ELSEVIER Journal of Econometrics 63 (1994) 341–388

JOURNAL OF Econometrics

Semiparametric instrumental variable estimation of simultaneous equation sample selection models

Lung-Fei Lee

Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220, USA

(Received June 1991; final version received May 1993)

Abstract

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 with normal disturbances. The necessary order condition for identification in the semiparametric model corresponds to the overidentification 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 two-stage least squares estimator is shown to be asymptotically efficient in a class of semiparametric instrumental variable estimators.

Key words: Semiparametric model; Sample selection; Simultaneity; Index model; Identification; Instrumental variables; Asymptotic efficiency
JEL classification: C14; C34

1. Introduction

For the estimation of simultaneous equation sample selection models with parametric (normal) disturbances, several methods are available in the

I appreciate having financial support from NSF under grant no. SES-9010516 and SES-9296071 for my research. I would like to thank an anonymous referee for valuable comments and suggestions. The issues in this article are partially motivated by an empirical project under way with my colleagues Professors Mark Rosenzweig and Mark M. Pitt.

SSDI 0304407693015713

econometric literature, e.g., Lee, Maddala, and Trost (1980), Lee (1981), Amemiya (1983), Newey (1987), and Blundell and Smith (1989). 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), Newey (1987), and Blundell and Smith (1989).

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 on 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.

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:

$$y^* = z^*\alpha_0 + xJ\gamma_0 + u_1, \tag{2.1}$$

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 row 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

$$d^* = x\zeta_0 + \varepsilon, \tag{2.3}$$

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 $x\zeta_0$. Such a framework is slightly more general than the case where the disturbances are independent of x.

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

$$y_i = z_i\alpha_0 + x_iJ\gamma_0 + \mathrm{E}(u_{1i} \mid x_i\zeta_0, d_i^* > 0) + u_{di}, \tag{2.4}$$

where

$$u_{di} = u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0, d_i^* > 0). \tag{2.5}$$

Let $K(\cdot)$ be a kernel function with a bandwidth parameter a_n (Silverman, 1986, or Bierens, 1985). Let $w = (z, xJ)$ and $\beta_0 = (\alpha_0, \gamma_0)$. For any possible value (β, ζ) of (β_0, ζ_0), the conditional expectation function $\mathrm{E}(y - w\beta \mid x\zeta = x_i\zeta, d^* > 0)$ of $y - w\beta$, conditional on the random variables $x\zeta$ and $d^* > 0$ evaluated at the point $x_i\zeta$, can be estimated by the following nonparametric regression function:

$$\mathrm{E}_n(y - w\beta \mid x_i\zeta) = \frac{\sum_{j \neq i}^{n} (y_j - w_j\beta) K\left(\frac{x_i\zeta - x_j\zeta}{a_n}\right)}{\sum_{j \neq i}^{n} K\left(\frac{x_i\zeta - x_j\zeta}{a_n}\right)}, \tag{2.6}$$

where n is the sample size for the observations of (y, z, x) conditional on $d^* > 0$ (Nadaraja, 1964; Watson, 1964). Given a $\sqrt{n}$-consistent estimate $\hat{\zeta}$ of ζ, Powell (1987) has proposed an instrumental variable method for the estimation of β_0 from the following equation:

$$y_i - \mathrm{E}_n(y \mid x_i\hat{\zeta}) = (w_i - \mathrm{E}_n(w \mid x_i\hat{\zeta}))\beta_0 + \hat{u}_{ni}, \tag{2.7}$$

where

$$\mathrm{E}_n(s \mid x_i\zeta) = \sum_{j \neq i}^{n} s_j K\left(\frac{x_i\zeta - x_j\zeta}{a_n}\right) \Big/ \sum_{j \neq i}^{n} K\left(\frac{x_i\zeta - x_j\zeta}{a_n}\right)$$

for any random variable s (see also Robinson, 1988). Instrumental variable methods require instrumental variables for $w_i - \mathrm{E}_n(w \mid x_i\hat{\zeta})$ (Powell, 1987). A simple instrumental variable estimator with instrumental variables p can be

$$\left(\sum_{i=1}^{n} p_i'(w_i - \mathrm{E}_n(w \mid x_i\hat{\zeta}))\right)^{-1} \sum_{i=1}^{n} p_i'(y_i - \mathrm{E}_n(y \mid x_i\hat{\zeta})). \tag{2.8}$$

However, due to the technical difficulty of handling the denominator in the nonparametric regression function in (2.6), some modifications are needed to overcome the 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). 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; Ichimura and Lee, 1991). In this article, 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 (see Sections A.2 and A.3 of the Appendix). Suppose that $t_n(x_i\hat{\zeta}, \hat{\xi}_n)$ is a trimming indicator with value 0 when $x_i\hat{\zeta}$ is deleted, where $\hat{\xi}_n$ is a vector of sample quantiles. A simple unweighted instrumental variable estimator is

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

The trimming procedure is preferred to the Powell's procedure. The weighting scheme in Powell (1987) has nothing to do with the variance of the disturbance $\hat{u}_{ni}$ in (2.7). With the trimming procedure, weighting estimation method which incorporates heteroskedastic variances can be introduced in subsequent sections.

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 polychotomous 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; Ichimura and Lee, 1991). $x\zeta$ 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 (2.2), (2.3), and (3.1) has been considered in Ichimura (1987), Chamberlain (1986), Powell (1987), and Ichimura and Lee (1991). Conditional on $d^* > 0$ and x,

$$\mathrm{E}(y \mid x, d^* > 0) = x\pi_1 + \mathrm{E}(v_1 \mid x\zeta_0, x\zeta_0 > -\varepsilon), \tag{3.2}$$

and

$$\mathrm{E}(z \mid x, d^* > 0) = x\Pi_2 + \mathrm{E}(v_2 \mid x\zeta_0, x\zeta_0 > -\varepsilon). \tag{3.3}$$

As shown in Ichimura (1987) for the single-index model, ζ_0 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_0 cannot even be identified up to a scale. Therefore, we consider only the model where a relevant continuous exogenous variable is present in the index $x\zeta_0$. As the coefficients in the index can only be identified up to a scale, normalization is needed. A convenient normalization (Ichimura, 1987) is to set the coefficient of a continuous exogenous variable to the 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 $x\zeta_0$ contain the same set of variables x and they cannot be distinguished from each other in (3.2). Similarly, this is so for $x\Pi_2$ and $x\zeta_0$. This identification problem has been studied in Ichimura and Lee (1991) 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

$$x\zeta_0 = x_{(1)} + x_{(2)}\delta_0,$$

where $x_{(1)}$ is a relevant continuous exogenous variable in $x = (x_{(1)}, x_{(2)})$ and $x_{(2)}$ is the remaining subvector of x. Conformably, $x\pi_1 = x_{(1)}\pi_{11} + x_{(2)}\pi_{12}$, $x\Pi_2 = x_{(1)}\pi'_{21} + x_{(2)}\Pi_{22}$, and $x\gamma_0 = x_{(1)}\gamma_{0,1} + x_{(2)}\gamma_{0,2}$. The reduced form equation (3.1) can be rewritten into

$$y^* = x_{(2)}\pi_1^* + v_1^*, \tag{3.4}$$

where $\pi_1^* = \pi_{12} - \delta_0\pi_{11}$ and $v_1^* = v_1 + x\zeta_0\pi_{11}$. Similarly,

$$z^* = x_{(2)}\Pi_2^* + v_2^*, \tag{3.5}$$

where $\Pi_2^* = \Pi_{22} - \delta_0 \pi'_{21}$ and $v_2^* = v_2 + x\zeta_0 \pi'_{21}$. It follows that

$$E(y \mid x, d^* > 0) = x_{(2)}\pi_1^* + E(v_1^* \mid x\zeta_0, d^* > 0) \tag{3.6}$$

and

$$E(z \mid x, d^* > 0) = x_{(2)}\Pi_2^* + E(v_2^* \mid x\zeta_0, d^* > 0). \tag{3.7}$$

The index $x\zeta_0$ is distinguishable from $x_{(2)}\pi_1^*$ and $x_{(2)}\Pi_2^*$, because $x_{(1)}$ appears only in $x\zeta_0$ but neither in $x_{(2)}\pi_1^*$ nor in $x_{(2)}\Pi_2^*$. The transformed parameters π_1^* and Π_2^* and δ_0 are identifiable.

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

$$\begin{aligned} y^* &= z^*\alpha_0 + x\bar{\gamma}_0 + u_1 \\ &= (x_{(2)}\Pi_2^* + v_2^*)\alpha_0 + x_{(1)}\gamma_{0,1} + x_{(2)}\gamma_{0,2} + u_1 \\ &= x_{(2)}(\Pi_2^*\alpha_0 - \gamma_{0,1}\delta_0 + \gamma_{0,2}) + u_1^*, \end{aligned} \tag{3.8}$$

where $u_1^* = u_1 + v_2^*\alpha_0 + \gamma_{0,1}x\zeta_0$ and $\bar{\gamma}_0 = (\gamma'_{0,1}, \gamma'_{0,2})'$ is the vector of coefficients of x before exclusion restrictions on x are imposed. Comparing (3.8) with (3.4),

$$\pi_1^* = \Pi_2^*\alpha_0 - \gamma_{0,1}\delta_0 + \gamma_{0,2}. \tag{3.9}$$

From (3.9), we see that the identification of the structural parameters α_0, $\gamma_{0,1}$, and $\gamma_{0,2}$ requires restrictions on the structural equation (2.1). With exclusion restrictions in (2.1), we have $\bar{\gamma}_0 = J\gamma_0$ and (3.9) becomes $\pi_1^* = [\Pi_2^*, (-\delta_0, I)J](\alpha'_0, \gamma'_0)'$. From this relation, we see that the rank identification of the structural parameters in $y = z\alpha_0 + xJ\gamma_0 + u_1$ is that $[\Pi_2^*, (-\delta_o, I)J]$ has full (column) rank. For the order identification condition, it is convenient to consider separately the two cases of exclusion restrictions 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 $\gamma_{0,1} = 0$. 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., $\gamma_{0,2} = (\gamma'_{0,21}, 0)'$ where $\gamma_{0,21}$ is of dimension k_1. Conformably, let $\pi_1^* = (\pi_{11}^{*\prime}, \pi_{12}^{*\prime})'$, $\Pi_2^* = (\Pi_{21}^{*\prime}, \Pi_{22}^{*\prime})'$, and $\delta_0 = (\delta'_{0,1}, \delta'_{0,2})'$. Since $\gamma_{0,1} = 0$, (3.9) is equivalent to

$$\pi_{11}^* = \Pi_{21}^*\alpha_0 + \gamma_{0,21} \tag{3.10}$$

and

$$\pi_{12}^* = \Pi_{22}^*\alpha_0. \tag{3.11}$$

It follows from (3.10) and (3.11) that the rank condition is equivalent to rank $\Pi_{22}^* = G_1$. The necessary order condition is $K - 1 - k_1 \geqslant G_1$, 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). Eq. (3.9) is now equivalent to

