times N(\delta_o)} p(t|\delta)$. Similarly, $p - F_{\delta_n}(\xi_p(\delta_n) - \epsilon_n) \geq b\epsilon_n$. Therefore, for large n, $\exp(-n\delta^2_{\epsilon_n}) \leq \exp(-nb^2\epsilon_n^2) = \exp(-b^2c^2)$, and

$$P(\sqrt{n}|\hat{\xi}_{np}(\delta_n) - \xi_p(\delta_n)| > c) \leq 2\exp(-b^2c^2). \tag{A.2.3}$$

Since c is arbitrary, it follows that $\sqrt{n}|\hat{\xi}_{np}(\delta_n) - \xi_p(\delta_n)| = O_p(1)$ for any sequence $\{\delta_n\}$ which converges to δ_o. For any finite constant $M > 0$, define $\Delta_{M,n} = \{\delta | \parallel \delta - \delta_o \parallel < \frac{M}{\sqrt{n}} \text{ and } \delta \in R^k_n\}$. $\Delta_{M,n}$ has the interesting property that its cardinality is finite and bounded, say by $\bar{M}$, independently of n. This is so since the cardinality of $\{(i_1, \cdots, i_k) | i_1, \cdots, i_k \text{ are integers and } |i_l - \sqrt{n}\delta_{o,l}| < M,\ l = 1, \cdots, k\}$ is finite and bounded, independently of n. It is obvious that for any sequence $\{\delta_n\}$ with $\delta_n \in \Delta_{M,n}$, it converges to δ_o. The finiteness of $\Delta_{M,n}$ and (A.2.3) imply that

$$P(\sup_{\delta_n \in \Delta_{M,n}} \sqrt{n}|\hat{\xi}_{np}(\delta_n) - \xi_p(\delta_n)| > c) \leq 2\bar{M}\exp(-b^2c^2) \tag{A.2.4}$$

for large n, and therefore

$$\sup_{\delta_n \in \Delta_{M,n}} \sqrt{n}|\hat{\xi}_{np}(\delta_n) - \xi_p(\delta_n)| = O_P(1).$$

Since $\bar{\delta}_n$ is in R^k_n and is a $\sqrt{n}$-consistent estimate of δ_o, $\bar{\delta}_n$ will lie in $\Delta_{M,n}$ with probability close to one for large n. Hence

$$\sqrt{n}|\hat{\xi}_{np}(\bar{\delta}_n) - \xi_p(\bar{\delta}_n)| = O_P(1). \tag{A.2.5}$$

By the mean value theorem, $\xi_p(\bar{\delta}_n) = \xi_p(\delta_o) + \frac{\partial \xi_p(\delta_n^*)}{\partial \delta'}(\bar{\delta}_n - \delta_o)$. Since $p(t|\delta)$ is continuous, $\frac{\partial \xi_p(\delta)}{\partial \delta}$ is continuous at δ_o, and $\sqrt{n}(\bar{\delta}_n - \delta_o) = O_p(1)$, it follows that

$$\begin{aligned} \sqrt{n}|\hat{\xi}_{np}(\bar{\delta}_n) - \xi_p(\delta_o)| &\leq \sqrt{n}|\hat{\xi}_{np}(\bar{\delta}_n) - \xi_p(\bar{\delta}_n)| + \parallel \frac{\partial \xi_p(\delta_n^*)}{\partial \delta'} \parallel \cdot \sqrt{n}|\bar{\delta}_n - \delta_o| \\ &= O_p(1), \end{aligned} \tag{A.2.6}$$

i.e., $\hat{\xi}_{np}$ is a $\sqrt{n}$-consistent estimate of $\xi_p(\delta_o)$. Similarly, $\hat{\xi}_{n(1-p)}$ is a $\sqrt{n}$-consistent estimate of $\xi_{(1-p)}(\delta_o)$.

A.3 Smooth Trimming

The trimming can be smoothed by down weighting the observations $x_i\hat{\zeta}$ near the sample quantiles $\hat{\xi}_{np}$ and $\hat{\xi}_{n(1-p)}$. Let h_n be a sequence of positive numbers which converge to zero with a rate such that

$\lim_{n\to\infty} \frac{n}{\ln n} h_n^{(1+4/r)} = \infty$, and let $q(\cdot)$ be a continuous density function on [0 1] such that $q(0) = q(1) = 0$.[10] Denote $\hat{\xi}_n = (\hat{\xi}_{np},\ \hat{\xi}_{n(1-p)})$. Define the following smooth trimming functions t_n of $x\hat{\zeta}$:

$$t_n(x\hat{\zeta}, \hat{\xi}_n) = \begin{cases} 0 & \text{if } x\hat{\zeta} < \hat{\xi}_{np} \\ \int_0^{\frac{x\hat{\zeta}-\hat{\xi}_{np}}{h_n}} q(w)dw & \text{if } \hat{\xi}_{np} \le x\hat{\zeta} \le \hat{\xi}_{np} + h_n \\ 1 & \text{if } \hat{\xi}_{np} + h_n < x\hat{\zeta} < \hat{\xi}_{n(1-p)} - h_n \\ 1 - \int_0^{\frac{x\hat{\zeta}-\hat{\xi}_{n(1-p)}}{h_n}+1} q(w)dw & \text{if } \hat{\xi}_{n(1-p)} - h_n \le x\hat{\zeta} \le \hat{\xi}_{n(1-p)} \\ 0 & \text{if } \hat{\xi}_{n(1-p)} \le x\hat{\zeta}. \end{cases} \qquad (A.3.1)$$

This function is continuously differentiable in $\hat{\zeta}$, $\hat{\xi}_{np}$ and $\hat{\xi}_{n(1-p)}$, with

$$\frac{\partial t_n(x\hat{\zeta}, \hat{\xi}_n)}{\partial \delta'} = \begin{cases} 0 & \text{if } x\hat{\zeta} < \hat{\xi}_{np} \\ \frac{1}{h_n} q\left(\frac{x\hat{\zeta}-\hat{\xi}_{np}}{h_n}\right) x_2 & \text{if } \hat{\xi}_{np} \le x\hat{\zeta} \le \hat{\xi}_{np} + h_n \\ 0 & \text{if } \hat{\xi}_{np} + h_n < x\hat{\zeta} < \hat{\xi}_{n(1-p)} - h_n \\ -\frac{1}{h_n} q\left(\frac{x\hat{\zeta}-\hat{\xi}_{n(1-p)}}{h_n} + 1\right) x_2 & \text{if } \hat{\xi}_{n(1-p)} - h_n \le x\hat{\zeta} \le \hat{\xi}_{n(1-p)} \\ 0 & \text{if } \hat{\xi}_{n(1-p)} \le x\hat{\zeta}, \end{cases} \qquad (A.3.2)$$

$$\frac{\partial t_n(x\hat{\zeta}, \hat{\xi}_n)}{\partial \xi_{np}} = \begin{cases} 0 & \text{if } x\hat{\zeta} < \hat{\xi}_{np} \\ -\frac{1}{h_n} q\left(\frac{x\hat{\zeta}-\hat{\xi}_{np}}{h_n}\right) & \text{if } \hat{\xi}_{np} \le x\hat{\zeta} \le \hat{\xi}_{np} + h_n \\ 0 & \text{if } \hat{\xi}_{np} + h_n < x\hat{\zeta}, \end{cases} \qquad (A.3.3)$$

and

$$\frac{\partial t_n(x\hat{\zeta}, \hat{\xi}_n)}{\partial \xi_{n(1-p)}} = \begin{cases} 0 & \text{if } x\hat{\zeta} < \hat{\xi}_{n(1-p)} - h_n \\ \frac{1}{h_n} q\left(\frac{x\hat{\zeta}-\hat{\xi}_{n(1-p)}}{h_n} + 1\right) & \text{if } \hat{\xi}_{n(1-p)} - h_n \le x\hat{\zeta} \le \hat{\xi}_{n(1-p)} \\ 0 & \text{if } \hat{\xi}_{n(1-p)} < x\hat{\zeta}. \end{cases} \qquad (A.3.4)$$

As n tends to infinity, $t_n(x\hat{\zeta}, \hat{\xi}_n)$ will converge in probability to an indicator function $I_T(x\zeta_o)$, where $T = [\xi_p(\delta_o),\ \xi_{(1-p)}(\delta_o)]$.

The above trimming procedure can be generalized to models with multiple indices. For each index, it can be trimmed with the univariate function $t_n(x\hat{\zeta}, \hat{\xi}_n)$ above. The smooth trimming function can then be the product of all such univariate trimming functions.

[10] The rate of convergence of h_n is designed to justify our asymptotic analysis.

A.4 Some Useful Asymptotic Properties of Nonparametric Functions:

The following Lemma 1 provides a uniform law of large numbers for functions with a bandwidth sequence of parameters. Its proof relies on Hoeffding inequality (or Bernstein inequality) and can be found in Ichimura and Lee [1988]. Lemmas 2, 3 and 4 provide results on the asymptotic biases of functions involving a kernel function and its first two-order derivatives. The proofs of Lemmas 2, 3 and 4 are also abstracted from Ichimura and Lee [1988].

Lemma 1. *Let $g(y, a_n, \beta)$ be a measurable function which can be represented in the form*

$$g(y, a_n, \beta) = \frac{1}{a_n^d} c(y, \beta) h\left[y, \beta, \frac{s(y,\beta)}{a_n}\right],$$

where $a_n = O(\frac{1}{n^p})$, $p > 0$, $d > 0$, $\beta \in B$, and $s(y, \beta)$ is a finite dimensional vector value function. Let $\{y_i\}$ be a sequence of i.i.d. random vectors. Suppose that the following conditions are satisfied:

(i) B is a compact subset of a finite dimensional Euclidean space.

(ii) $c(y, \beta)$ is bounded by a δ-order polynomial of y uniformly in β, where $\delta \geq 0$.

(iii) the first $\delta \cdot r$ moments of y exist, where $r \geq 2$.

(iv) $h(\cdot)$ is a bounded function.

(v) $E[c^2(y, \beta) h^2(y, \beta, \frac{s(y,\beta)}{a_n}))= O(a_n^{\bar{d}})$ uniformly in $\beta \in B$, where $\bar{d} \leq d$.

(vi) $h(y, \beta, s)$ satisfies the Lipschizian condition of order 1 with respect to β and s uniformly in y. $s(y, \beta)$ satisfies the Lipschizian condition of order 1 with respect to β uniformly in y.