$$\pi_{11}^* = \Pi_{21}^* \alpha_0 - \gamma_{0,1}\delta_{0,1} + \gamma_{0,21} \tag{3.12}$$

and

$$\pi_{12}^* = \Pi_{22}^* \alpha_0 - \gamma_{0,1}\delta_{0,2}. \tag{3.13}$$

The rank identification for this case is rank $[\Pi_{22}^*, \ -\delta_{0,2}] = G_1 + 1$, and the necessary order condition is $K - 1 - k_1 \geqslant G_1 + 1$. In any event, the identification condition is stronger than the identification condition for the classical simultaneous equation model. The exact identification of (2.1) for the classical model becomes underidentification for the semiparametric model. The order identification condition for the semiparametric model corresponds to the overidentification condition in the classical simultaneous equation model. The stronger condition for the identification of the semiparametric model is apparently due to the addition of a sample selection bias term of an unknown form in the bias corrected structural equation. Exogenous variables which are excluded from the structural equation (before bias correction), but appear in the selection bias term through the index $x\zeta_0$, identify the selection bias term. Intuitively, the included bias correction term introduces excluded exogenous variables back into the (bias) corrected structural equation, and the effective number of the included exogenous variables in this equation is the number of originally included exogenous variables plus one. Therefore, the order condition for identification requires stronger exclusion restrictions than the classical simultaneous equation model. For the parametric simultaneous equation sample selection model under normal disturbances, the bias correction term has a known nonlinear functional form, which introduces nonlinear restriction into the bias corrected structural equation. The nonlinear bias correction term with a particular known form due to normality helps identification even though the excluded exogenous variables are implicitly introduced back into the bias corrected structural equation. If the bias correction term in a parametric simultaneous equation sample selection model were a linear function of the index (Olsen, 1980), then stronger identification condition similar to the one of our semiparametric model would be needed. Putting it in another way, stronger identification condition is needed for our semiparametric model because it does not exclude the parametric specification of Olsen (1980).

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, 1991). Suppose $x\zeta$ contains m indices. With normalization, $x\zeta = (x_{(11)} + x_{(2)}\delta_{(1)}, \ldots,$

$x_{(1m)} + x_{(2)}\delta_{(m)}$), where $x_{(11)}, \ldots, x_{(1m)}$ are m distinct continuous variables not contained in the subvector $x_{(2)}$. For this general model, $\delta = [\delta_{(1)}, \ldots, \delta_{(m)}]$ is now a matrix. The rank identification condition is that $[\Pi_2^*, (-\delta_0, I)J]$ has full (column) rank.

4. Semiparametric two-stage least squares estimation

For the classical simultaneous equation model, the most 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:

$$\hat{X}_2 = \begin{pmatrix} t_n(x_1\hat{\zeta}, \hat{\xi}_n)(x_{(2)1} - \mathrm{E}_n(x_{(2)} \mid x_1\hat{\zeta})) \\ \vdots \\ t_n(x_n\hat{\zeta}, \hat{\xi}_n)(x_{(2)n} - \mathrm{E}_n(x_{(2)} \mid x_n\hat{\zeta})) \end{pmatrix},$$

$$\hat{W} = \begin{pmatrix} t_n(x_1\hat{\zeta}, \hat{\xi}_n)(w_1 - \mathrm{E}_n(w \mid x_1\hat{\zeta})) \\ \vdots \\ t_n(x_n\hat{\zeta}, \hat{\xi}_n)(w_n - \mathrm{E}_n(w \mid x_n\hat{\zeta})) \end{pmatrix},$$

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

where $t_n(x\hat{\zeta}, \hat{\xi}_n)$ is a smooth quantile trimming function of $x\hat{\zeta}$ introduced in Sections A.2 and A.3 of the Appendix. The trimming function is differentiable with respect to its arguments $x\hat{\zeta}$ and $\hat{\xi}_n$ so that one can easily investigate the impact of the randomness of $\hat{\delta}$ and the sample quantiles $\hat{\xi}_n$ on the asymptotic properties of the derived estimators. A semiparametric two-stage least squares estimator (S2LS) is

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

This estimator can be interpreted as being derived from a two-stage estimation 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 $z - \mathrm{E}_n(z \mid x\hat{\zeta})$ are used as instrumental variables for $z - \mathrm{E}_n(z \mid x\hat{\zeta})$. 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 $x_{(1)} - \mathrm{E}_n(x_{(1)} \mid x\hat{\zeta})$ will be used as an instrumental variable for $x_{(1)} - \mathrm{E}_n(x_{(1)} \mid x\hat{\zeta})$ for the estimation of (2.7). This two-stage estimator has a two-stage semiparametric least squares interpretation. To see these, define an auxiliary equation for $x_{(1)}$:

$$x_{(1)} = -x_{(2)}\delta_0 + v_{1,*}, \tag{4.2}$$

where $v_{1,*} = x\zeta_0$. Since $w^* = (z^*, xJ)$, (3.5) and (4.2) imply that

$$w^* = x_{(2)}[\Pi_2^*, (-\delta_0, I)J] + v^*, \tag{4.2'}$$

where $v^* = (v_2^*, (v_{1,*}, 0)J)$, and conditional on $d^* > 0$ and x,

$$w = x_{(2)}[\Pi_2^*, (-\delta_0, I)J] + \mathrm{E}(v^* \mid x, d^* > 0) + v_d, \tag{4.3}$$

where $v_d = v^* - \mathrm{E}(v^* \mid x, d^* > 0)$. Similarly, as in (2.7), since the distribution of v^* is a function of $x\zeta_0$,

$$w_i - \mathrm{E}_n(w \mid x_i\hat{\zeta}) = (x_{(2)i} - \mathrm{E}_n(x_{(2)} \mid x_i\hat{\zeta}))\Pi_w + \hat{v}_{ni}, \tag{4.4}$$

where $\Pi_w = [\Pi_2^*, (-\delta_0, I)J]$. A semiparametric least squares (SLS) estimator of Π_w with trimming function t_n is $\hat{\Pi}_w = (\hat{X}'_2\hat{X}_2)^{-1}\hat{X}'_2\hat{W}$. It follows that $\hat{\beta}_{\mathrm{S2LS}} = [(\hat{X}_2\hat{\Pi}_w)'(\hat{X}_2\hat{\Pi}_w)]^{-1}(\hat{X}_2\hat{\Pi}_w)'\hat{Y}$, 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.

The propositions in Appendix Section A.4 can be used to derive the asymptotic properties of our estimator. Some of the detailed derivations will be referred to Appendix Section A.5. Since sample observations for (2.4) are available only after selection, all expectations will be taken as conditional expectations conditional on $d^* > 0$. To simplify notation, the conditional argument $d^* > 0$ will be suppressed. Thus $\mathrm{E}(\cdot \mid x\zeta_0)$ stands for $\mathrm{E}(\cdot \mid x\zeta_0, d^* > 0)$ in subsequent presentation and the Appendix. Proposition 1 in Appendix Section A.4 implies that

$$\frac{1}{n}\hat{X}'_2\hat{X}_2 = \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}_n)(x_{(2)i} - \mathrm{E}_n(x_{(2)} \mid x_i\hat{\zeta}))'(x_{(2)i} - \mathrm{E}_n(x_{(2)} \mid x_i\hat{\zeta})) \tag{4.5}$$

$$\xrightarrow{p} C,$$

where

$$C = \mathrm{E}[I_T(x\zeta_0)(x_{(2)} - \mathrm{E}(x_{(2)}\,|\,x\zeta_0))'(x_{(2)} - \mathrm{E}(x_{(2)}\,|\,x\zeta_0))], \tag{4.6}$$

and I_T is the indicator function of T, where $T = [\xi_p(\delta_0), \xi_{(1-p)}(\delta_0)]$ with $\xi_p(\delta_0)$ and $\xi_{(1-p)}(\delta_0)$ being respectively the pth and $(1-p)$th quantiles of $x\zeta_0$. On the other hand, since, from (4.2)', $v^* = [v_2^*, (x\zeta_0, 0)J]$ and $\mathrm{E}(v_2^*\,|\,x) = \mathrm{E}(v_2^*\,|\,x\zeta_0)$, it follows that $\mathrm{E}[v^* - \mathrm{E}(v^*\,|\,x\zeta_0)|x] = 0$ and, by Proposition 1 in Appendix Section A.4,

$$\begin{aligned}\frac{1}{n}\hat{W}'\hat{X}_2 &= \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}_n)(w_i - \mathrm{E}_n(w\,|\,x_i\hat{\zeta}))'(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\hat{\zeta}))\\ &= \Pi'_w \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}_n)(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\hat{\zeta}))'(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\hat{\zeta}))\\ &\quad + \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}_n)(v_i^* - \mathrm{E}_n(v^*|x_i\hat{\zeta}))'(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\hat{\zeta}))\\ &\xrightarrow{p} \Pi'_w C.\end{aligned} \tag{4.7}$$

It follows that under the assumption [Assumption 1(6)] 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. Let $\hat{U}_n = (t_n(x_1\hat{\zeta}, \hat{\xi}_n)\hat{u}_{n1}, \cdots, t_n(x_n\hat{\zeta}, \hat{\xi}_n)\hat{u}_{nn})'$. Eqs. (2.7) and (4.1) imply that

$$\hat{\beta}_{\mathrm{S2LS}} - \beta_0 = [\hat{W}'\hat{X}_2(\hat{X}'_2\hat{X}_2)^{-1}\hat{X}'_2\hat{W}]^{-1}\hat{W}'\hat{X}_2(\hat{X}'_2\hat{X}_2)^{-1}\hat{X}'_2\hat{U}_n. \tag{4.9}$$

Since $\mathrm{plim}_{n\to\infty}(1/n)\hat{X}'_2\hat{U}_n = \mathrm{E}(I_T(x\zeta_0)[x_{(2)} - \mathrm{E}(x_{(2)}|x\zeta_0)]'[u_1 - \mathrm{E}(u_1\,|\,x\zeta_0)]) = 0$, $\hat{\beta}_{\mathrm{S2LS}}$ is a consistent estimator of β_0.

The asymptotic distribution of $\hat{\beta}_{\mathrm{S2LS}}$ can be derived from (4.9). Let $\omega(x_i\zeta_0, \delta_0)$ denote the conditional variance of u_{di}. Let $\nabla\mathrm{E}(u_1\,|x\zeta_0)$ denote the first-order derivative of $\mathrm{E}(u_1\,|\,x\zeta_0)$ with respect to the argument $x\zeta_0$. As shown in (A.5.1) of the Appendix,

$$\begin{aligned}\frac{1}{\sqrt{n}}\hat{X}'_2\hat{U}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}_n)(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\hat{\zeta}))'(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\hat{\zeta}))\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)(x_{(2)i} - \mathrm{E}(x_{(2)i}\,|\,x_i\zeta_0))'(u_{1i} - \mathrm{E}(u_{1i}\,|\,x_i\zeta_0))\\ &\quad - \mathrm{E}(I_T(x\zeta_0)(x_{(2)} - \mathrm{E}(x_{(2)}\,|\,x\zeta_0))'(x_{(2)} - \mathrm{E}(x_{(2)}\,|\,x\zeta_0))\\ &\quad \times \nabla\mathrm{E}(u_1\,|\,x\zeta_0))\sqrt{n}(\hat{\delta} - \delta_0) + o_p(1).\end{aligned} \tag{4.10}$$

The first term on the right-hand side of (4.10) captures the impact of the disturbance u_d in the structural equation (2.4) on the limiting distribution of $\hat{\beta}_{S2LS}$, and the second component captures the randomness of the first-stage estimator $\hat{\delta}$ of δ_0 on the limiting distribution of $\hat{\beta}_{S2LS}$. Under the assumption [Assumption 1(4) in the Appendix] that $\sqrt{n}(\hat{\delta} - \delta_0)$ is asymptotically normal, $N(0, V_\delta)$, and is asymptotically uncorrelated with $u_1 - E(u_1 \mid x\zeta_0)$, the Lindeberg–Feller and multivariate central limit theorems imply that

$$\frac{1}{\sqrt{n}} \hat{X}'_2 \hat{U}_n \xrightarrow{D} N(0, \Delta),$$

where

$$\Delta = E[I_T(x\zeta_0)(x_{(2)} - E(x_{(2)} \mid x\zeta_0))'\omega(x\zeta_0, \delta_0)(x_{(2)} - E(x_{(2)} \mid x\zeta_0))] + DV_\delta D' \tag{4.11}$$

and

$$D = E[I_T(x\zeta_0)(x_{(2)} - E(x_{(2)} \mid x\zeta_0))'(x_{(2)} - E(x_{(2)} \mid x\zeta_0)) \nabla E(u_1 \mid x\zeta_0)]. \tag{4.12}$$

The above assumption [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 Lee (1991) for polychotomous choice models. In general, for a choice model with L alternatives and T sample observations where $T > n$, if

$$\sqrt{n}(\hat{\delta} - \delta_0) = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} f(x_i, I_{1,i}, \ldots, I_{L,i}) + o_p(1),$$

where I_l is a dichotomous choice indicator for the l alternative and $f(x, I_1, \ldots, I_L)$ is some measurable function with zero mean, is asymptotically normal, Assumption 1(4) will be satisfied. This is so, because

$$\begin{aligned} E^*&\left\{ \sum_{j=1}^{T} f(x_j, I_{1,j}, \ldots, I_{L,j})(u_{1i} - E(u_{1i} \mid x_i\zeta_0)) \right\} \\ &= E^*\{f(x_i, I_{1,i}, \ldots, I_{L,i})(u_{1i} - E(u_{1i} \mid x_i\zeta_0))\} \\ &= E^*\{E[u_{1i} - E(u_{1i} \mid x_i\zeta_0) \mid x_i, I_{1i} = 1] f(x_i, I_{1,i}, \ldots, I_{L,i})\}, \\ &= 0, \end{aligned}$$

where E* denotes the unconditional expectation taken with the whole population (not just the subpopulation with $d^* > 0$).

In conclusion,

$$\sqrt{n}(\hat{\beta}_{S2LS} - \beta_0) \xrightarrow{D} N(0, \Omega), \tag{4.13}$$

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:

$$\begin{aligned}\hat{u}_{ni} &= u_{1i} - E_n(u_1 \mid x_i\hat{\zeta}) \\ &= (u_{1i} - E(u_{1i} \mid x_i\zeta_0)) - (E_n(u_1 \mid x_i\hat{\zeta}) - E_n(u_1 \mid x_i\zeta_0)) \\ &\quad - (E_n(u_1 \mid x_i\zeta_0) - E(u_{1i} \mid x_i\zeta_0)).\end{aligned} \tag{5.1}$$

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 $E_n(u_1 \mid x_i\zeta_0)$ by replacing ζ_0 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; Lee, Maddala, and Trost, 1980).[1] This is also the case for the semiparametric model in (2.1) and (2.3). 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 details, in particular, Propositions 4 and 6 of Section A.4).

[1] 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)].

The variance of u_{di} in (2.4) is a function of $x_i\zeta_0$. It can be estimated nonparametrically. Let H be a kernel function with a bandwidth b_n.[2] Define

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

where $\hat{\beta}$ is an initial consistent estimate of β [e.g., $\hat{\beta}_{\mathrm{S2LS}}$ in (4.1)], $\hat{\omega}_n(\hat{\beta}, \hat{\delta})$ provides a nonparametric estimate of the variance $\omega(x_i\zeta_0, \delta_0)$ of u_{di} at $x_i\zeta_0$. The above arguments suggest the following covariance matrix:

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

where $\hat{\Lambda}_n$ is a diagonal matrix with diagonal elements $\hat{\omega}_{ni}(\hat{\beta}, \hat{\delta})$, $i = 1, \ldots, n$, $\hat{V}_{n,\delta}/n$ is a consistent estimate of the limiting covariance matrix V_δ of $\sqrt{n}(\hat{\delta} - \delta_0)$, and

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

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:

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

For the parametric simultaneous equation sample selection model, several generalized two-stage least squares estimators (G2LS) have been introduced (Lee, 1981; Amemiya, 1983). For the semiparametric model, the following estimator is a semiparametric generalized two-stage least squares estimator (SG2LS):

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

[2] 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.

An alternative SG2LS estimator is

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

It can be shown that these two estimators have the same asymptotic distribution.[3] 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),

$$\hat{\beta}_{SG} - \beta_0 = [\hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{W}]^{-1} \times \hat{W}'\hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{U}_n \tag{5.8}$$

and

$$\tilde{\beta}_{SG} - \beta_0 = [\hat{W}'\hat{\Sigma}^{-1}\hat{X}_2(\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{W}]^{-1} \times \hat{W}'\hat{\Sigma}^{-1}\hat{X}_2(\hat{X}_2'\hat{\Sigma}^{-1}\hat{X}_2)^{-1}\hat{X}_2'\hat{\Sigma}^{-1}\hat{U}_n. \tag{5.9}$$

Proposition 1 of Appendix Section A.4 implies that

$$\frac{1}{n}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{X}_2 = \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}_n)(x_{(2)i} - \mathrm{E}_n(x_{(2)}|x_i\hat{\zeta}))' \times \omega_{ni}^{-1}(\hat{\beta}, \hat{\delta})(x_{(2)i} - \mathrm{E}_n(x_{(2)}|x_i\hat{\zeta})) \xrightarrow{p} C_\omega,$$

where

$$C_\omega = \mathrm{E}[I_T(x\zeta_0)(x_{(2)} - \mathrm{E}(x_{(2)}|x\zeta_0))'\omega^{-1}(x\zeta_0, \delta_0)(x_{(2)} - \mathrm{E}(x_{(2)}|x\zeta_0))]$$

and

$$\frac{1}{n}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{W} = \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}_n)(x_{2i} - \mathrm{E}_n(x_{(2)}|x_i\hat{\zeta}))'\omega_{ni}^{-1}(\hat{\beta}, \hat{\delta})(w_i - \mathrm{E}_n(w|x_i\hat{\zeta})) \xrightarrow{p} C_\omega \Pi_w.$$

Furthermore, let $\partial\mathrm{E}(u_1 | x\zeta_0, \delta_0)/\partial\delta$ denote $\partial\mathrm{E}(u_1 | x\zeta)/\partial\delta|_{\delta=\delta_0}$. An explicit expression for this derivative [Appendix Eq. (A.4.11)] is

$$\frac{\partial\mathrm{E}(u_1 | x\zeta_0, \delta_0)}{\partial\delta} = (x_{(2)} - \mathrm{E}(x_{(2)} | x\zeta_0))'\nabla\mathrm{E}(u_1 | x\zeta_0).$$

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

Proposition 1 implies that

$$\frac{1}{n}\hat{X}'_2\hat{\Lambda}_n^{-1}\frac{\partial \mathrm{E}_n(\hat{\beta},\hat{\zeta})}{\partial \delta'} \xrightarrow{p} D_\omega,$$

where

$$D_\omega = \mathrm{E}\left[I_T(x\zeta_0)(x_{(2)} - \mathrm{E}(x_{(2)}\,|\,x\zeta_0))'\omega^{-1}(x\zeta_0,\delta_0)\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\partial \delta'}\right]$$

and

$$\frac{1}{n}\frac{\partial \mathrm{E}_n(\hat{\beta},\hat{\zeta})}{\partial \delta}\hat{\Lambda}_n^{-1}\frac{\partial \mathrm{E}_n(\hat{\beta},\hat{\zeta})}{\partial \delta'} \xrightarrow{p} E_\omega,$$

where

$$\mathrm{E}_\omega = \mathrm{E}\left[I_T(x\zeta_0)\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\sqrt{\partial \delta}}\omega^{-1}(x\zeta_0,\delta_0)\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\partial \delta'}\right].$$

Similarly,

$$\frac{1}{n}\frac{\partial \mathrm{E}_n(\hat{\beta},\hat{\zeta})}{\partial \delta}\hat{\Lambda}_n^{-1}\hat{W}_n \xrightarrow{p} D'_\omega \Pi_w.$$

With (5.5), by combining the above relations,

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

and

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

where

$$\Gamma = C_\omega - D_\omega(V_\delta^{-1} + E_\omega)^{-1}D'_\omega. \tag{5.12}$$

Furthermore, from the derivations in Appendix Section A.5 [Eq. (A.5.2)],

$$\frac{1}{\sqrt{n}}\hat{X}'_2\hat{\Sigma}^{-1}\hat{U}_n \xrightarrow{D} \mathrm{N}(0,\Gamma). \tag{5.13}$$

Eqs. (5.9)–(5.13) imply that

$$\sqrt{n}(\tilde{\beta}_{\mathrm{SG}} - \beta_0) \xrightarrow{D} \mathrm{N}(0,(\Pi'_w\Gamma\Pi_w)^{-1}). \tag{5.14}$$

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 $\sqrt{n}(\hat{\beta}_{\mathrm{SG}} - \beta_0)$ has the same limiting distribution of $\sqrt{n}(\tilde{\beta}_{\mathrm{SG}} - \beta_0)$. The asymptotic covariance matrix of $\hat{\beta}_{\mathrm{SG}}$ (or $\tilde{\beta}_{\mathrm{SG}}$) 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 Schwartz inequality, $\hat{X}'_2\hat{\Sigma}^{-1}\hat{X}_2 \geqslant \hat{X}'_2\hat{Q}(\hat{Q}'\hat{\Sigma}\hat{Q})^{-1}\hat{Q}'\hat{X}_2$. Hence

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

by (4.7), (4.14), and the fact that $(1/n)\hat{X}'_2\hat{\Sigma}\hat{X}_2 \xrightarrow{p} \Delta$. Hence $\tilde{\beta}_{\mathrm{SG}}$ is asymptotically efficient relative to $\hat{\beta}_{\mathrm{S2LS}}$.