If $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{2(1+\frac{\delta}{r})d-\bar{d}} = \infty$, then $\frac{1}{n}\sum_{i=1}^n [g(y_i, a_n, \beta) - E(g(y_i, a_n, \beta))]$ converges in probability to zero uniformly in $\beta \in B$.

Furthermore, in addition to the above conditions, if

(vii) $E(g(y, a_n, \beta))$ converges to a limit function $g^*(\beta)$ uniformly in $\beta \in B$, then $\frac{1}{n}\sum_{i=1}^n g(y_i, a_n, \beta)$ converges in probability to $g^*(\beta)$ uniformly in $\beta \in B$.

Lemma 2. *Let $K(v)$ be a function on R^m with a bounded support D such that $\int_D |K(v)| dv < \infty$. $t(z, \theta)$ is a continuous m-dimensional random vector. Suppose that $E(c(z, z_i, \theta)|t, z_i, \theta) g(t|\theta)$, where $g(t|\theta)$ is the density function of $t(z,\theta)$, is uniformly continuous at t on R^m uniformly in (θ, z_i). Then*

$$\sup_{z_i,\theta} \left| E\left[c(z, z_i, \theta) \frac{1}{a_n^m} K\left(\frac{t(z_i, \theta) - t(z, \theta)}{a_n}\right) \Big| z_i\right] - E[c(z, z_i, \theta)|t(z_i, \theta), z_i, \theta] g(t(z_i, \theta)|\theta)\right| \longrightarrow 0.$$

Furthermore, if $K(v)$ has zero moments up to the order s^*; i.e., $\int_D v_1^{i_1}\cdots v_m^{i_m}K(v)dv = 0$, for all $i_j \geq 0$, where $j = 1,\ldots,m$, $i_1+\cdots+i_m < s^*$, and $\int_D \|v\|^{s^*}|K(v)|dv < \infty$; and $E(c(z,z_i,\theta)|t,z_i,\theta)g(t|\theta)$ is differentiable with t on R^m to the order s^* and the s^* order derivatives are uniformly bounded, then

$$\sup_{z_i,\theta}\left|E\left[c(z,z_i,\theta)\frac{1}{a_n^m}K\left(\frac{t(z_i,\theta)-t(z,\theta)}{a_n}\right)\bigg|z_i\right] - E[c(z,z_i,\theta)|t(z_i,\theta),z_i,\theta]g(t(z_i,\theta)|\theta)\right| = O(a_n^{s^*}).$$

Lemma 3. *Let $K(v)$ be a function on R^m with a bounded support D such that $K(v)$ goes to zero at the boundary of D and its gradient $\frac{\partial K(v)}{\partial v}$ is bounded. Suppose that $\frac{\partial}{\partial t}[E(c(z,z_i,\theta)|t,z_i,\theta)g(t|\theta)]$, where $g(t|\theta)$ is the density function $t(z,\theta)$, are uniformly continuous at t everywhere uniformly in (z_i,θ). Then*

$$\lim_{n\to\infty}\sup_{z_i,\theta}\left|E\left[c(z,z_i,\theta)\frac{1}{a_n^{m+1}}\frac{\partial K\left(\frac{t(z_i,\theta)-t(z,\theta)}{a_n}\right)}{\partial v}\bigg|z_i\right] - \frac{\partial}{\partial t}\{E[c(z,z_i,\theta)|t(z_i,\theta),z_i,\theta]g(t(z_i,\theta)|\theta)\}\right| = 0.$$

Furthermore, if $K(v)$ has zero moments up to the order s^*, $E(c(z,z_i,\theta)|t,z_i,\theta)g(t|\theta)$ is differentiable at t everywhere to the order s^*+1, and these derivatives are uniformly bounded, then

$$\sup_{z_i,\theta}\left|E\left[c(z,z_i,\theta)\frac{1}{a_n^{m+1}}\frac{\partial K\left(\frac{t(z_i,\theta)-t(z,\theta)}{a_n}\right)}{\partial v}\bigg|z_i\right] - \frac{\partial}{\partial t}\{E[c(z,z_i,\theta)|t(z_i,\theta),z_i,\theta]g(t(z_i,\theta)|\theta)\}\right| = O(a_n^{s^*}).$$

Lemma 4. *Let $K(v)$ be a twice differentiable function on R^m with a bounded support D such that $K(v)$ and its gradient $\frac{\partial K(v)}{\partial v}$ go to zero at the boundary of D, and the gradient $\frac{\partial K(v)}{\partial v}$ and its hessian matrix $\frac{\partial^2 K(v)}{\partial v\partial v'}$ are bounded. Suppose that $\frac{\partial^2}{\partial t\partial t'}[E(c(z,z_i,\theta)|t,z_i,\theta)g(t|\theta)]$ are uniformly continuous at t everywhere uniformly in (z_i,θ). Then*

$$\lim_{n\to\infty}\sup_{z_i,\theta}\left|E\left[c(z,z_i,\theta)\frac{1}{a_n^{m+2}}\frac{\partial^2 K\left(\frac{t(z_i,\theta)-t(z,\theta)}{a_n}\right)}{\partial v\partial v'}\bigg|z_i\right] - \frac{\partial^2}{\partial v\partial v'}\{E[c(z,z_i,\theta)|t(z_i,\theta),z_i,\theta]g(t(z_i,\theta)|\theta)\}\right| = 0.$$

Let $x\zeta$ be an m-dimensional vector of indices and K be an m-dimensional kernel function. Denote

$$A_n(s|x_i\zeta) = \frac{1}{(n-1)a_n^m}\sum_{j\neq i}^{n} s_j K\left(\frac{x_i\zeta - x_j\zeta}{a_n}\right), \tag{A.4.1}$$

where $s = 1, y, z$ or x. Under Assumption 3(3), the variance of $A_n(w|x_i\zeta)$ has the familar order $O(\frac{1}{na_n^m})$ uniformly on $C\times\Theta$, where C is any compact subset of $x\zeta$ in R^m; i.e.,

$$\sup_{C\times\Theta} var(A_n(s|x_i\zeta)|x_i\zeta,\delta) = O(\frac{1}{na_n^m}). \tag{A.4.2}$$

Assumption 2 implies that $K(\cdot)$ is bounded. It follows from Lemma 1 that if $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{(1+2/r)m} = \infty$,

$$\sup_{(x_i\zeta,\theta)\in C\times\Theta} \left| A_n(s|x_i\zeta) - E[A_n(s|x_i\zeta)|x_i\zeta,\delta] \right| \xrightarrow{p} 0. \qquad (A.4.3)$$

Under Assumptions 3(2) and 3(3), Lemma 2 guarantees that

$$\sup_{C\times\Theta} \left| E[A_n(s|x_i\zeta)|x_i\zeta,\delta] - A(s|x_i\zeta,\delta) \right| = O(a_n^{s^*}), \qquad (A.4.4)$$

where

$$A(s|x\zeta,\delta) = E(s|x\zeta,\delta)p(x\zeta|\delta). \qquad (A.4.5)$$

Since by LeCam's device T_n can be treated as a nonstochastic set contained in a compact subset of R^m for large n, $A_n(s|x_i\zeta)$ converges to $A(s|x_i\zeta,\delta)$ on $T_n \times \Theta$. Since $p(x\zeta|\delta)$ is continuous and is positive everywhere, it is bounded away from zero on $T_n \times \Theta$. Hence the uniform convergence of $A_n(1|x_i\zeta)$ implies that $\inf_{X_n\times\Theta} A_n(1|x_i\zeta)$ is bounded away from zero in probability.[11] $E_n(s|x_i\zeta)$ as a ratio of $A_n(s|x_i\zeta)$ over $A_n(1|x_i\zeta)$ will converge in probability to $E(s|x_i\zeta,\delta)$ uniformly on $T_n \times \Theta$.

The first order derivative of $A_n(s|x_i\zeta)$ is

$$\frac{\partial A_n(s|x_i\zeta)}{\partial\delta_k} = \frac{1}{(n-1)a_n^{m+1}} \sum_{j\neq i}^{n} s_j(x_i - x_j) \frac{\partial\zeta}{\partial\delta_k} \frac{\partial K(\frac{x_i\zeta - x_j\zeta}{a_n})}{\partial v}. \qquad (A.4.6)$$

Under Assumption 3(3), the variance of $\frac{\partial A_n(s|x_i\zeta)}{\partial\delta}$ has the familiar order $O(\frac{1}{na_n^{m+2}})$ uniformly in $(x_i\zeta,\delta) \in C\times\Theta$. Assumption 3(2) justifies the conditions in Lemma 3 and hence

$$\sup_{C\times\Theta} \left| E\left[\frac{\partial A_n(s|x_i\zeta)}{\partial\delta_k}|x_i\zeta,\delta\right] - \frac{\partial A(s|x_i\zeta,\delta)}{\partial\delta_k} \right| = O(a_n^{s^*}), \qquad (A.4.7)$$

where

$$\frac{\partial A(s|x_i\zeta)}{\partial\delta_k} = tr\left\{ \frac{\partial E(s(x_i - x)|x_i\zeta,\delta)}{\partial t} \frac{\partial\zeta}{\partial\delta_k} \right\} \cdot p(x_i\zeta|\delta) + E(s(x_i - x)|x_i\zeta,\delta) \frac{\partial\zeta}{\partial\delta_k} \frac{\partial p(x_i\zeta|\delta)}{\partial t}, \qquad (A.4.8)$$

where $t = (x\zeta)'$ [see Ichimura and Lee (1988)]. Lemmas 1 and 3 imply that if $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{(1+4/r)(m+1)+1} = \infty$,

$$\sup_{T_n\times\Theta} \left| \frac{\partial A_n(s|x_i\zeta)}{\partial\delta} - \frac{\partial A(s|x_i\zeta)}{\partial\delta} \right| \xrightarrow{p} 0. \qquad (A.4.9)$$

[11] The trimming of indices is designed mainly for this purpose. Otherwise, trimming would not be needed.

Similarly, under Assumptions 3(3) and 3(4), Lemmas 1 and 4 imply that if $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{(1+6/r)(m+2)+2} = \infty$,

$$\sup_{T_n\times\Theta} \left| \frac{\partial^2 A_n(s|x_i\zeta)}{\partial\delta\partial\delta'} - \frac{\partial^2 A(s|x_i\zeta)}{\partial\delta\partial\delta'} \right| \xrightarrow{p} 0. \tag{A.4.10}$$

The following propositions and lemmas will be used repeatedly for our subsequently asymptotic analysis in Appendix A.5. They are summarized here because they are of interest on their own and provide convenient reference.

Proposition 1. *Let $\xi_n = (\xi_{np}, \xi_{n(1-p)})$ be a deterministic sequence which converges to $\xi = (\xi_p, \xi_{(1-p)})$ where $\xi_{(1-p)} > \xi_p$. Suppose that*

(1) $t_n(x_i\zeta, \xi_n)$ is a bounded smooth trimming function defined in Appendix A.3, which vanishes at any $x_i\zeta$ outside the compact set T_n where $T_n = [\xi_{np}, \xi_{n(1-p)}]$.

(2) $A_n(s|x_i\zeta, \delta) \xrightarrow{p} A(s|x_i\zeta, \delta)$ uniformly in $(x_i\zeta, \delta) \in T_n \times \Theta$.

(3) $f(d)$ is uniformly continuous and bounded on A_Δ, where $A_\Delta = \{d|d = A(s|x\zeta, \delta),\ (x\zeta, \delta) \in [\xi_p - \Delta, \xi_{(1-p)} + \Delta] \times \Theta\}$ for some $\Delta > 0$.

(4) $a(s)$ is bounded by a square integrable polynomial of finite order.

Then,

$$\operatorname{plim}_{n\to\infty} \sup_{\Theta} \left| \frac{1}{n} \sum_{i=1}^{n} t_n(x_i\zeta, \xi_n) a(s_i) f(A_n(s|x_i\zeta)) - E\{t_n(x\zeta, \xi_n) a(s) f(A(s|x\zeta, \delta))\} \right| = 0.$$

In addition to the above assumptions, if $A(s|x\zeta, \delta)$ is continuous in δ and $x\zeta$ a.e., and $\hat{\delta}$ is a consistent estimate of δ_o, then

$$\operatorname{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \xi_n) a(s_i) f(A_n(s|x_i\hat{\zeta})) = E\{I_T(x\zeta_o) a(s) f(A(s|x\zeta_o, \delta_o))\},$$

where $T = [\xi_p, \xi_{(1-p)}]$.

Proof:

$$\begin{aligned}
&\sup_{\delta\in\Theta} \left| \frac{1}{n} \sum_{i=1}^{n} t_n(x_i\zeta, \xi_n) a(s_i) f(A_n(s|x_i\zeta, \delta)) - \frac{1}{n} \sum_{i=1}^{n} t_n(x_i\zeta, \xi_n) a(s_i) f(A(s|x_i\zeta, \delta)) \right| \\
&\le \frac{1}{n} \sum_{i=1}^{n} |a(s_i)| \cdot \sup_{(x_i\zeta, \delta)\in T_n\times\Theta} \left| f(A_n(s|x_i\zeta, \delta)) - f(A(s|x_i\zeta, \delta)) \right| \\
&\xrightarrow{p} 0,
\end{aligned}$$

because t_n is bounded by one, $f(A_n(s|x_i\zeta, \delta))$ converges in probability to $f(A(s|x_i\zeta, \delta))$ uniformly on $T_n \times \Theta$, and $\frac{1}{n}\sum_{i=1}^{n} |a(s_i)| = O_p(1)$. Because $\sup_\Theta |t_n(x\zeta, \xi_n) f(A(s|x\zeta, \delta))| \le M$ is bounded and $a(s)$ is bounded by a polynomial of s, the uniform law of large numbers in Lemma 1 with $d = \bar{d} = 0$ implies that as n goes to infinity,

$$\frac{1}{n}\sum_{i=1}^{n}\{t_n(x_i\zeta,\xi_n)a(s_i)f(A(s|x_i\zeta,\delta)) - E[t_n(x_i\zeta,\xi_n)a(s_i)f(A(s|x_i\zeta,\delta))]\} \xrightarrow{p} 0$$

uniformly on Θ. The second part of the conclusion follows as $E[t_n(x\zeta,\xi_n)a(s)f(A(s|x\zeta,\delta))]$ is continuous in δ and ξ_n by the Lebesque dominated convergence theorem.

Q.E.D.

Proposition 2. *Let $\xi_n = (\xi_{np},\xi_{n(1-p)})$, where $\xi_{np} = (\xi_{1,np},\xi_{(2),np})$, be a deterministic sequence which converges to $\xi = (\xi_p,\xi_{(1-p)})$, where $\xi_{(1-p)} > \xi_p$. Suppose that*

(1) q is a continuously differentiable density function having a support $[0,1]$.

(2) $A_n(s|x_i\zeta,\delta) \xrightarrow{p} A(s|x_i\zeta,\delta)$ uniformly in $(x_i\zeta,\delta) \in T_n \times \Theta$, where $T_n = [\xi_{np},\xi_{n(1-p)}]$ and $A(s|x\zeta,\delta)$ is continuous in δ and $x\zeta$.

(3) $f(d)$ is uniformly continuous on A_Δ, where $A_\Delta = \{d|d = A(s|x\zeta,\delta),\ (x\zeta,\delta) \in [\xi_p - \Delta,\xi_{(1-p)} + \Delta] \times \Theta\}$ for some $\Delta > 0$.

(4) $a(s)$ is bounded by a polynomial of order δ, $\delta \geq 0$. The $r \cdot \delta$ order moment of s exists, where $r \geq 2$.

(5) The functions $F_n(x\zeta_{(2)},\xi_n)$ are uniformly bounded (by a constant M, say) and $\lim_{n\to\infty} F_n(x\zeta_{(2)},\xi_n) = F(x\zeta_{(2)},\xi)$, a.e., where $x\zeta = (x\zeta_1,x\zeta_{(2)})$.

(6) $E(a(s)|x\zeta,\delta)g(x\zeta|\delta)$ and $E(|a(s)||x\zeta,\delta)g(x\zeta|\delta)$ are uniformly continuous at $x\zeta$ and δ, where $g(t|\delta)$ is the density function of $x\zeta$.

(7) The function $\sup_{\delta\in\Theta}|E[a(s)|\xi_{1,p},x\zeta_{(2)},\delta]g(\xi_{1,p},x\zeta_{(2)}|\delta)f(A(s|\xi_{1,p},x\zeta_{(2)},\delta))|$, where $\xi_p = (\xi_{1,p},\xi_{(2),p})$, is dominated by a Lebesque integrable function of $x\zeta_{(2)}$.

Then, under the rate that $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} \frac{nh_n^{(1+2\delta/r)}}{\ln n} = \infty$,

$$\operatorname{plim}_{n\to\infty}\sup_{\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}q\left(\frac{x_i\zeta_1 - \xi_{1,np}}{h_n}\right)F_n(x_i\zeta_{(2)},\xi_n)a(s_i)f(A_n(s|x_i\zeta,\delta)) - \int E[a(s)|\xi_{1,p},t_{(2)},\delta]f(A(s|\xi_{1,p},t_{(2)},\delta))F(t_{(2)},\xi)g(\xi_{1,p},t_{(2)}|\delta)dt_{(2)}\right| = 0.$$

In addition to the above assumptions, if $\hat{\delta}$ is a consistent estimate of δ_o, then

$$\operatorname{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}q\left(\frac{x_i\hat{\zeta}_1 - \xi_{1,np}}{h_n}\right)F_n(x_i\hat{\zeta}_{(2)},\xi_n)a(s_i)f(A_n(s|x_i\hat{\zeta},\hat{\delta})) = \int E[a(s)|\xi_{1,p},t_{(2)},\delta_o]f(A(s|\xi_{1,p},t_{(2)},\delta_o))F(t_{(2)},\xi)g(\xi_{1,p},t_{(2)}|\delta_o)dt_{(2)}.$$

Proof:

$$\sup_{\delta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}q(\frac{x_i\zeta_1-\xi_{1,np}}{h_n})F_n(x_i\zeta_{(2)},\xi_n)a(s_i)\left[f(A_n(s|x_i\zeta,\delta))-f(A(s|x_i\zeta,\delta))\right]\right|$$
$$\leq\frac{1}{n}\sum_{i=1}^{n}|a(s_i)|\frac{1}{h_n}q(\frac{x_i\zeta_1-\xi_{1,np}}{h_n})\cdot M\cdot\sup_{(x_i\zeta,\delta)\in T_n\times\Theta}\left|f(A_n(s|x_i\zeta,\delta))-f(A(s|x_i\zeta))\right|. \qquad (*)$$

Let $N(\xi_{1,p})$ be a compact neighborhood of $\xi_{1,p}$. By Lemma 1, as $\frac{nh_n^{(1+2\delta/r)}}{\ln n}=\infty$,

$$\sup_{(\delta,\bar{\xi})\in\Theta\times N(\xi_{1,p})}\left|\frac{1}{n}\sum_{i=1}^{n}|a(s_i)|\frac{1}{h_n}q\left(\frac{x_i\zeta_1-\bar{\xi}}{h_n}\right)-E\left\{|a(s)|\frac{1}{h_n}q\left(\frac{x_i\zeta_1-\bar{\xi}}{h_n}\right)\right\}\right|\xrightarrow{p}0.$$

By Lemma 2,

$$\sup_{(\delta,\bar{\xi})\in\Theta\times N(\xi_{1,p})}\left|E\left\{|a(s)|\frac{1}{h_n}q\left(\frac{x_i\zeta_1-\bar{\xi}}{h_n}\right)\right\}-E(|a(s)||\bar{\xi},\delta)g_1(\bar{\xi}|\delta)\right|\to 0,$$

where $g_1(\cdot|\delta)$ is the marginal density of $x\zeta_1$. Since $E(|a(s)||\xi_{1,n},\delta)g_1(\xi_{1,n}|\delta)$ is bounded on Θ for large n, it follows that

$$\sup_{\delta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}|a(s_i)|\frac{1}{h_n}q(\frac{x_i\zeta_1-\xi_{1,np}}{h_n})=O_p(1),$$

and hence (*) goes to zero in probability. Similarly, by Lemma 1,

$$\sup_{\delta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}q\left(\frac{x_i\zeta_1-\xi_{1,np}}{h_n}\right)F_n(x_i\zeta_{(2)},\xi_n)a(s_i)f(A(s|x_i\zeta,\delta))\right.$$
$$\left.-E\left[\frac{1}{h_n}q\left(\frac{x\zeta_1-\xi_{1,np}}{h_n}\right)F_n(x\zeta_{(2)},\xi_n)a(s)f(A(s|x\zeta,\delta))\right]\right|\xrightarrow{p}0.$$