The SG2LS estimators $\hat{\beta}_{\mathrm{SG}}$ and $\tilde{\beta}_{\mathrm{SG}}$ are not only asymptotically efficient relative to the S2LS estimator $\hat{\beta}_{\mathrm{S2LS}}$. They have also an asymptotically optimal property. They are asymptotically efficient IV estimators for the estimation of (2.7) (conditional on the choice of first-stage estimator of δ_0 and the same trimmed version). 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

$$P = [t_n(x_1\hat{\zeta}, \hat{\xi}_n)p'_1, \ldots, t_n(x_n\hat{\zeta}, \hat{\xi}_n)p'_n]'. \tag{5.19}$$

Eq. (2.7) implies that

$$\hat{\beta}_p - \beta_0 = (P'\hat{W})^{-1}P'\hat{U}_n. \tag{5.20}$$

For this case, the component $E_n(u_1 \mid x_i\zeta_0) - E(u_1 \mid x\zeta_0 = x_i\zeta_0)$ in $\hat{u}_{ni}$ has a significant impact on the asymptotic distribution of $\hat{\beta}_p$. From (A.5.4) of the Appendix,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) p_i'(u_{1i} - E_n(u_1 \mid x_i\zeta_0))$$

$$\overset{D}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_0)(p_i - E(p_i \mid x_i\zeta_0))'(u_{1i} - E(u_{1i} \mid x_i\zeta_0)), \tag{5.21}$$

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

$$\frac{1}{n} P'\hat{W} \overset{P}{\to} E(I_T(x\zeta_0)p'(x_{(2)} - E(x_{(2)} \mid x\zeta_0))\Pi_w$$

$$= E(I_T(x\zeta_0)(p - E(p \mid x\zeta_0))'(x_{(2)} - E(x_{(2)} \mid x\zeta_0))\Pi_w. \tag{5.22}$$

The latter equality in (5.22) holds because $I_T(x\zeta_0)$ is a function of $x\zeta_0$ alone. From (A.5.3) and (A.5.4) in Appendix Section A.5,

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

where

$$\begin{aligned}\Delta_p = {} & E[I_T(x\zeta_0)(p - E(p \mid x\zeta_0))'\omega(x\zeta_0, \delta_0)(p - E(p \mid x\zeta_0))] \\ & + E[I_T(x\zeta_0)(p - E(p \mid x\zeta_0))'(x_{(2)} - E(x_{(2)} \mid x\zeta_0))\nabla E(u_1 \mid x\zeta_0)] \\ & \times V_\delta E[I_T(x\zeta_0)(p - E(p \mid x\zeta_0))'(x_{(2)} - E(x_{(2)} \mid x\zeta_0)\nabla E(u_1 \mid x\zeta_0)]'.\end{aligned} \tag{5.24}$$

It follows that

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

where

$$\begin{aligned}\Omega_p = {} & \{E[I_T(x\zeta_0)(p - E(p \mid x\zeta_0))'(x_{(2)} - E(x_{(2)} \mid x\zeta_0))\Pi_w\}^{-1} \\ & \times \Delta_p\{E(I_T(x\zeta_0)(p - E(p \mid x\zeta_0))'(x_{(2)} - E(x_{(2)} \mid x\zeta_0))\Pi_w\}^{-1\prime}.\end{aligned} \tag{5.26}$$

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

$$\hat{P} = \begin{pmatrix} t_n(x_1\zeta_0, \xi_0)(p_1 - \mathrm{E}_n(p\,|\,x_1\hat{\zeta})) \\ \vdots \\ t_n(x_n\zeta_0, \xi_0)(p_n - \mathrm{E}_n(p\,|\,x_n\hat{\zeta})) \end{pmatrix}. \tag{5.28}$$

This indicates that for the estimation of the semiparametric simultaneous equation model, the residuals $p_i - \mathrm{E}_n(p\,|\,x_i\zeta_0)$ and $p - \mathrm{E}(p\,|\,x\zeta_0)$ 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 $\tilde{\Omega}_{p,n} = [\hat{P}'\hat{W}]^{-1}\hat{P}'\hat{\Sigma}\hat{P}[\hat{W}'\hat{P}]^{-1}$. The asymptotic efficiency of $\tilde{\beta}_{SG}$ relative to $\tilde{\beta}_p$ follows from the following inequality:

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

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 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 (conditional on the same first-stage estimator of the parameters of the selection equation and the trimming of observations with low index densities) 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 indices 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), semiparametric efficiency bound has been derived in Chamberlain (1986). For efficient estimation, asymptotic efficient estimators might be derived from some semiparametric maximum likelihood methods. For the semiparametric sample selection model without simultaneity, Lee (1990) has considered such a procedure.[4] 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, 1983). One might conjecture that such a structural estimator could be an asymptotically efficient semiparametric estimator. At any rate, such estimation method is not a simple instrumental variable method.

Appendix

A.1. Model assumptions

Assumption 1

(1) The samples s_i, where $s_i = (y_i, z_i, x_i)$, $i = 1, \dots, n$, are i.i.d. x is the vector consisting of all exogenous variables in the equation system. The moments of order $3 \times r$, where $r \geqslant 2$, of s exist.

(2) The parameter space Θ of δ is a compact subset of a finite-dimensional Euclidean space, and δ_0 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, $\sqrt{n}(\hat{\delta} - \delta_0)$ is asymptotically normal, $\mathrm{N}(0, V_\delta)$, where V_δ is a positive definite matrix, and is asymptotically uncorrelated with $u_{1i} - \mathrm{E}(u_{1i} \mid x_i \zeta_0)$ for all i.

[4] In Lee (1990), both the density of index in the selection equation and the density of the disturbances in the outcome equations are assumed to be bounded away from zero. For such a situation, there is no need to trim the observations with lower index densities. Unfortunately, such a strong assumption rules out many cases. The author is currently investigating how to relax such an assumption. Trimming procedure introduced in this article may be valuable.

(5) The matrix Π_w in (4.4) has full column rank.

(6) The matrices C in (4.6) and Γ in (5.12) 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.,

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

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

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

Assumption 3

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

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

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

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

---

[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 of order 1 if there exists a constant c such that $\| h(x_1) - h(x_2) \| \leqslant c \| x_1 - x_2 \|$ for all x_1 and x_2.

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

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

Assumption 5

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

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

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

As pointed out in Section 4 of the text, Assumption 1(4) will, in general, be satisfied with parametric or semiparametric estimators of discrete choice models. 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: $\int_D |K(v)|\,\mathrm{d}v$ and $\int_D \|v\|^{s^*}|K(v)|\,\mathrm{d}v$ are finite; $K(v)$ and its first-order derivatives are bounded; and $K(v)$ and $\partial K(v)/\partial v$ 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 $f(x_{(1)} \mid x_{(2)})$ be the density function of $x_{(1)}$ conditional on $x_{(2)}$. Since $t = x_{(1)} + x_{(2)}\delta$,

$$p(t \mid \delta) = \int f(t - x_{(2)}\delta \mid x_{(2)})\,\mathrm{d}v(x_{(2)}), \tag{A.1.1}$$

where $v(x_{(2)})$ is the distribution measure of $x_{(2)}$. If $f(x_{(1)} \mid x_{(2)})$ is continuous and bounded, the bounded convergence theorem will imply that $p(t \mid \delta)$ 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 Section 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.

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 (LeCam, 1960). Let $\{\hat{\delta}_n\}$ be a sequence of $\sqrt{n}$-consistent estimate of δ_0. Let $\|\delta\| = \max_{l=1,\ldots,k}|\delta_l|$, where $\delta = (\delta_1, \ldots, \delta_k)'$ is the norm of δ in the k-dimensional Euclidean space R^k. Let $R_n^k = \{(1/\sqrt{n}) \times (i_1, \ldots, i_k) \mid i_1, \ldots, i_k$ are integers$\}$ and let $\bar{\delta}_n$ be a point in R_n^k closest to $\hat{\delta}_n$ under $\|\cdot\|$. The $\{\bar{\delta}_n\}$ is a discretized sequence of estimates of δ_0. 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 $x_i\bar{\zeta}$, $i = 1, \ldots, n$, 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 asymptotic analysis. First of all, $\hat{\xi}_{np}$ and $\hat{\xi}_{n(1-p)}$ can be shown to be $\sqrt{n}$-consistent. Since $\hat{\delta}_n$ is $\sqrt{n}$-consistent and $\|\hat{\delta}_n - \bar{\delta}_n\| \leqslant 1/\sqrt{n}$,

$$\sqrt{n}(\bar{\delta}_n - \delta_0) = \sqrt{n}(\bar{\delta}_n - \hat{\delta}_n) + \sqrt{n}(\hat{\delta}_n - \delta_0) = O_p(1),$$

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

$$P(|\xi_{np}(\delta) - \xi_p(\delta)| > \varepsilon) \leqslant 2\exp(-n\delta_\varepsilon^2), \qquad n \geqslant 1, \tag{A.2.1}$$

where $\delta_\varepsilon = \min[F_\delta(\xi_p(\delta) + \varepsilon) - p, p - F_\delta(\xi_p(\delta) - \varepsilon)]$ and F_δ denotes the distribution of $x\zeta$.[8] Under Assumption 3(1), the density function $p(t \mid \delta)$ of $x\zeta$ is positive everywhere, and hence is bounded away from zero on any compact neighborhood $\mathcal{N}(\xi_p(\delta_0)) \times \mathcal{N}(\delta_0)$ of $(\xi_p(\delta_0), \delta_0)$. With (A.2.1), we can show that for any sequence $\{\delta_n\}$ such that $\lim_{n\to\infty}\delta_n = \delta_0$,

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

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

Let $\varepsilon_n = c/\sqrt{n}$, where c is an arbitrary constant. By the mean value theorem, $F_{\delta_n}(\xi_p(\delta_n) + \varepsilon_n) - p = p(\xi_p(\delta_n) + \lambda_n \varepsilon_n | \delta_n)\varepsilon_n$ for some λ_n, $0 \leqslant \lambda_n \leqslant 1$. By the continuity of $\xi_p(\delta)$ at $\delta = \delta_0$, $(\xi_p(\delta_n) + \lambda_n \varepsilon_n, \delta_n) \in \mathcal{N}(\xi_p(\delta_0)) \times \mathcal{N}(\delta_0)$ for sufficiently large n. It follows that for large n, $F_{\delta_n}(\xi_p(\delta_n) + \varepsilon_n) - p \geqslant b\varepsilon_n$, where $b = \inf_{(t,\delta)\in\mathcal{N}(\xi_p(\delta_0))\times\mathcal{N}(\delta_0)} p(t|\delta) > 0$. Similarly, $p - F_{\delta_n}(\xi_p(\delta_n) - \varepsilon_n) \geqslant b\varepsilon_n$. Therefore, for large n, $\exp(-n\delta^2_{\varepsilon_n}) \leqslant \exp(-nb^2\varepsilon_n^2) = \exp(-b^2c^2)$, and

$$P(\sqrt{n}\,|\,\xi_{np}(\delta_n) - \xi_p(\delta_n)| > c) \leqslant 2\exp(-b^2c^2). \tag{A.2.3}$$