Since

$$E\left[\frac{1}{h_n}q\left(\frac{x\zeta_1-\xi_{1,np}}{h_n}\right)F_n(x\zeta_{(2)},\xi_n)a(s)f(A(s|x\zeta,\delta))\right]$$
$$=\int\left\{\int E[a(s)|\xi_{1,np}+h_nw_1,t_{(2)},\delta]f(A(s|\xi_{1,np}+h_nw_1,t_{(2)},\delta))g(\xi_{1,np}+h_nw_1,t_{(2)}|\delta)q(w_1)dw_1\right\}$$
$$\cdot F_n(t_{(2)},\xi_n)dt_{(2)},$$

it follows that

$$\sup_{\delta\in\Theta}\left|E\left[\frac{1}{h_n}q\left(\frac{x\zeta_1-\xi_{1,np}}{h_n}\right)F_n(x\zeta_{(2)},\xi_n)a(s)f(A(s|x\zeta,\delta))\right]\right.$$
$$\left.-\int E[a(s)|\xi_{1,p},t_{(2)},\delta]f(A(s|\xi_{1,p},t_{(2)},\delta))g(\xi_{1,p},t_{(2)}|\delta)F(t_{(2)},\xi)dt_{(2)}\right|$$
$$\leq\sup_{\delta\in\Theta,w_1,t_{(2)}}\left|E[a(s)|\xi_{1,np}+h_nw_1,t_{(2)},\delta]f(A(s|\xi_{1,np}+h_nw_1,t_{(2)},\delta))g(\xi_{1,np}+h_nw_1,t_{(2)}|\delta)\right.$$
$$\left.-E[a(s)|\xi_{1,p},t_{(2)},\delta]f(A(s|\xi_{1,p},t_{(2)},\delta))g(\xi_{1,p},t_{(2)}|\delta)\right|\cdot\int|F(t_{(2)},\xi)|dt_{(2)}$$
$$+\int\sup_{\delta\in\Theta}\left|E[a(s)|\xi_{1,p},t_{(2)},\delta]f(A(s|\xi_{1,p},t_{(2)},\delta))g(\xi_{1,p},t_{(2)}|\delta)\right|\cdot|F_n(t_{(2)},\xi_n)-F(t_{(2)},\xi)|dt_{(2)}$$
$$\longrightarrow 0$$

by the uniform continuity and the Lebesque convergence theorem.

The first part of the result follows from the above convergences. The second part of the result follows since $\hat{\delta}$ is a consistent estimate of δ.

Q.E.D.

Lemma 5. *Let $C_n(s_1, \cdots, s_{i-1}, s_{i+1}, \cdots, s_n;\ s_i)$ be a sequence of measurable functions of an i.i.d. sample $\{s_i\}$ and $d_n(s)$ be a measurable function such that $E(I_{Z_n}(s)d_n^2(s)) < \infty$ uniformly in n. Suppose that,*

(1) $\sup_{Z_n} |E(C_n(s_1, \cdots, s_n;\ s_i)|s_i) - C(s_i)| = O(a_n^s)$, *for some measurable function $C(s_i)$, and*

(2) $\sup_{Z_n} var(C_n(s_1, \cdots, s_n;\ s_i)|s_i) = O(\frac{1}{na_n^r})$.

If $s > r/2$, $\lim_{n\to\infty} na_n^{2r} = \infty$, and $\lim_{n\to\infty} na_n^{4s} = 0$, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{Z_n}(s_i) d_n(s_i) \left|C_n(s_1, \cdots, s_n; s_i) - C(s_i)\right|^2 \xrightarrow{p} 0.$$

Proof: The proof of this lemma can be found in Lee [1990].

Proposition 3. *Let $\nabla_a f(s,a)$ and $\nabla_a^2 f(s,a)$ denote the first and second order derivatives of $f(s,a)$ with respect to a. Suppose that*

(1) the nonparametric function $A_n(s|x_i\zeta_o, \delta_o)$ satisfies the conditions in Lemma 5, and

(2)

$$\sup_{x_i\zeta_o \in T} \left|\nabla_a^2 f_n(s_i, \bar{A}_n(s|x_i\zeta_o))\right| \leq O_p(1) \cdot d_n(s_i),$$

where $\bar{A}_n(s|x_i\zeta_o)$ is a consistent estimate of $A(s|x_i\zeta_o, \delta_o)$, and $d_n(s_i)$ is a measurable function such that $E(I_T(x_i\zeta_o)d_n^2(s_i)) < \infty$ uniformly in n. Then

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_o) f_n(s_i, A_n(s|x_i\zeta_o)) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_o) f_n(s_i, A(s|x_i\zeta_o, \delta_o)) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_o) \nabla_a f_n(s_i, A(s|x_i\zeta_o, \delta_o)) \cdot (A_n(s|x_i\zeta_o) - A(s|x_i\zeta_o, \delta_o)) + o_p(1).
\end{aligned}$$

Proof: By a Taylor expansion,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_o) f_n(s_i, A_n(s|x_i\zeta_o)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_o) f_n(s_i, A(s|x_i\zeta_o, \delta_o)) + L_n + R_n,$$

where

$$L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_o) \nabla_a f_n(s_i, A(s|x_i\zeta_o, \delta_o))(A_n(s|x_i\zeta_o) - A(s|x_i\zeta_o, \delta_o))$$

and

$$R_n = \frac{1}{2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_o) \nabla_a^2 f_n(s_i, \bar{A}_n(s|x_i\zeta_o))(A_n(s|x_i\zeta_o) - A(s|x_i\zeta_o, \delta_o))^2,$$

where $\bar{A}_n(s|x_i\zeta_o)$ lies between $A_n(s|x_i\zeta_o)$ and $A(s|x_i\zeta_o, \delta_o)$. Since

$$|R_n| \leq O_p(1) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_o) d_n(s_i)[A_n(s|x_i\zeta_o) - A(s|x_i\zeta_o, \delta_o)]^2$$

by our assumption, Lemma 5 implies that R_n converges to zero in probability.

Q.E.D.

Lemma 6. *Let $\{s_i\}$ be an i.i.d. sample and $\Phi_n(s_1, s_2, a_n)$ be a sequence of vector valued random functions with bandwidth a_n. Suppose that*

(1) there exist square integrable functions $h_j(s)$, $j = 1, 2$ such that

$$\left| E(\Phi_n(s_1, s_2, a_n)|s_j) \right| \leq h_j(s_j), \qquad j = 1, 2,$$

(2) $E(\Phi_n(s_1, s_2, a_n)) = O(a_n^s)$ and $var(\Phi_n(s_1, s_2, a_n)) = O(\frac{1}{a_n^r})$,

(3) $\lim_{n\to\infty} E(\Phi_n(s_1, s_2, a_n)|s_j) = \psi_j(s_j)$, a.e., for some measurable functions ψ_j, $j = 1, 2$, and

(4) $\lim_{n\to\infty} \sqrt{n} a_n^s = 0$ and $\lim_{n\to\infty} n a_n^r = \infty$.

If $\psi_1(s)$ and $\psi_2(s)$ are zero a.e., then

$$\frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \Phi_n(s_i, s_j, a_n) \xrightarrow{p} 0.$$

On the other hand, if $E\{[\psi_1(s) + \psi_2(s)][\psi_1(s) + \psi_2(s)]'\} = \Sigma$ which is nonzero, then

$$\frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \Phi_n(s_i, s_j, a_n) \xrightarrow{D} N(0, \Sigma).$$

Proof: The proof of this lemma can be found in Lee [1990]. See also Powell, Stock and Stoker [1989] for a result similar to the second part of the lemma.

Proposition 4: *Suppose that K is an r-dimensional kernel function with a bounded support D such that $\int_D |K(w)|dw < \infty$ and with a bandwidth a_n. Let $A_n(s|x_i\zeta_o) = \frac{1}{(n-1)a_n^r} \sum_{j\neq i}^{n} s_j K\left(\frac{x_i\zeta_o - x_j\zeta_o}{a_n}\right)$ and let $g(x\zeta_o|\delta_o)$ be the density of $x\zeta_o$. Denote $A(s|x\zeta_o, \delta_o) = E(s|x\zeta_o, \delta_o)g(x\zeta_o|\delta_o)$. Suppose that*

(1) $E(s|x\zeta_o, \delta_o)g(x\zeta_o|\delta_o)$ is continuous at $x\zeta_o$ a.e., and

(2) $E(f_n(s, x\zeta_o)|x\zeta_o, \delta_o) = 0$ a.e., for all n.

If $\lim_{n\to\infty} na_n^r = \infty$, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_n(s_i, x_i\zeta_o)[A_n(s|x_i\zeta_o) - A(s|x_i\zeta_o, \delta_o)] \xrightarrow{p} 0.$$

Proof: Define

$$\Phi_n(s_i, s_j, a_n) = f_n(s_i, x_i\zeta_o)\left[\frac{1}{a_n^r} s_j K\left(\frac{x_i\zeta_o - x_j\zeta_o}{a_n}\right) - A(s|x_i\zeta, \delta_o)\right].$$

It follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_n(s_i, x_i\zeta_o)[A_n(s|x_i\zeta_o) - A(s|x_i\zeta_o, \delta_o)] = \frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Phi_n(s_i, s_j, a_n).$$

By (1), $E[s_j \frac{1}{a_n^r} K(\frac{x_i\zeta_o - x_j\zeta_o}{a_n})|x_i\zeta_o]$ converges to $A(s|x_i\zeta_o, \delta_o)$ at $x_i\zeta_o$ a.e. Hence, $E(\Phi_n(s_i, s_j, a_n)|s_i)$ converges to zero a.e. On the other hand,

$$\begin{aligned}
&E(\Phi_n(s_j, s_i, a_n)|s_i) \\
&= s_i E\left[f_n(s_j, x_j\zeta_o)\frac{1}{a_n^r}K\left(\frac{x_j\zeta_o - x_i\zeta_o}{a_n}\right)|x_i\right] - E[f_n(s_j, x_j\zeta_o)A(s|x_j\zeta_o, \delta_o)] \\
&= s_i \int E[f_n(s, x_i\zeta_o + a_n w)|x_i\zeta_o + a_n w, \delta_o]g(x_i\zeta_o + a_n w|\delta_o)K(w)dw \\
&\quad - E\{E[f_n(s, x\zeta_o|x\zeta_o, \delta_o)]A(s|x\zeta_o, \delta_o)\} \\
&= 0.
\end{aligned}$$

The conclusion of the proposition follows from Lemma 6.

Q.E.D.