Since c is arbitrary, it follows that $\sqrt{n}|\xi_{np}(\delta_n) - \xi_p(\delta_n)| = O_p(1)$ for any sequence $\{\delta_n\}$ which converges to δ_0. For any finite constant $M > 0$, define $\Delta_{M,n} = \{\delta \,|\, \|\delta - \delta_0\| < M/\sqrt{n}$ and $\delta \in R_n^k\}$. $\Delta_{M,n}$ 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 $\{(i_1, \ldots, i_k) | i_1, \ldots, i_k$ are integers and $|i_l - \sqrt{n}\delta_{0,l}| < M$, $l = 1, \ldots, k\}$ 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_0. The finiteness of $\Delta_{M,n}$ and (A.2.3) imply that

$$P\left(\sup_{\delta_n \in \Delta_{M,n}} \sqrt{n}\,|\,\xi_{np}(\delta_n) - \xi_p(\delta_n)| > c\right) \leqslant 2\bar{M}\exp(-b^2c^2), \tag{A.2.4}$$

for large n, and therefore

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

Since $\bar{\delta}_n$ is in R_n^k and is a $\sqrt{n}$-consistent estimate of δ_0, $\bar{\delta}_n$ will lie in $\Delta_{M,n}$ with probability close to one for large n. Hence

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

By the mean value theorem, $\xi_p(\bar{\delta}_n) = \xi_p(\delta_0) + (\partial\xi_p(\delta_n^*)/\partial\delta')(\bar{\delta}_n - \delta_0)$. Since $p(t|\delta)$ is continuous, $\partial\xi_p(\delta)/\partial\delta$ is continuous at δ_0, and $\sqrt{n}(\bar{\delta}_n - \delta_0) = O_p(1)$, it follows that

$$\begin{aligned}\sqrt{n}\,|\,\xi_{np}(\bar{\delta}_n) - \xi_p(\delta_0)| &\leqslant \sqrt{n}\,|\,\xi_{np}(\bar{\delta}_n) - \xi_p(\bar{\delta}_n)| + \left\|\frac{\partial\xi_p(\delta_n^*)}{\partial\delta'}\right\|\sqrt{n}|\bar{\delta}_n - \delta_0| \\ &= O_p(1),\end{aligned} \tag{A.2.6}$$

i.e., $\hat{\xi}_{np}$ is a $\sqrt{n}$-consistent estimate of $\xi_p(\delta_0)$. Similarly, $\hat{\xi}_{n(1-p)}$ is a $\sqrt{n}$-consistent estimate of $\xi_{(1-p)}(\delta_0)$.

The LeCam discretization device is also useful in the following way. Denote $\hat{\xi}_n = (\hat{\xi}_{np}, \hat{\xi}_{n(1-p)})$ and $\xi_0 = (\xi_p(\delta_0), \xi_{(1-p)}(\delta_0))$. For any statistic $S_n(\hat{\xi}_n)$ constructed from the sample and $\hat{\xi}_n$, if we want to show that $S_n(\hat{\xi}_n)$ converges to 0 in probability, it is sufficient to show that $S_n(\xi_0 + (1/\sqrt{n})h_n)$, where $\{h_n\}$ is any nonstochastic bounded sequence, converges to 0 in probability. Effectively, one can replace the stochastic sequence $\{\hat{\xi}_n\}$ by nonstochastic sequence $\{\xi_0 + (1/\sqrt{n})h_n\}$ in the proof of convergence in probability (see LeCam, 1960; Manski, 1984).

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 $\lim_{n\to\infty}(n/\ln n)\,h_n^{(1+4/r)} = \infty$, and let $q(\cdot)$ be a continuous density function on [0 1] such that $q(0) = q(1) = 0$.[9] Denote $\xi_n = (\xi_{np}, \xi_{n(1-p)})$. Define the following smooth trimming function t_n of $x\zeta$:

$$t_n(x\zeta, \xi_n) = \begin{cases} 0 & \text{if } x\zeta < \xi_{np}, \\ \int_0^{\frac{x\zeta - \xi_{np}}{h_n}} q(w)\,\mathrm{d}w & \text{if } \xi_{np} \leqslant x\zeta \leqslant \xi_{np} + h_n, \\ 1 & \text{if } \xi_{np} + h_n < x\zeta < \xi_{n(1-p)} - h_n, \\ 1 - \int_0^{\frac{x\zeta - \xi_{n(1-p)}}{h_n}+1} q(w)\,\mathrm{d}w & \text{if } \xi_{n(1-p)} - h_n \leqslant x\zeta \leqslant \xi_{n(1-p)}, \\ 0 & \text{if } \xi_{n(1-p)} \leqslant x\zeta. \end{cases} \tag{A.3.1}$$

This function is continuously differentiable in δ, ξ_{np}, and $\xi_{n(1-p)}$, with

$$\frac{\partial t_n(x\zeta, \xi_n)}{\partial \delta'} = \begin{cases} 0 & \text{if } x\zeta < \xi_{np}, \\ \frac{1}{h_n} q\left(\frac{x\zeta - \xi_{np}}{h_n}\right) x_{(2)} & \text{if } \xi_{np} \leqslant x\zeta \leqslant \xi_{np} + h_n, \\ 0 & \text{if } \xi_{np} + h_n < x\zeta < \xi_{n(1-p)} - h_n, \\ -\frac{1}{h_n} q\left(\frac{x\zeta - \xi_{n(1-p)}}{h_n} + 1\right) x_{(2)} & \text{if } \xi_{n(1-p)} - h_n \leqslant x\zeta \leqslant \xi_{n(1-p)}, \\ 0 & \text{if } \xi_{n(1-p)} \leqslant x\zeta, \end{cases} \tag{A.3.2}$$

$$\frac{\partial t_n(x\zeta, \xi_n)}{\partial \xi_{np}} = \begin{cases} 0 & \text{if } x\zeta < \xi_{np}, \\ -\frac{1}{h_n} q\left(\frac{x\zeta - \xi_{np}}{h_n}\right) & \text{if } \xi_{np} \leqslant x\zeta \leqslant \xi_{np} + h_n, \\ 0 & \text{if } \xi_{np} + h_n < x\zeta, \end{cases} \tag{A.3.3}$$

[9] The rate of convergence of h_n is designed to justify our asymptotic analysis (see Proposition 2 of Section A.4).

and

$$\frac{\partial t_n(x\zeta,\xi_n)}{\partial \xi_{n(1-p)}} = \begin{cases} 0 & \text{if} \quad x\zeta < \xi_{n(1-p)} - h_n, \\ \dfrac{1}{h_n} q\left(\dfrac{x\zeta - \xi_{n(1-p)}}{h_n} + 1\right) & \text{if} \quad \xi_{n(1-p)} - h_n \leqslant x\zeta \leqslant \xi_{n(1-p)}, \\ 0 & \text{if} \quad \xi_{n(1-p)} < x\zeta. \end{cases} \tag{A.3.4}$$

As n tends to infinity, $t_n(x\hat{\zeta},\hat{\xi}_n)$ will converge in probability to the indicator function $I_T(x\zeta_0)$, where $T = [\xi_p(\delta_0), \xi_{(1-p)}(\delta_0)]$.

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.

A.4. Some useful asymptotic properties of nonparametric estimators of unknown 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 (1991). Lemmas 2, 3, and 4 provide results on the asymptotic biases of functions involving a kernel function and its first- and second-order derivatives. The proofs of Lemmas 2, 3, and 4 are also abstracted from Ichimura and Lee (1991).

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

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

where $a_n = O(1/n^p)$, $p>0$, $d>0$, $\beta \in B$, and $s(z,\beta)$ is a finite-dimensional vector value function. Let $\{z_i\}$ 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) $c(z,\beta)$ is bounded by an l-order polynomial of z uniformly in β, where $l \geqslant 0$.*
- *(iii) The first $l \times r$ moments of z exist, where $r \geqslant 2$.*
- *(iv) $h(\cdot)$ is a bounded function.*
- *(v) $\mathrm{E}[c^2(z,\beta)h^2(z,\beta,s(z,\beta)/a_n)] = O(a_n^{\bar{d}})$ uniformly in $\beta \in B$, where $\bar{d} \leqslant d$.*
- *(vi) $h(z,\beta,s)$ satisfies the Lipschizian condition of order 1 with respect to β and s uniformly in z; $s(z,\beta)$ satisfies the Lipschizian condition of order 1 with respect to β uniformly in z.*

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

Furthermore, in addition to the above conditions, if:

(*vii*) $\mathrm{E}(g(z, a_n, \beta))$ *converges to a limit function* $g^*(\beta)$ *uniformly in* $\beta \in B$, *then* $(1/n)\sum_{i=1}^n g(z_i, a_n, \beta)$ *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* $\int_D |K(v)|\,\mathrm{d}v < \infty$. $t(z,\theta)$ *is a continuous m-dimensional random vector.* z_i *is a point in the support of z. Suppose that* $\mathrm{E}(c(z, z_i, \theta) \mid t(z,\theta) = t, z_i) g(t \mid \theta)$, *where* $g(t \mid \theta)$ *is the density function of* $t(z, \theta)$, *is uniformly continuous at t on* R^m *uniformly in* (θ, z_i). *Then*

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

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

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

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* $\partial K(v)/\partial v$ *is bounded. Suppose that* $(\partial/\partial t)[\mathrm{E}(c(z, z_i, \theta) | t(z,\theta) = t, z_i) g(t \mid \theta)]$, *where* $g(t \mid \theta)$ *is the density function* $t(z, \theta)$, *are uniformly continuous at t uniformly in* (z_i, θ). *Then*

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

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

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

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 $\partial K(v)/\partial v$ go to zero at the boundary of D, and the gradient $\partial K(v)/\partial v$ and its Hessian matrix $\partial^2 K(v)/\partial v \partial v'$ are bounded. Suppose that $(\partial^2/\partial t \partial t')[\mathrm{E}(c(z, z_i, \theta) | t(z, \theta) = t, z_i) g(t | \theta)]$ are uniformly continuous at t everywhere uniformly in (z_i, θ). Then

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

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 familiar order $\mathrm{O}(1/na_n^m)$ uniformly on $C \times \Theta$, where C is any compact subset of $x\zeta$ in R^m; i.e.,

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

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

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

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

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

where

$$A(s\,|\,x_i\zeta,\delta) = [\mathrm{E}(s\,|\,x\zeta)p(x\zeta\,|\,\delta)]|_{x\zeta=x_i\zeta}, \tag{A.4.5}$$

i.e., $A(s\,|\,x_i\zeta,\delta)$ is the product of $p(x_i\zeta\,|\,\delta)$ and the conditional expectation of s conditional on $x\zeta$ evaluated at the point $x_i\zeta$. To simplify notation in subsequent presentation, we adopt the convention that $\mathrm{E}(s\,|\,x\zeta = x_i\zeta)$ denotes the conditional expectation $\mathrm{E}(s\,|\,x\zeta)$ evaluated at the point $x_i\zeta$, i.e., $\mathrm{E}(s\,|\,x\zeta = x_i\zeta) = \mathrm{E}(s\,|\,x\zeta)|_{x\zeta=x_i\zeta}$, for any random variable s. Since with probability close to one, T_n will be contained in a compact subset of R^m for large n, $A_n(s\,|\,x_i\zeta)$ converges in probability 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 $\inf_{T_n\times\Theta}A_n(1\,|\,x_i\zeta)$ is bounded away from zero in probability.[10] $\mathrm{E}_n(s\,|\,x_i\zeta)$ as a ratio of $A_n(s\,|\,x_i\zeta)$ over $A_n(1\,|\,x_i\zeta)$ will converge in probability to $\mathrm{E}(s\,|\,x\zeta = x_i\zeta)$ uniformly on $T_n\times\Theta$.

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

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

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

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

where

$$\begin{aligned}\frac{\partial A(s\,|\,x_i\zeta,\delta)}{\partial\delta_k} &= \mathrm{tr}\left\{\frac{\partial\mathrm{E}(s(x_i - x)|x\zeta = x_i\zeta, x_i)}{\partial t}\,\frac{\partial\zeta}{\partial\delta_k}\right\}p(x_i\zeta\,|\,\delta)\\ &\quad + \mathrm{E}(s(x_i - x)|x\zeta = x_i\zeta, x_i)\frac{\partial\zeta}{\partial\delta_k}\,\nabla p(x_i\zeta\,|\,\delta),\end{aligned} \tag{A.4.8}$$

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

where $\nabla p(x\zeta \mid \delta)$ denotes the first-order derivative of $p(x\zeta \mid \delta)$ with respect to $x\zeta$ (see Ichimura and Lee, 1991). Lemmas 1 and 3 imply that if $\lim_{n\to\infty} (n/\ln n)\, a_n^{(1+4/r)(m+1)+1} = \infty$,

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

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

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

Since $\mathrm{E}_n(s \mid x_i\zeta) = A_n(s \mid x_i\zeta)/A_n(1 \mid x_i\zeta)$,

$$\frac{\partial \mathrm{E}_n(s \mid x_i\zeta)}{\partial \delta} = \left(\frac{\partial A_n(s \mid x_i\zeta)}{\partial \delta} - \mathrm{E}_n(s \mid x_i\zeta) \frac{\partial A_n(1 \mid x_i\zeta)}{\partial \delta} \right) \Big/ A_n(1 \mid x_i\zeta)$$

converges in probability to $\partial \mathrm{E}(s \mid x\zeta = x_i\zeta)/\partial\delta$, where

$$\frac{\partial \mathrm{E}(s \mid x\zeta)}{\partial \delta} = \left(\frac{\partial A(s \mid x\zeta, \delta)}{\partial \delta} - \mathrm{E}(s \mid x\zeta) \frac{\partial A(1 \mid x\zeta, \delta)}{\partial \delta} \right) \Big/ A(1 \mid x\zeta, \delta)$$

from (A.4.3), (A.4.4), and (A.4.9). Since $\mathrm{E}(u_1 \mid x) = \mathrm{E}(u_1 \mid x\zeta_0)$, as shown in Ichimura and Lee (1991, Lemma 4), we have explicitly the following expression:

$$\left. \frac{\partial \mathrm{E}(u_1 \mid x\zeta)}{\partial \delta_k} \right|_{\delta=\delta_0} = (x - \mathrm{E}(x \mid x\zeta_0)) \frac{\partial \zeta(\delta_0)}{\partial \delta_k} \nabla \mathrm{E}(u_1 \mid x\zeta_0), \tag{A.4.11}$$

where $\nabla \mathrm{E}(u_1 \mid x\zeta_0)$ is the first-order derivative of $\mathrm{E}(u_1 \mid x\zeta_0)$ with respect to $x\zeta_0$.

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

Proposition 1. Let $\mathcal{N}(\xi_0)$ be a compact neighborhood of $\xi_0 = (\xi_p(\delta_0), \xi_{(1-p)}(\delta_0))$ such that for any $\xi = (\xi_p, \xi_{(1-p)}) \in \mathcal{N}(\xi_0)$, $[\xi_p, \xi_{(1-p)}] \subseteq \bar{T}$ where $\bar{T} = [\xi_p(\delta_0) - \Delta, \xi_{(1-p)}(\delta_0) + \Delta]$ for some $\Delta > 0$ (a constant not depended on ξ). Suppose that:

(*1*) *$t_n(x_i\zeta, \xi)$ is a bounded smooth trimming function defined in Section A.3, which vanishes at any $x_i\zeta$ outside the interval $[\xi_p, \xi_{(1-p)}]$.*
(*2*) *$A_n(s \mid x_i\zeta) \xrightarrow{p} A(s \mid x_i\zeta, \delta)$ uniformly in $(x_i\zeta, \delta) \in \bar{T} \times \Theta$, where Θ is a compact neighborhood of δ_0, and $A(s \mid x\zeta, \delta)$ is continuous in δ and $x\zeta$.*
(*3*) *$f(d)$ is uniformly continuous and bounded on A_Δ, where $A_\Delta = \{d \mid d = A(s \mid x\zeta, \delta), (x\zeta, \delta) \in \bar{T} \times \Theta\}$.*

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

Then

$$\operatorname*{plim}_{n\to\infty} \sup_{\Theta\times\mathcal{N}(\xi_0)} \left| \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\zeta,\xi)a(s_i)f(A_n(s\mid x_i\zeta)) - \mathrm{E}\{t_n(x\zeta,\xi)a(s)f(A(s\mid x\zeta,\delta))\} \right| = 0.$$

In addition to the above assumptions, if $A(s\mid x\zeta,\delta)$ is continuous in δ and $x\zeta$ a.e., $\hat{\delta}$ is a consistent estimate of δ_0, and $\hat{\xi}_n$ is a consistent estimate of ξ_0, then

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi}_n)a(s_i)f(A_n(s\mid x_i\hat{\zeta})) = \mathrm{E}\{I_T(x\zeta_0)a(s)f(A(s\mid x\zeta_0,\delta_0))\},$$

where $T = [\xi_p(\delta_0), \xi_{(1-p)}(\delta_0)]$.

Proof.

$$\sup_{\Theta\times\mathcal{N}(\xi_0)} \left| \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\zeta,\xi)a(s_i)f(A_n(s\mid x_i\zeta)) - \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\zeta,\xi)a(s_i)f(A(s\mid x_i\zeta,\delta)) \right|$$

$$\leqslant \frac{1}{n}\sum_{i=1}^{n} |a(s_i)| \sup_{(x_i\zeta,\delta)\in\bar{T}\times\Theta} |f(A_n(s\mid x_i\zeta)) - f(A(s\mid x_i\zeta,\delta))|$$

$$\xrightarrow{p} 0,$$

because t_n is bounded by one, $f(A_n(s\mid x_i\zeta))$ converges in probability to $f(A(s\mid x_i\zeta,\delta))$ uniformly on $\bar{T}\times\Theta$, and $(1/n)\sum_{i=1}^{n}|a(s_i)| = \mathrm{O}_p(1)$. Because $\sup_{\Theta\times\mathcal{N}(\xi_0)}|t_n(x\zeta,\xi)f(A(s\mid x\zeta,\delta))|$ 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,

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

uniformly on $\Theta\times\mathcal{N}(\xi_0)$. The second part of the conclusion follows by the Lebesgue dominated convergence theorem. Q.E.D.

Proposition 2. Let $\mathcal{N}(\xi_0)$ be a compact neighborhood of $\xi_0 = (\xi_p(\delta_0), \xi_{(1-p)}(\delta_0))$ such that for any $\xi = (\xi_p, \xi_{(1-p)}) \in \mathcal{N}(\xi_0)$, $[\xi_p, \xi_{(1-p)}] \subseteq \bar{T}$ where $\bar{T} = [\xi_p(\delta_0) - \Delta, \xi_{(1-p)}(\delta_0) + \Delta]$ for some $\Delta > 0$ (a constant not dependent on ξ). Let $x\zeta = (x\zeta_1, x\zeta_{(2)})$ and, conformably, $\xi = (\xi_1, \xi_{(2)})$, where $x\zeta_1$ is a single index. Let $t_n(x\zeta_{(2)}, \xi_{(2)})$ be the trimming function for the indices $x\zeta_{(2)}$ defined in

Section A.3. Suppose that:

(*1*) *q is a continuously differentiable density function having a support* $[0, 1]$.
(*2*) $A_n(s \mid x_i\zeta) \xrightarrow{p} A(s \mid x_i\zeta, \delta)$ *uniformly in* $(x_i\zeta, \delta) \in \bar{T} \times \Theta$, *where* Θ *is a compact neighborhood of* δ_0, *and* $A(s \mid x\zeta, \delta)$ *is continuous in* δ *and* $x\zeta$.
(*3*) $f(d)$ *is uniformly continuous on* A_Δ, *where* $A_\Delta = \{d \mid d = A(s \mid x\zeta, \delta), (x\zeta, \delta) \in \bar{T} \times \Theta\}$.
(*4*) $a(s)$ *is bounded by a polynomial of order* l, $l \geqslant 0$. *The* $r \times l$*-order moment of* s *exists, where* $r \geqslant 2$.
(*5*) *The functions* $q((x\zeta_{1,p} - \xi_{1,p})/h_n)\, t_n(x\zeta_{(2)}, \xi_{(2)})$ *of* $x\zeta$ *have finite supports contained in* $\bar{T}$ *for all* $\xi \in \mathcal{N}(\xi_0)$ *for large n.*
(*6*) $\mathrm{E}[t_n(x\zeta_{(2)}, \xi_{(2)})a(s)f(A(s \mid x\zeta, \delta)) \mid x\zeta_1]\, g_1(x\zeta_1 \mid \delta)$ *is uniformly continuous at* $x\zeta_1$, *uniformly on* $\Theta \times \mathcal{N}(\xi_0)$ *and n, where* $g_1(t \mid \delta)$ *is the marginal density of* $x\zeta_1$.