Proposition 5: *Suppose that K is an r-dimensional kernel function with a bounded support D such that $\int_D |K(w)|dw < \infty$ and with a bandwidth a_n. Let $A_n(s|x_i\zeta_o) = \frac{1}{(n-1)a_n^r}\sum_{j\neq i}^{n} s_j K\left(\frac{x_i\zeta_o - x_j\zeta_o}{a_n}\right)$ and let $g(x\zeta_o|\delta_o)$ be the density of $x\zeta_o$. Denote $A(s|x\zeta_o, \delta_o) = E(s|x\zeta_o, \delta_o)g(x\zeta_o|\delta_o)$. Suppose that*

(1) $E(s|x\zeta_o, \delta_o)g(x\zeta_o|\delta_o)$ is continuous at $x\zeta_o$, a.e.,

(2) $E(f_n(s, x\zeta_o)|x\zeta_o, \delta_o)g(x\zeta_o|\delta_o)$ is continuous in $x\zeta_o$ a.e. uniformly in n,

(3) there exists a measurable function $h(x\zeta_o)$ such that

$$\left|E(f_n(s, x\zeta_o)|x\zeta_o, \delta_o)\right| \leq h(x\zeta_o)$$

with $E\left|h(x\zeta_o)A(s|x\zeta_o,\delta_o)\right| < \infty$ *for large* n, *and*

(4) $\lim_{n\to\infty} E(f_n(s,x\zeta_o)|x\zeta_o,\delta_o) = c(x\zeta_o)$ *a.e.*

If $\lim_{n\to\infty} na_n^r = \infty$ *and* $\lim_{n\to\infty} na_n^{2s} = 0$, *then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_n(s_i,x_i\zeta_o)[A_n(s|x_i\zeta_o) - A(s|x_i\zeta_o,\delta_o)] \xrightarrow{D} N(0,\Phi),$$

where

$$\begin{aligned}\Phi =& E\big\{\big[s\cdot c(x\zeta_o)g(x\zeta_o|\delta_o) - E\big(c(x\zeta_o)E(s|x\zeta_o,\delta_o)g(x\zeta_o|\delta_o)\big)\big]' \\ &\cdot\big[s\cdot c(x\zeta_o)g(x\zeta_o|\delta_o) - E\big(c(x\zeta_o)E(s|x\zeta_o,\delta_o)g(x\zeta_o|\delta_o)\big)\big]\big\}.\end{aligned}$$

Proof:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_n(s_i,x_i\zeta_o)[A_n(s|x_i\zeta_o) - A(s|x_i\zeta_o,\delta_o)] = \frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Phi_n(s_i,s_j,a_n),$$

where

$$\Phi_n(s_i,s_j,a_n) = f_n(s_i,x_i\zeta_o)\left[\frac{1}{a_n^r}s_jK\left(\frac{x_i\zeta_o - x_j\zeta_o}{a_n}\right) - A(s|x_i\zeta,\delta_o)\right].$$

By (1), $E(\Phi_n(s_i,s_j,a_n)|s_i)$ converges to zero a.e. On the other hand,

$$\begin{aligned}&E(\Phi_n(s_j,s_i,a_n)|s_i)\\ &= s_iE\left[f_n(s_j,x_j\zeta_o)\frac{1}{a_n^r}K\left(\frac{x_j\zeta_o - x_i\zeta_o}{a_n}\right)|x_i\right] - E[f_n(s_j,x_j\zeta_o)A(s|x_j\zeta_o,\delta_o)]\\ &= s_i\int E[f_n(s,x_i\zeta_o + a_nw)|x_i\zeta_o + a_nw,\delta_o]g(x_i\zeta_o + a_nw|\delta_o)K(w)dw\\ &\quad - E\{E[f_n(s,x\zeta_o)|x\zeta_o,\delta_o]A(s|x\zeta_o,\delta_o)\}.\end{aligned}$$

Since

$$\begin{aligned}&\left|\int E[f_n(s,x_i\zeta_o + a_nw)|x_i\zeta_o + a_nw,\delta_o]g(x_i\zeta_o + a_nw|\delta_o)K(w)dw - E[f_n(s,x\zeta_o)|x_i\zeta_o,\delta_o]g(x_i\zeta_o|\delta_o)\right|\\ &= \left|\int\{E[f_n(s,x_i\zeta_o + a_nw)|x_i\zeta_o + a_nw,\delta_o]g(x_i\zeta_o + a_nw|\delta_o) - E[f_n(s,x_i\zeta_o)|x_i\zeta_o,\delta_o]g(x_i\zeta_o|\delta_o)\}K(w)dw\right|\\ &\leq \sup_w\left|E[f_n(s,x_i\zeta_o + a_nw)|x_i\zeta_o + a_nw,\delta_o]g(x_i\zeta_o + a_nw|\delta_o) - E[f_n(s,x_i\zeta_o)|x_i\zeta_o,\delta_o]g(x_i\zeta_o|\delta_o)\right|\\ &\qquad\cdot\int|K(w)|dw\\ &\to 0,\end{aligned}$$

by condition (2), it follows that, by the Lebesque convergence theorem and condition (4),

$$\lim_{n\to\infty} E(\Phi_n(s_j,s_i,a_n)|s_i) = s_ic(x_i\zeta_o)g(x_i\zeta_o|\delta_o) - E\big(c(x\zeta_o)E(s|x\zeta_o,\delta_o)g(x\zeta_o|\delta_o)\big) \quad a.e.$$

The result of the proposition follows from the second part of Lemma 6.

Q.E.D.

Proposition 6: *Suppose that K is an r-dimensional kernel function with a bounded support D such that $\int_D |K(w)|dw < \infty$ and with a bandwidth a_n. Let $D_n(s|x_i\zeta_o) = \frac{1}{(n-1)a_n^{r+1}} \sum_{j\neq i}^n s_j \nabla K\left(\frac{x_i\zeta_o - x_j\zeta_o}{a_n}\right)$, where $\nabla K(t) = \frac{\partial K(t)}{\partial t}$ is the gradient vector of K, and let $g(x\zeta_o|\delta_o)$ be the density of $x\zeta_o$. Denote $D(s|x\zeta_o, \delta_o) = \frac{\partial}{\partial(x\zeta_o)}\{E(s|x\zeta_o, \delta_o)g(x\zeta_o|\delta_o)\}$. Suppose that*

(1) $K(w)$ vanishes at the boundary of D,

(2) $\frac{\partial}{\partial(x\zeta_o)}\{E(s|x\zeta_o, \delta_o)g(x\zeta_o|\delta_o)\}$ is continuous at $x\zeta_o$ a.e.,

(3) $E(f_n(s, x\zeta_o)|x\zeta_o, \delta_o) = 0$ a.e., for all n.

If $\lim_{n\to\infty} na_n^{r+2} = \infty$, then $\frac{1}{\sqrt{n}}\sum_{i=1}^n f_n(s_i, x_i\zeta_o)[D_n(s|x_i\zeta_o) - D(s|x_i\zeta_o, \delta_o)] \xrightarrow{p} 0$.

Proof: Define

$$\Phi_n(s_i, s_j, a_n) = f_n(s_i, x_i\zeta_o)\left[\frac{1}{a_n^{r+1}} s_j \nabla K\left(\frac{x_i\zeta_o - x_j\zeta_o}{a_n}\right) - D(s|x_i\zeta, \delta_o)\right].$$

It follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n f_n(s_i, x_i\zeta_o)[D_n(s|x_i\zeta_o) - D(s|x_i\zeta_o, \delta_o)] = \frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^n\sum_{j\neq i}^n \Phi_n(s_i, s_j, a_n).$$

By condition (2), Lemma 3 implies that $E[s_j \frac{1}{a_n^{r+1}}\nabla K(\frac{x_i\zeta_o - x_j\zeta_o}{a_n})|x_i\zeta_o]$ converges to $D(s|x_i\zeta_o, \delta_o)$ at $x_i\zeta_o$ a.e. Hence $E(\Phi_n(s_i, s_j, a_n)|s_i)$ converges to zero a.e. On the other hand, by the integration by parts and under conditions (1) and (3),

$$\begin{aligned}
&E(\Phi_n(s_j, s_i, a_n)|s_i) \\
&= s_i \cdot E\left[f_n(s_j, x_j\zeta_o)\frac{1}{a_n^{r+1}}\nabla K\left(\frac{x_j\zeta_o - x_i\zeta_o}{a_n}\right)|x_i\right] - E[f_n(s_j, x_j\zeta_o)D(s|x_j\zeta_o, \delta_o)] \\
&= s_i \cdot \frac{1}{a_n}\int_D E[f_n(s, x_i\zeta_o + a_n w)|x_i\zeta_o + a_n w, \delta_o]g(x_i\zeta_o + a_n w|\delta_o)\nabla K(w)dw \\
&\quad - E\{E[f_n(s, x\zeta_o)|x\zeta_o, \delta_o]D(s|x\zeta_o, \delta_o)\} \\
&= -s_i \cdot \frac{1}{a_n}\int_D \nabla\{E[f_n(s, x_i\zeta_o + a_n w)|x_i\zeta_o + a_n w, \delta_o]g(x_i\zeta_o + a_n w|\delta_o)\}K(w)dw \\
&\quad - E\{E[f_n(s, x\zeta_o)|x\zeta_o, \delta_o]D(s|x\zeta_o, \delta_o)\} \\
&= 0.
\end{aligned}$$

The result of the proposition follows from the first part of Lemma 6.

Q.E.D.

A.5 Asymptotic Properties:

The propositions in Appendix A.4 can be used to derive the asymptotic properties of our estimators. Proposition 1 implies that

$$\begin{aligned}\frac{1}{n}\hat{X}_2'\hat{X}_2 &= \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi}_n)(x_{2i} - E_n(x_2|x_i\hat{\zeta}))'(x_{2i} - E_n(x_2|x_i\hat{\zeta})) \\ &\xrightarrow{p} E[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))'(x_2 - E(x_2|x\zeta_o,\delta_o))].\end{aligned} \tag{A.5.1}$$

From $(4.2)'$, $v^* = [v_2^*,\ (x\zeta_o, 0)J]$. Since $E(v_2^*|x) = E(v_2^*|x\zeta_o,\delta_o)$, it follows that $E[v^* - E(v^*|x\zeta_o,\delta_o)|x] = 0$. Therefore,

$$\begin{aligned}\frac{1}{n}\hat{W}'\hat{X}_2 &= \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi}_n)(w_i - E_n(w|x_i\hat{\zeta}))'(x_{2i} - E_n(x_2|x_i\hat{\zeta})) \\ &= \Pi_w'\frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi}_n)(x_{2i} - E_n(x_2|x_i\hat{\zeta}))'(x_{2i} - E_n(x_2|x_i\hat{\zeta})) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi}_n)(v_i^* - E_n(v^*|x_i\hat{\zeta}))'(x_{2i} - E_n(x_2|x_i\hat{\zeta})) \\ &\xrightarrow{p} \Pi_w' E[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))'(x_2 - E(x_2|x\zeta_o,\delta_o))].\end{aligned} \tag{A.5.2}$$

Similarly, for the SG2LS estimator,

$$\begin{aligned}\frac{1}{n}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{X}_2 &= \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi}_n)(x_{2i} - E_n(x_2|x_i\hat{\zeta}))'\omega_{ni}^{-1}(\hat{\beta},\hat{\delta})(x_{2i} - E_n(x_2|x_i\hat{\zeta})) \\ &\xrightarrow{p} E[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))'\omega^{-1}(x\zeta_o,\delta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))],\end{aligned} \tag{A.5.3}$$

and

$$\begin{aligned}\frac{1}{n}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{W} &= \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi}_n)(x_{2i} - E_n(x_2|x_i\hat{\zeta}))'\omega_{ni}^{-1}(\hat{\beta},\hat{\delta})(w_i - E_n(w|x_i\hat{\zeta})) \\ &\xrightarrow{p} E[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))'\omega^{-1}(x\zeta_o,\delta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))]\Pi_w.\end{aligned} \tag{A.5.4}$$

Furthermore,

$$\frac{1}{n}\hat{X}_2'\hat{\Lambda}_n^{-1}\frac{\partial E_n(\hat{\beta},\hat{\zeta})}{\partial\delta'} \xrightarrow{p} E\left[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))'\omega^{-1}(x\zeta_o,\delta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right], \tag{A.5.5}$$

$$\frac{1}{n}\frac{\partial E_n(\hat{\beta},\hat{\zeta})}{\partial\delta}\hat{\Lambda}_n^{-1}\frac{\partial E_n(\hat{\beta},\hat{\zeta})}{\partial\delta'} \xrightarrow{p} E\left[I_T(x\zeta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta}\omega^{-1}(x\zeta_o,\delta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right], \tag{A.5.6}$$

and

$$\frac{1}{n}\frac{\partial E_n(\hat{\beta},\hat{\zeta})}{\partial\delta}\hat{\Lambda}_n^{-1}\hat{W}_n \xrightarrow{p} E\left[I_T(x\zeta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta}\omega^{-1}(x\zeta_o,\delta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))\right]\Pi_w. \tag{A.5.7}$$

The asymptotic distributions of $\frac{1}{\sqrt{n}}\hat{X}_2'\hat{U}_n$, $\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{U}_n$ and $\frac{1}{\sqrt{n}}\frac{\partial E_n(\hat{\beta},\hat{\zeta})}{\partial\delta}\hat{\Lambda}_n^{-1}\hat{U}_n$ can be derived from Propositions 2-4 and 6. Let $\xi_o = (\xi_p(\delta_o), \xi_{(1-p)}(\delta_o))$.

$$\begin{aligned}\frac{1}{\sqrt{n}}\hat{X}_2'\hat{U}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi}_n)(x_{2i} - E_n(x_2|x_i\hat{\zeta}))'(u_{1i} - E_n(u_1|x_i\hat{\zeta})) \\ &= C_{0,n} + \{C_{1,n} + C_{2,n} + C_{3,n}\}\sqrt{n}(\hat{\delta} - \delta_o) + C_{4,n}\sqrt{n}(\hat{\xi} - \xi_o),\end{aligned}$$

where

$$C_{0,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o)(x_{2i} - E_n(x_2|x_i\zeta_o))'(u_{1i} - E_n(u_1|x_i\zeta_o)),$$

$$C_{1,n} = -\frac{1}{n} \sum_{i=1}^{n} t_n(x_i\bar{\zeta}, \bar{\xi})(x_{2i} - E_n(x_2|x_i\bar{\zeta}))' \frac{\partial E_n(u_1|x_i\bar{\zeta})}{\partial \delta'},$$

$$C_{2,n} = -\frac{1}{n} \sum_{i=1}^{n} t_n(x_i\bar{\zeta}, \bar{\xi})(u_{1i} - E_n(u_1|x_i\bar{\zeta}))' \frac{\partial E_n(x_2'|x_i\bar{\zeta})}{\partial \delta'},$$

$$C_{3,n} = \frac{1}{n} \sum_{i=1}^{n} (x_{2i} - E_n(x_2|x_i\bar{\zeta}))'(u_{1i} - E_n(u_1|x_i\bar{\zeta})) \frac{\partial t_n(x_i\bar{\zeta}, \bar{\xi})}{\partial \delta'},$$

and

$$C_{4,n} = \frac{1}{n} \sum_{i=1}^{n} (x_{2i} - E_n(x_2|x_i\bar{\zeta}))'(u_{1i} - E_n(u_1|x_i\bar{\zeta})) \frac{\partial t_n(x_i\bar{\zeta}, \bar{\xi})}{\partial \xi'}.$$

Proposition 1 implies that

$$C_{1,n} \xrightarrow{p} -E\left(t_n(x\zeta_o, \xi_o)(x_2 - E(x_2|x\zeta_o, \delta_o))' \frac{\partial E(u_1|x\zeta_o, \delta_o)}{\partial \delta'}\right),$$

and $C_{2,n}$ converges to zero in probability. As $\lim_{n\to\infty} \frac{n}{\ln n} h_n^{(1+4/r)} = \infty$, Proposition 2 implies that $C_{3,n}$ and $C_{4,n}$ converge to zero in probability. To simplify notations, let $A_{ni} = \frac{1}{(n-1)a_n^m} \sum_{j\neq i}^{n} x_{2j} K\left(\frac{x_i\zeta_o - x_j\zeta_o}{a_n}\right)$, $B_{ni} = \frac{1}{(n-1)a_n^m} \sum_{j\neq i}^{n} K\left(\frac{x_i\zeta_o - x_j\zeta_o}{a_n}\right)$, and $C_{ni} = \frac{1}{(n-1)a_n^m} \sum_{j\neq i}^{n} u_{1j} K\left(\frac{x_i\zeta_o - x_j\zeta_o}{a_n}\right)$. Let $A_i = E(x_2|x_i\zeta_o, \delta_o)p(x_i\zeta_o|\delta_o)$, $B_i = p(x_i\zeta_o|\delta_o)$, and $C_i = E(u_1|x_i\zeta_o, \delta_o)p(x_i\zeta_o|\delta_o)$ be their limits. By a Taylor expansion,

$$C_{0,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o)(x_{2i} - E(x_2|x_i\zeta_o, \delta_o))'(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) + L_n + R_n,$$

where

$$\begin{aligned}
L_n = &-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \frac{1}{B_i}(u_{1i} - E(u_1|x_i\zeta_o, \delta_o))(A_{ni} - A_i)' \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o)(x_{2i} - E(x_2|x_i\zeta_o, \delta_o))' \frac{1}{B_i}(C_{ni} - C_i) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o)\left[E(x_2|x_i\zeta_o, \delta_o)'(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) + (x_{2i} - E(x_2|x_i\zeta_o, \delta_o))'E(u_1|x_i\zeta_o, \delta_o)\right] \\
&\cdot \frac{1}{B_i}(B_{ni} - B_i),
\end{aligned}$$

and

$$
\begin{aligned}
R_n =& \frac{2}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)\frac{1}{\tilde{B}_{ni}^2}(A_{ni}-A_i)'(C_{ni}-C_i)\\
&+\frac{2}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)[(u_{1i}\tilde{B}_{ni}-\tilde{C}_{ni})-\tilde{C}_{ni}]\frac{1}{\tilde{B}_{ni}^3}(A_{ni}-A_i)'(B_{ni}-B_i)\\
&+\frac{2}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)[-\tilde{A}_{ni}'+(x_{2i}\tilde{B}_{ni}-\tilde{A}_{ni})']\frac{1}{\tilde{B}_{ni}^3}(B_{ni}-B_i)(C_{ni}-C_i)\\
&+\frac{2}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)[-\tilde{A}_{ni}'(u_{1i}\tilde{B}_{ni}-\tilde{C}_{ni})+\tilde{A}_{ni}'\tilde{C}_{ni}-(x_{2i}\tilde{B}_{ni}-\tilde{A}_{ni})'\tilde{C}_{ni}]\frac{1}{\tilde{B}_{ni}^4}(B_{ni}-B_i)^2.
\end{aligned}
$$

Since $\lim_{n\to\infty} na_n^{2m}=\infty$ and $\lim_{n\to\infty} na_n^{4s^*}=0$, R_n converges to zero in probability by Proposition 3. As $\lim_{n\to\infty} na_n^m=\infty$, L_n converges to zero in probability by Proposition 4. Hence

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\hat{X}_2'\hat{U}_n =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)(x_{2i}-E(x_2|x_i\zeta_o,\delta_o))'(u_{1i}-E(u_1|x_i\zeta_o,\delta_o))\\
&-E\left(I_T(x\zeta_o)(x_2-E(x_2|x\zeta_o,\delta_o))'\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right)\sqrt{n}(\hat{\delta}-\delta_o)+o_p(1),
\end{aligned}
\tag{A.5.8}
$$

which is asymptotically normal $N(0,\Delta)$, where Δ is defined in (4.11), under the property that $\sqrt{n}(\hat{\zeta}-\zeta_o)$ is asymptotically uncorrelated with $(u_1-E(u_1|x\zeta_o,\delta_o))$. For the asymptotic distribution of the SG2LS,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{U}_n\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi})(x_{2i}-E_n(x_2|x_i\hat{\zeta}))'\omega_{ni}^{-1}(\hat{\beta},\hat{\delta})(u_{1i}-E_n(u_1|x_i\hat{\zeta}))\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)(x_{2i}-E_n(x_2|x_i\zeta_o))'\omega_{ni}^{-1}(\beta_o,\delta_o)(u_{1i}-E_n(u_1|x_i\zeta_o))\\
&\quad-\frac{1}{n}\sum_{i=1}^{n} t_n(x_i\bar{\zeta},\bar{\xi})(x_{2i}-E_n(x_2|x_i\bar{\zeta}))'\omega_{ni}^{-1}(\bar{\beta},\bar{\delta})\frac{\partial E_n(u_1|x_i\bar{\zeta})}{\partial\delta'}\cdot\sqrt{n}(\hat{\delta}-\delta_o)\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}(u_{1i}-E_n(u_1|x_i\bar{\zeta}))\frac{\partial}{\partial\delta'}[t_n(x_i\bar{\zeta},\bar{\xi})(x_{2i}-E_n(x_2|x_i\bar{\zeta}))'\omega_{ni}^{-1}(\bar{\beta},\bar{\delta})]\cdot\sqrt{n}(\hat{\delta}-\delta_o)\\
&\quad+\frac{1}{n}\sum_{i=1}^{n} t_n(x_i\bar{\zeta},\bar{\xi})(u_{1i}-E_n(u_1|x_i\bar{\zeta}))(x_{2i}-E_n(x_2|x_i\bar{\zeta}))'\frac{\partial\omega_{ni}^{-1}(\bar{\beta},\bar{\delta})}{\partial\beta'}\cdot\sqrt{n}(\hat{\beta}-\beta_o)\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}(u_{1i}-E_n(u_1|x_i\bar{\zeta}))(x_{2i}-E_n(x_2|x_i\bar{\zeta}))'\omega_{ni}^{-1}(\bar{\beta},\bar{\delta})\frac{\partial t_n(x_i\bar{\zeta},\bar{\xi})}{\partial\delta'}\cdot\sqrt{n}(\hat{\xi}-\xi_o)\\
&=C_{0,n}^{(\omega)}-E\left[I_T(x\zeta_o)(x_2-E(x_2|x\zeta_o,\delta_o))'\omega^{-1}(x\zeta_o,\delta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right]\cdot\sqrt{n}(\hat{\delta}-\delta_o)+o_p(1),
\end{aligned}
$$