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

$$\operatorname*{plim}_{n\to\infty} \sup_{\Theta\times\mathcal{N}(\xi_0)} \Bigg| \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_n} q\left(\frac{x_i\zeta_1 - \xi_{1,p}}{h_n}\right) t_n(x_i\zeta_{(2)}, \xi_{(2)}) a(s_i) f(A_n(s \mid x_i\zeta))$$
$$- \mathrm{E}[t_n(x\zeta_{(2)}, \xi_{(2)}) a(s) f(A(s \mid x\zeta)) \mid x\zeta_1 = \xi_{1,p}]\, g_1(\xi_{1,p} \mid \delta) \Bigg| = 0.$$

In addition to the above assumptions, if $\hat{\delta}$ *is a consistent estimate of* δ_0 *and* $\hat{\xi}_n$ *is a consistent estimate of* ξ_0, *then*

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_n} q\left(\frac{x_i\hat{\zeta}_1 - \hat{\xi}_{1,pn}}{h_n}\right) t_n(x_i\hat{\zeta}_{(2)}, \xi_{(2),n}) a(s_i) f(A_n(s \mid x_i\hat{\zeta}))$$
$$= \mathrm{E}[I_{T_{(2)}}(x\zeta_{(2)}(\delta_0)) a(s) f(A(s \mid x\zeta_0)) \mid x\zeta_1(\delta_0) = \xi_{1,p}(\delta_0)]\, g_1(\xi_{1,p}(\delta_0) \mid \delta_0),$$

where $T_{(2)} = [\xi_{(2),p}(\delta_0), \xi_{(2),(1-p)}(\delta_0)]$.

Proof

$$\sup_{\Theta\times\mathcal{N}(\xi_0)} \Bigg| \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_n} q\left(\frac{x_i\zeta_1 - \xi_{1,p}}{h_n}\right) t_n(x_i\zeta_{(2)}, \xi_{(2)})$$
$$\times a(s_i)[f(A_n(s \mid x_i\zeta)) - f(A(s \mid x_i\zeta, \delta))] \Bigg|$$
$$\leqslant \sup_{\Theta\times\mathcal{N}(\xi_0)} \frac{1}{n}\sum_{i=1}^{n} |a(s_i)| \frac{1}{h_n} q\left(\frac{x_i\zeta_1 - \xi_{1,p}}{h_n}\right)$$
$$\times \sup_{(x_i\zeta, \delta)\in\bar{T}\times\Theta} |f(A_n(s \mid x_i\zeta)) - f(A(s \mid x_i\zeta, \delta))|. \qquad (*)$$

By Lemma 1, as $nh_n^{(1+2l/r)}/\ln n = \infty$,

$$\sup_{\Theta\times\mathcal{N}(\xi_0)}\left|\frac{1}{n}\sum_{i=1}^{n}|a(s_i)|\frac{1}{h_n}q\left(\frac{x_i\zeta_1-\xi_{1,p}}{h_n}\right)-\mathrm{E}\left\{|a(s)|\frac{1}{h_n}q\left(\frac{x_i\zeta_1-\xi_{1,p}}{h_n}\right)\right\}\right|\xrightarrow{p}0.$$

By Lemma 2,

$$\sup_{\Theta\times\mathcal{N}(\xi_0)}\left|\mathrm{E}\left\{|a(s)|\frac{1}{h_n}q\left(\frac{x\zeta_1-\xi_{1,p}}{h_n}\right)\right\}-\mathrm{E}(|a(s)|\,|x\zeta_1=\xi_{1,p})g_1(\xi_{1,p}|\delta)\right|$$

$$\to 0.$$

Since $\mathrm{E}(|a(s)|\,|x\zeta_1=\xi_{1,p})g_1(\xi_{1,p}|\delta)$ is bounded on $\Theta\times\mathcal{N}(\xi_0)$, it follows that

$$\sup_{\Theta\times\mathcal{N}(\xi_0)}\frac{1}{n}\sum_{i=1}^{n}|a(s_i)|\frac{1}{h_n}q\left(\frac{x_i\zeta_1-\xi_{1,p}}{h_n}\right)=\mathrm{O}_p(1),$$

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

$$\sup_{\Theta\times\mathcal{N}(\xi_0)}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}q\left(\frac{x_i\zeta_1-\xi_{1,p}}{h_n}\right)t_n(x_i\zeta_{(2)},\xi_{(2)})a(s_i)f(A(s\,|\,x_i\zeta,\delta))\right.$$
$$\left.-\mathrm{E}\left[\frac{1}{h_n}q\left(\frac{x\zeta_1-\xi_{1,p}}{h_n}\right)t_n(x\zeta_{(2)},\xi_{(2)})a(s)f(A(s\,|\,x\zeta,\delta))\right]\right|\xrightarrow{p}0.$$

Since

$$\mathrm{E}\left[\frac{1}{h_n}q\left(\frac{x\zeta_1-\xi_{1,p}}{h_n}\right)t_n(x\zeta_{(2)},\xi_{(2)})a(s)f(A(s\,|\,x\zeta,\delta))\right]$$
$$=\int\mathrm{E}\,[t_n(x\zeta_{(2)},\xi_{(2)})a(s)f(A(s\,|\,x\zeta,\delta))|x\zeta_1=\xi_{1,p}+h_nv]g_1(\xi_{1,p}+h_nv|\delta)q(v)\mathrm{d}v,$$

it follows that

$$\sup_{\Theta\times\mathcal{N}(\delta_0)}\left|\mathrm{E}\left[\frac{1}{h_n}q\left(\frac{x\zeta_1-\xi_{1,p}}{h_n}\right)t_n(x\zeta_{(2)},\xi_{(2)})a(s)f(A(s\,|\,x\zeta,\delta))\right]\right.$$
$$\left.-\mathrm{E}\,[t_n(x\zeta_{(2)},\xi_{(2)})a(s)f(A(s\,|\,x\zeta,\delta))|x\zeta_1=\xi_{1,p}]g_1(\xi_{1,p}|\delta)\right|$$
$$\leqslant\int\sup_{\Theta\times\mathcal{N}(\delta_0),n}\left|\mathrm{E}\,[t_n(x\zeta_{(2)},\xi_{(2)})a(s)f(A(s\,|\,x\zeta,\delta))|x\zeta_1=\xi_{1,p}+h_nv]\right.$$
$$\times g_1(\xi_{1,p}+h_nv|\delta)-\mathrm{E}\,[t_n(x\zeta_{(2)},\xi_{(2)})a(s)f(A(s\,|\,x\zeta,\delta))|x\zeta_1=\xi_{1,p}]$$
$$\left.\times g_1(\xi_{1,p}|\delta)\right|q(v)\,\mathrm{d}v$$
$$\to 0$$

by the uniform continuity. The first part of the result follows from the above convergences. The second part of the result follows by the Lebesque dominated convergence theorem. Q.E.D.

Lemma 5. Let $C_n(\bar{s}_1, \ldots, \bar{s}_{i-1}, \bar{s}_{i+1}, \ldots, \bar{s}_n; \bar{s}_i)$ be a sequence of measurable functions of an i.i.d. sample $\{\bar{s}_i\}$ and $d_n(\bar{s})$ be a measurable function such that $\mathrm{E}(|d_n(\bar{s})|) < \infty$ uniformly in n. Suppose that:

(1) $\sup |\mathrm{E}(C_n(\bar{s}_1, \cdots, \bar{s}_n; \bar{s}_i) | \bar{s}_i) - C(\bar{s}_i)| = \mathrm{O}(a_n^{s^*})$, for some measurable function $C(\bar{s}_i)$, and

(2) $\sup \operatorname{var}(C_n(\bar{s}_1, \cdots, \bar{s}_n; \bar{s}_i) | \bar{s}_i) = \mathrm{O}(1/na_n^r)$.

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_n(\bar{s}_i) | C_n(\bar{s}_1, \ldots, \bar{s}_n; \bar{s}_i) - C(\bar{s}_i)|^2 \xrightarrow{p} 0.$$

Proof. This result can be easily proved by the Markov inequality (Lee, 1992, p. 79, Lemma 6).

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

(1) The nonparametric function $A_n(s | x_i\zeta_0)$ satisfies the conditions in Lemma 5, and

(2) $\sup_{x_i\zeta_0 \in T} | \nabla_a^2 f_n(s_i, \bar{A}_n(s | x_i\zeta_0))| \leqslant \mathrm{O}_p(1) d_n(s_i)$, where $\bar{A}_n(s | x_i\zeta_0)$ is a consistent estimate of $A(s | x_i\zeta_0, \delta_0)$ and $d_n(s_i)$ is a measurable function such that $\mathrm{E}(I_T(x_i\zeta_0) d_n(s_i)) < \infty$ uniformly in n.

Then

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_0) f_n(s_i, A_n(s | x_i\zeta_0)) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_0) f_n(s_i, A(s | x_i\zeta_0, \delta_0)) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_0) \nabla_a f_n(s_i, A(s | x_i\zeta_0, \delta_0)) \\
&\quad \times (A_n(s | x_i\zeta_0) - A(s | x_i\zeta_0, \delta_0)) + \mathrm{o}_p(1).
\end{aligned}$$

Proof. By a Taylor expansion,

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_0) f_n(s_i, A_n(s | x_i\zeta_0)) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_0) f_n(s_i, A(s | x_i\zeta_0, \delta_0)) + L_n + R_n,
\end{aligned}$$

where

$$L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_0)\, \nabla_a f_n(s_i, A(s\,|\,x_i\zeta_0,\delta_0))(A_n(s\,|\,x_i\zeta_0) - A(s\,|\,x_i\zeta_0,\delta_0))$$

and

$$R_n = \frac{1}{2}\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_0)\, \nabla_a^2 f_n(s_i, \bar{A}_n(s\,|\,x_i\zeta_0))(A_n(s\,|\,x_i\zeta_0) - A(s\,|\,x_i\zeta_0,\delta_0))^2,$$

where $\bar{A}_n(s\,|\,x_i\zeta_0)$ lies between $A_n(s\,|\,x_i\zeta_0)$ *and* $A(s\,|\,x_i\zeta_0,\delta_0)$. Since

$$|R_n| \leqslant O_p(1)\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_T(x_i\zeta_0) d_n(s_i)[A_n(s\,|\,x_i\zeta_0) - A(s\,|\,x_i\zeta_0,\delta_0)]^2$$

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

Lemma 6. Let $\{\bar{s}_i\}$ be an i.i.d. sample and $\Phi_n(\bar{s}_1,\bar{s}_2,a_n)$ be a sequence of vector-valued random functions with bandwidth a_n. Suppose that:

(*1*) *There exist square integrable functions $h_j(\bar{s})$ such that $|E(\Phi_n(\bar{s}_1,\bar{s}_2,a_n)|\bar{s}_j)| \leqslant h_j(\bar{s}_j)$ for $j = 1,2$.*
(*2*) $E(\Phi_n(\bar{s}_1,\bar{s}_2,a_n)) = O(a_n^{s^*})$ *and* $\mathrm{var}(\Phi_n(\bar{s}_1,\bar{s}_2,a_n)) = O(1/a_n^r)$.
(*3*) $\lim_{n\to\infty} E(\Phi_n(\bar{s}_1,\bar{s}_2,a_n)|\bar{s}_j) = \psi_j(\bar{s}_j)$, *a.e., for some measurable functions ψ_j, $j = 1, 2$.*
(*4*) $\lim_{n\to\infty}\sqrt{n}a_n^{s^*} = 0$ *and* $\lim_{n\to\infty} na_n^r = \infty$.