where

$$
C_{0,n}^{(\omega)}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)(x_{2i}-E_n(x_2|x_i\zeta_o))'\omega_{ni}^{-1}(\beta_o,\delta_o)(u_{1i}-E_n(u_1|x_i\zeta_o)).
$$

To simplify notations, let $H_{ni} = \frac{1}{(n-1)b_n^m}\sum_{j\neq i}^n H\left(\frac{x_i\zeta_o - x_j\zeta_o}{b_n}\right)$, $R_{ni} = \frac{1}{(n-1)b_n^m}\sum_{j\neq i}^n u_{1j} H\left(\frac{x_i\zeta_o - x_j\zeta_o}{b_n}\right)$, and $S_{ni} = \frac{1}{(n-1)b_n^m}\sum_{j\neq i}^n u_{1j}^2 H\left(\frac{x_i\zeta_o - x_j\zeta_o}{b_n}\right)$. Let $H_i = B_i$, $R_i = C_i$ and $S_i = E(u_1^2|x_i\zeta_o, \delta_o)p(x_i\zeta_o|\delta_o)$ be their limits respectively. By Proposition 3, as $\lim_{n\to\infty} na_n^{2m} = \infty$, $\lim_{n\to\infty} na_n^{4s^*} = 0$, $\lim_{n\to\infty} nb_n^{2m} = \infty$, and $\lim_{n\to\infty} nb_n^{4h^*} = 0$,

$$C_{0,n}^{(\omega)} = \frac{1}{\sqrt{n}}\sum_{i=1}^n t_n(x_i\zeta_o, \xi_o)(x_{2i} - E(x_2|x_i\zeta_o, \delta_o))'\omega^{-1}(x\zeta_o, \delta_o)(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) + L_n^{(\omega)} + o_p(1),$$

where

$$\begin{aligned}
L_n^{(\omega)} = &-\frac{1}{\sqrt{n}}\sum_{i=1}^n t_n(x_i\zeta_o, \xi_o)\omega^{-1}(x\zeta_o, \delta_o)(u_{1i} - E(u_1|x_i\zeta_o, \delta_o))\frac{1}{B_i}(A_{ni} - A_i)' \\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^n t_n(x_i\zeta_o, \xi_o)\omega^{-1}(x_i\zeta_o, \delta_o)(x_{2i} - E(x_2|x_i\zeta_o, \delta_o))'\frac{1}{B_i}(C_{ni} - C_i) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n t_n(x_i\zeta_o, \xi_o)\omega^{-1}(x_i\zeta_o, \delta_o)\big[E(x_2|x_i\zeta_o, \delta_o)'(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) \\
&\quad + (x_{2i} - E(x_2|x_i\zeta_o, \delta_o))'E(u_1|x_i\zeta_o, \delta_o)\big] \cdot \frac{1}{B_i}(B_{ni} - B_i) \\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^n t_n(x_i\zeta_o, \xi_o)\omega^{-2}(x_i\zeta_o, \delta_o)(x_{2i} - E(x_2|x_i\zeta_o, \delta_o))'(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) \\
&\cdot \frac{1}{H_i}\big\{(S_{ni} - S_i) - 2E(u_1|x_i\zeta_o, \delta_o)(R_{ni} - R_i) - [E(u_1^2|x_i\zeta_o, \delta_o) - 2E(u_1|x_i\zeta_o, \delta_o)](H_{ni} - H_i)\big\}.
\end{aligned}$$

Since $E[x_{2i} - E(x_2|x_i\zeta_o, \delta_o)|x_i\zeta_o, \delta_o] = 0$, $E[u_{1i} - E(u_1|x_i\zeta_o, \delta_o)|x_i] = 0$, and the other functions in $L_n^{(\omega)}$ are functions of $x_i\zeta_o$, Proposition 4 implies that $L_n^{(\omega)}$ converges to zero in probability when $\lim_{n\to\infty} na_n^m = \infty$ and $\lim_{n\to\infty} nb_n^m = \infty$. Similarly,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\frac{\partial E_n(\hat\beta, \hat\zeta)}{\partial\delta}\hat\Lambda_n^{-1}\hat U_n \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n t_n(x_i\hat\zeta, \hat\xi)\frac{\partial E_n(y - w\hat\beta|x_i\hat\zeta)}{\partial\delta}\omega_{ni}^{-1}(\hat\beta, \hat\delta)(u_{1i} - E_n(u_1|x_i\hat\zeta)) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n t_n(x_i\zeta_o, \xi_o)\frac{\partial E_n(u_1|x_i\zeta_o)}{\partial\delta}\omega_{ni}^{-1}(\beta_o, \delta_o)(u_{1i} - E_n(u_1|x_i\zeta_o)) \\
&\quad - \frac{1}{n}\sum_{i=1}^n t_n(x_i\bar\zeta, \bar\xi)\frac{\partial E_n(y - w\bar\beta|x_i\bar\zeta)}{\partial\delta}\omega_{ni}^{-1}(\bar\beta, \bar\delta)\frac{\partial E_n(u_1|x_i\bar\zeta)}{\partial\delta'} \cdot \sqrt{n}(\hat\delta - \delta_o) \\
&\quad + \frac{1}{n}\sum_{i=1}^n (u_{1i} - E_n(u_1|x_i\bar\zeta))\frac{\partial}{\partial\delta'}\left[t(x_i\bar\zeta, \bar\xi)\frac{\partial E_n(y - w\bar\beta|x_i\bar\zeta)}{\partial\delta}\omega_{ni}^{-1}(\bar\beta, \bar\delta)\right] \cdot \sqrt{n}(\hat\delta - \delta_o) \\
&\quad + \frac{1}{n}\sum_{i=1}^n t_n(x_i\bar\zeta, \bar\xi)(u_{1i} - E_n(u_1|x_i\bar\zeta))\frac{\partial}{\partial\beta'}\left[\frac{\partial E_n(y - w\bar\beta|x_i\bar\zeta)}{\partial\delta}\omega_{ni}^{-1}(\bar\beta, \bar\delta)\right] \cdot \sqrt{n}(\hat\beta - \beta_o) \\
&\quad + \frac{1}{n}\sum_{i=1}^n (u_{1i} - E_n(u_1|x_i\bar\zeta))\frac{\partial E_n(y - w\bar\beta|x_i\bar\zeta)}{\partial\delta}\omega_{ni}^{-1}(\bar\beta, \bar\delta)\frac{\partial t(x_i\bar\zeta, \bar\xi)}{\partial\delta'} \cdot \sqrt{n}(\hat\xi - \xi_o) \\
&= C_{0,n}^{(d)} - E\left[I_T(x\zeta_o)\frac{\partial E(u_1|x_i\zeta_o, \delta_o)}{\partial\delta}\omega^{-1}(x_i\zeta_o, \delta_o)\frac{\partial E(u_1|x\zeta_o, \delta_o)}{\partial\delta'}\right] \cdot \sqrt{n}(\hat\delta - \delta_o) + o_p(1),
\end{aligned}$$

where

$$C_{0,n}^{(d)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \frac{\partial E_n(u_1|x_i\zeta_o, \delta_o)}{\partial \delta} \omega_n^{-1}(\beta_o, \delta_o)(u_{1i} - E_n(u_1|x_i\zeta_o, \delta_o)).$$

Since $\lim_{n\to\infty} na_n^{2(m+2)} = \infty$, $\lim_{n\to\infty} na_n^{4s^*} = 0$, $\lim_{n\to\infty} nb_n^{2m} = \infty$, and $\lim_{n\to\infty} nb_n^{4h^*} = 0$, Proposition 3 implies that

$$C_{0,n}^{(d)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \frac{\partial E(u_1|x_i\zeta_o, \delta_o)}{\partial \delta} \omega^{-1}(x_i\zeta_o, \delta_o)(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) + L_n^{(d)} + o_p(1),$$

where

$$\begin{aligned}
L_{0,n}^{(d)} = &-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \frac{\partial E(u_1|x_i\zeta_o, \delta_o)}{\partial \delta} \omega^{-1}(x_i\zeta_o, \delta_o) \frac{1}{B_i}(A_{ni} - A_i) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \omega^{-1}(x_i\zeta_o, \delta_o) \Bigg[\frac{\partial E(u_1|x_i\zeta_o, \delta_o)}{\partial \delta} A_i - (u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) \\
&\cdot \left\{ \frac{\partial C_i}{\partial \delta} - 2E(u_1|x_i\zeta_o, \delta_o) \frac{\partial B_i}{\partial \delta} \right\} \Bigg] \frac{1}{B_i^2}(B_{ni} - B_i) \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \omega^{-1}(x_i\zeta_o, \delta_o)(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) \frac{1}{B_i^2} \frac{\partial B_i}{\partial \delta}(C_{ni} - C_i) \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \omega^{-1}(x_i\zeta_o, \delta_o)(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) E(u_1|x_i\zeta_o, \delta_o) \frac{1}{B_i} \left(\frac{\partial B_{ni}}{\partial \delta} - \frac{\partial B_i}{\partial \delta} \right) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \omega^{-1}(x_i\zeta_o, \delta_o)(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) \frac{1}{B_i} \left(\frac{\partial C_{ni}}{\partial \delta} - \frac{\partial C_i}{\partial \delta} \right) \\
&+ \frac{2}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \omega^{-2}(x_i\zeta_o, \delta_o)(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) \frac{\partial E(u_1|x_i\zeta_o, \delta_o)}{\partial \delta} E(u_1|x_i\zeta_o, \delta_o) \frac{1}{H_i}(R_{ni} - R_i) \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \frac{\partial E(u_1|x_i\zeta_o, \delta_o)}{\partial \delta} \omega^{-2}(x_i\zeta_o, \delta_o)(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) \frac{1}{H_i}(S_{ni} - S_i) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_o, \xi_o) \omega^{-2}(x_i\zeta_o, \delta_o)(u_{1i} - E(u_1|x_i\zeta_o, \delta_o)) \frac{\partial E(u_1|x_i\zeta_o, \delta_o)}{\partial \delta} \left(S_i - 2\frac{R_i^2}{H_i} \right) \frac{1}{H_i^2}(H_{ni} - H_i).
\end{aligned}$$

Since $\frac{\partial E(u_1|x\zeta_o, \delta_o)}{\partial \delta_k} = (x - E(x|x\zeta_o, \delta_o))' \frac{\partial \zeta(\delta_o)}{\partial \delta_k} \frac{\partial E(u_1|x\zeta_o, \delta_o)}{\partial t}$, by Propositions 4 and 6, $L_{0,n}^{(d)}$ converges to zero in probability as $\lim_{n\to\infty} na_n^{m+2} = \infty$ and $\lim_{n\to\infty} nb_n^m = \infty$. Hence it follows that

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\Sigma}^{-1}\hat{U}_n \\
&= \frac{1}{\sqrt{n}}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{U}_n - \frac{1}{n}\hat{X}_2'\hat{\Lambda}_n^{-1}\frac{\partial E_n(\hat{\beta},\hat{\zeta})}{\partial\delta'}\left(\hat{V}_{n,\delta}^{-1} + \frac{\partial E_n'(\hat{\beta},\hat{\zeta})}{\partial\delta}\hat{\Lambda}_n^{-1}\frac{\partial E_n(\hat{\beta},\hat{\zeta})}{\partial\delta'}\right)^{-1}\frac{\partial E_n'(\hat{\beta},\hat{\zeta})}{\partial\delta}\hat{\Lambda}_n^{-1}\hat{U}_n \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)(x_{2i} - E(x_2|x_i\zeta_o,\delta_o))'\omega^{-1}(x\zeta_o,\delta_o)(u_{1i} - E(u_1|x_i\zeta_o,\delta_o)) \\
&\quad - E\left[I_T(x\zeta_o)(x_2 - E(x_2|x\zeta_o,\delta_o))'\omega^{-1}(x\zeta_o,\delta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right] \\
&\quad \cdot\left(V_\delta^{-1} + \frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta}\omega^{-1}(x\zeta_o,\delta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right)^{-1} \\
&\quad \cdot\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)\frac{\partial E(u_1|x_i\zeta_o,\delta_o)}{\partial\delta}\omega^{-1}(x_i\zeta_o,\delta_o)(u_{1i} - E(u_1|x_i\zeta_o,\delta_o)) + V_\delta^{-1}\sqrt{n}(\hat{\delta}-\delta_o)\right\} + o_p(1) \\
&\xrightarrow{D} N(0,\Gamma),
\end{aligned}
\tag{A.5.9}
$$

where Γ is in (5.13).

For the instrumental variable estimator in (5.18),

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}P'\hat{U}_n \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)p_i'(u_{1i} - E_n(u_1|x_i\zeta_o)) - \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\bar{\zeta},\bar{\xi})p_i'\frac{\partial E_n(u_1|x_i\bar{\zeta})}{\partial\delta'}\sqrt{n}(\hat{\delta}-\delta_o) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} p_i'(u_{1i} - E_n(u_1|x_i\bar{\zeta}))\frac{\partial t_n(x_i\bar{\zeta},\bar{\xi})}{\partial\delta'}\sqrt{n}(\hat{\delta}-\delta_o) + \frac{1}{n}\sum_{i=1}^{n} p_i'(u_{1i} - E_n(u_1|x_i\bar{\zeta}))\frac{\partial t_n(x_i\bar{\zeta},\bar{\xi})}{\partial\xi'}\sqrt{n}(\hat{\xi}-\xi_o) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)p_i'(u_{1i} - E_n(u_1|x_i\zeta_o)) - E\left[I_T(x\zeta_o)p'\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right]\sqrt{n}(\hat{\delta}-\delta_o) + o_p(1).
\end{aligned}
\tag{A.5.10}
$$

Since $\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta_k} = (x - E(x|x\zeta_o,\delta_o))'\frac{\partial\zeta(\delta_o)}{\partial\delta_k}\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial t}$, it follows that $E\left[E(p'|x\zeta_o,\delta_o)\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right] = 0$ and

$$
E\left[I_T(x\zeta_o)p'\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right] = E\left[I_T(x\zeta_o)(p - E(p|x\zeta_o,\delta_o))'\frac{\partial E(u_1|x\zeta_o,\delta_o)}{\partial\delta'}\right].
$$

On the other hand,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)p_i'(u_{1i} - E_n(u_1|x_i\zeta_o)) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)p_i'(u_{1i} - E(u_1|x_i\zeta_o,\delta_o)) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)p_i'\frac{1}{B_i}(C_{ni} - C_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)p_i'E(u_1|x_i\zeta_o,\delta_o)\frac{1}{B_i}(B_{ni} - B_i) + o_p(1).
\end{aligned}
$$

With a high order kernel of order s^* such that $\lim_{n\to\infty} a_n^{2s^*} = 0$,[12] Proposition 5 implies

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)p_i'\frac{1}{B_i}(C_{ni}-C_i)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_T(x_i\zeta_o)u_{1i}E(p_i'|x_i\zeta_o,\delta_o)-E[I_T(x_i\zeta_o)E(u_1|x_i\zeta_o,\delta_o)E(p_i'|x_i\zeta_o,\delta_o)]\}+o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)p_i'E(u_1|x_i\zeta_o,\delta_o)\frac{1}{B_i}(B_{ni}-B_i)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_T(x_i\zeta_o)E(p_i'|x_i\zeta_o,\delta_o)E(u_1|x_i\zeta_o,\delta_o)-E[I_T(x_i\zeta_o)E(p_i'|x_i\zeta_o,\delta_o)E(u_1|x_i\zeta_o,\delta_o)]\}+o_p(1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_o,\xi_o)p_i'(u_{1i}-E_n(u_1|x_i\zeta_o))\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_T(x_i\zeta_o)(p_i'-E(p_i'|x_i\zeta_o,\delta_o))(u_{1i}-E(u_1|x_i\zeta_o,\delta_o))+o_p(1).
\end{aligned}\tag{A.5.11}$$

[12] This rate requirement implies $\lim_{n\to\infty} na_n^{4s^*} = 0$ in Assumption 2. This stronger requirement is needed only here. It guarantees that the asymptotic bias of the following term will converge to zero. Using $p_i - E_n(p|x_i\hat{\zeta})$ instead of p_i will eliminate such an asymptotic bias and this stronger rate requirement will not be needed. This indicates the advantage of using $p_i - E_n(p|x_i\hat{\zeta})$ as an instrumental variable instead of p_i.

References

1. Amemiya, T. [1978], "The Estimation of a Simultaneous Equation Generalized Probit Model", *Econometrica* **46**: 1193-1206.

2. Amemiya, T. [1983], "A Comparison of the Amemiya GLS and the Lee-Maddala-Trost G2SLS in a Simultaneous-equations Tobit Model", *Journal of Econometrics* **23**: 295-300.

3. Bierens, H.J. [1985], "Kernel Estimators of Regression Functions", in *Advances in Econometrics, Fifth World Congress*, edited by T.W. Bewley, Vol. 1: 99-144.

4. Chamberlain, G.[1986], "Asymptotic Efficiency in Semi-parametric Models with Censoring", *Journal of Econometrics* **32**: 189-218.

5. Cosslett, S.[1987], "Efficiency Bounds for Distribution-Free Estimators of the Censored Regression Models", *Econometrica* **55**: 559-585.

6. Heckman, J.J. [1979], "Sample Selection Bias as a Specification Error", *Econometrica* **47**: 153-162.

7. Ichimura, H. [1987], "Estimation of Single Index Models", Ph.D. thesis, Department of Economics, MIT.

8. Ichimura, H. and L.F. Lee [1988], "Semiparametric Estimation of Multiple Index Models: Single Equation Estimation", forthcoming in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, edited by W.A. Marnett, J. Powell and G. Tauchen.

9. Klein, R.W. and R.H. Spady [1987], "An Efficient Semiparametric Estimator for Discrete Choice Models", Bellcore Economics Discussion Paper No. 67, Morristown, New Jersey.

10. Lee, L.F. [1981], "Simultaneous Equations Models with Discrete Endogenous Variables", Chapter 9 in *Structural Analysis of Discrete Data and Econometric Applications*, edited by C.F. Manski and D. McFadden, The MIT Press, Cambridge, Massachusetts.

11. Lee, L.F. [1990], "Efficient Semiparametric Scoring Estimation of Sample Selection Models", Discussion Paper No. 255, Center for Economic Research, U. of Minnesota.

12. Lee, L.F., G.S. Maddala and R.P. Trost [1980], "Asymptotic Covariance Matrices of Two-stage Probit and Two-stage Tobit methods for Simultaneous Equations Models with Selectivity", *Econometrica* **48**: 491-503.

13. Newey, W.K. [1987], "Efficient Estimation of Limited Dependent Variable Models with Endogenous

Explanatory Variables", *Journal of Econometrics* **36**: 231-250.

14. Powell, J.L. [1987], "Semiparametric Estimation of Bivariate Latent Variable Models", Discussion Paper No. 8704, Social Systems Research Institue, U. of Wisconsin.

15. Powell, J.L., J.H. Stock, and T.M. Stoker [1989], "Semiparametric Estimation of Index Coefficients", *Econometrica* **57**: 1403-1430.

16. Robinson, P.M. [1988], "Root-N-Consistent Semiparametric Regession", *Econometrica* **56**: 931-954.

17. Serfling R.J. [1980], *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, New York.

18. Silverman, B.W. [1986], *Density Estimation for Statistics and Data Analysis*, Chapman and Hall Ltd., New York.

19. Stoker, T.M. [1986], "Consistent Estimation of Scaled Coefficients", *Econometrica* **54**: 1461-1481.