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

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

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

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

Proof. These results follow from Powell, Stock and Stoker (1989) and the Lindeberg central limit theorem. Q.E.D.

Proposition 4. Suppose that K is an r-dimensional kernel function with a bounded support D such that $\int_D |K(v)|\,dv < \infty$ and with a bandwidth a_n. Let

$$A_n(s\,|\,x_i\zeta_0) = \frac{1}{(n-1)a_n^r}\sum_{j\neq i}^{n} s_j K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n}\right),$$

and let $g(x\zeta_0 \mid \delta_0)$ be the density of $x\zeta_0$. Denote $A(s \mid x\zeta_0, \delta_0) = \mathrm{E}(s \mid x\zeta_0) g(x\zeta_0 \mid \delta_0)$. Let $f_n(s, x\zeta_0)$ be measurable functions such that $\sup_n |f_n(s, x\zeta_0)|$ is square integrable. Suppose that:

(1) $\mathrm{E}(s \mid x\zeta_0) g(x\zeta_0 \mid \delta_0)$ *is uniformly continuous at* $x\zeta_0$, *and*
(2) $\mathrm{E}(f_n(s, x\zeta_0) \mid x\zeta_0) = 0$ *a.e., for all n.*

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_n(s_i, x_i\zeta_0)[A_n(s \mid x_i\zeta_0) - A(s \mid x_i\zeta_0, \delta_0)] \xrightarrow{p} 0.$$

Proof. Define

$$\Phi_n(\bar{s}_i, \bar{s}_j, a_n) = f_n(s_i, x_i\zeta_0)\left[\frac{1}{a_n^r} s_j K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n}\right) - A(s \mid x_i\zeta, \delta_0)\right],$$

where $\bar{s} = (s, x)$. It follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_n(s_i, x_i\zeta_0)[A_n(s \mid x_i\zeta_0) - A(s \mid x_i\zeta_0, \delta_0)]$$

$$= \frac{1}{(n-1)\sqrt{n}} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \Phi_n(\bar{s}_i, \bar{s}_j, a_n).$$

For any $\varepsilon > 0$, for large n Lemma 2 implies that

$$|\mathrm{E}(\Phi_n(\bar{s}_1, \bar{s}_2, a_n) | \bar{s}_1)|$$

$$\leqslant |f_n(s_1, x_1\zeta_0)| \left| \mathrm{E}\left[\frac{1}{a_n^r} s_2 K\left(\frac{x_1\zeta_0 - x_2\zeta_0}{a_n}\right) - A(s \mid x_1\zeta, \delta_0) | x_1\right]\right|$$

$$\leqslant \varepsilon \sup_n |f_n(s_1, x_1\zeta_0)|,$$

which is square integrable. Since ε is arbitary, the above relations imply that $\mathrm{E}(\Phi_n(s_i, s_j, a_n) | s_i)$ converges to zero. On the other hand,

$$\mathrm{E}(\Phi_n(\bar{s}_j, \bar{s}_i, a_n) | \bar{s}_i)$$

$$= s_i \mathrm{E}\left[f_n(s_j, x_j\zeta_0) \frac{1}{a_n^r} K\left(\frac{x_j\zeta_0 - x_i\zeta_0}{a_n}\right) \middle| x_i\right] - \mathrm{E}[f_n(s_j, x_j\zeta_0) A(s \mid x_j\zeta_0, \delta_0)]$$

$$= s_i \int \mathrm{E}[f_n(s, x_i\zeta_0 + a_n v) | x_j\zeta_0 = x_i\zeta_0 + a_n v, x_i] g(x_i\zeta_0 + a_n v \mid \delta_0) K(v)\,\mathrm{d}v$$

$$- \mathrm{E}\{\mathrm{E}[f_n(s, x\zeta_0) | x\zeta_0)] A(s \mid x\zeta_0, \delta_0)\}$$

$$= 0,$$

by the condition (2). 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 $\int_D |K(v)|\,dv < \infty$ *and with a bandwidth* a_n. *Let*

$$A_n(s \mid x_i\zeta_0) = \frac{1}{(n-1)a_n^r} \sum_{j \neq i}^{n} s_j K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n}\right),$$

and let $g(x\zeta_0 \mid \delta_0)$ *be the density of* $x\zeta_0$. *Denote* $A(s \mid x\zeta_0, \delta_0) = \mathrm{E}(s \mid x\zeta_0) \times g(x\zeta_0 \mid \delta_0)$. *Let* $f_n(s, x\zeta_0)$ *be measurable functions such that* $\sup_n |f_n(s, x\zeta_0)|$ *is square integrable. Suppose that*:

(*1*) $\mathrm{E}(s \mid x\zeta_0) g(x\zeta_0 \mid \delta_0)$ *is uniformly continuous at* $x\zeta_0$.
(*2*) $\mathrm{E}(f_n(s, x\zeta_0) \mid x\zeta_0) g(x\zeta_0 \mid \delta_0)$ *is continuous in* $x\zeta_0$ *a.e. uniformly in n.*
(*3*) *There exists a measurable function* $h(x\zeta_0)$ *such that* $|\mathrm{E}(f_n(s, x\zeta_0) \mid x\zeta_0)| \leqslant h(x\zeta_0)$, *with* $\mathrm{E}|h(x\zeta_0) A(s \mid x\zeta_0, \delta_0)| < \infty$ *for large n.*
(*4*) $\lim_{n\to\infty} \mathrm{E}(f_n(s, x\zeta_0) \mid x\zeta_0) = c(x\zeta_0)$ *a.e.*

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_n(s_i, x_i\zeta_0)[A_n(s \mid x_i\zeta_0) - A(s \mid x_i\zeta_0, \delta_0)] \xrightarrow{D} \mathrm{N}(0, \Phi),$$

where

$$\Phi = \mathrm{E}\{[s \cdot c(x\zeta_0) g(x\zeta_0 \mid \delta_0) - \mathrm{E}(s \cdot c(x\zeta_0) g(x\zeta_0 \mid \delta_0))]' \times [s \cdot c(x\zeta_0) g(x\zeta_0 \mid \delta_0) - \mathrm{E}(s \cdot c(x\zeta_0) g(x\zeta_0 \mid \delta_0))]\}.$$

Proof.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_n(s_i, x_i\zeta_0)[A_n(s \mid x_i\zeta_0) - A(s \mid x_i\zeta_0, \delta_0)] = \frac{1}{(n-1)\sqrt{n}} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \Phi_n(s_i, s_j, a_n),$$

where

$$\Phi_n(\bar{s}_i, \bar{s}_j, a_n) = f_n(s_i, x_i\zeta_0)\left[\frac{1}{a_n^r} s_j K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n}\right) - A(s \mid x_i\zeta, \delta_0)\right].$$

By (1), $\mathrm{E}(\Phi_n(\bar{s}_i, \bar{s}_j, a_n) \mid \bar{s}_i)$ converges to zero. On the other hand,

$$\mathrm{E}(\Phi_n(\bar{s}_j, \bar{s}_i, a_n) \mid \bar{s}_i) = s_i \mathrm{E}\left[f_n(s_j, x_j\zeta_0) \frac{1}{a_n^r} K\left(\frac{x_j\zeta_0 - x_i\zeta_0}{a_n}\right) \middle| x_i \right] - \mathrm{E}\{\mathrm{E}[f_n(s, x\zeta_0) \mid x\zeta_0] A(s \mid x\zeta_0, \delta_0)\}.$$

Since

$$\lim_{n\to\infty}\left\{\mathrm{E}\left[f_n(s_j, x_j\zeta_0)\frac{1}{a_n^r}K\left(\frac{x_j\zeta_0 - x_i\zeta_0}{a_n}\right)\bigg| x_i\right] - \mathrm{E}[f_n(s, x\zeta_0)|x\zeta_0 = x_i\zeta_0]g(x_i\zeta_0|\delta_0)\right\} = 0$$

by condition (2), it follows that, by conditions (3) and (4) and the Lebesque convergence theorem,

$$\lim_{n\to\infty}\mathrm{E}(\Phi_n(\bar{s}_j, \bar{s}_i, a_n)|\bar{s}_i) = s_i c(x_i\zeta_0)g(x_i\zeta_0|\delta_0) - \mathrm{E}(c(x\zeta_0)\mathrm{E}(s|x\zeta_0)g(x\zeta_0\,|\,\delta_0)) \text{ a.e.}$$

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 $\int_D |K(v)|\,\mathrm{d}v < \infty$ *and with a bandwidth* a_n. *Let*

$$D_n(s\,|\,x_i\zeta_0) = \frac{1}{(n-1)a_n^{r+1}}\sum_{j\neq i}^{n} s_j \nabla K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n}\right),$$

where $\nabla K(v) = \partial K(v)/\partial v$ *is the gradient vector of K, and let* $g(x\zeta_0\,|\,\delta_0)$ *be the density of* $x\zeta_0$. *Denote*

$$D(s\,|\,x\zeta_0, \delta_0) = \frac{\partial}{\partial(x\zeta_0)}\{\mathrm{E}(s\,|\,x\zeta_0)g(x\zeta_0\,|\,\delta_0)\}.$$

Let $f_n(s, x\zeta_0)$ *be measurable functions such that* $\sup_n |f_n(s, x\zeta_0)|$ *is square integrable. Suppose that:*

(1) $K(v)$ *vanishes at the boundary of D.*
(2) $(\partial/\partial(x\zeta_0))\{\mathrm{E}(s\,|\,x\zeta_0)g(x\zeta_0\,|\,\delta_0)\}$ *is uniformly continuous at* $x\zeta_0$.
(3) $\mathrm{E}(f_n(s, x\zeta_0)\,|\,x\zeta_0) = 0$ *a.e., for all n.*

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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_n(s_i, x_i\zeta_0)[D_n(s\,|\,x_i\zeta_0) - D(s\,|\,x_i\zeta_0, \delta_0)] \xrightarrow{p} 0.$$

Proof. Define

$$\Phi_n(\bar{s}_i, \bar{s}_j, a_n) = f_n(s_i, x_i\zeta_0)\left[\frac{1}{a_n^{r+1}} s_j \nabla K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n}\right) - D(s|x_i\zeta_0, \delta_0)\right].$$

It follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_n(s_i, x_i\zeta_0)[D_n(s\,|\,x_i\zeta_0) - D(s\,|\,x_i\zeta_0, \delta_0)] = \frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j\neq 1}^{n}\Phi_n(\bar{s}_i, \bar{s}_j, a_n).$$

By conditions (1) and (2), Lemma 3 implies that

$$\mathrm{E}\left[s_j \frac{1}{a_n^{r+1}} \nabla K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n}\right)\bigg| x_i\right] \quad \text{converges to} \quad D(s\,|\,x_i\zeta_0, \delta_0).$$

Hence $\mathrm{E}(\Phi_n(\bar{s}_i, \bar{s}_j, a_n)|\bar{s}_i)$ converges to zero. On the other hand, under condition (3),

$$\begin{aligned}
\mathrm{E}(\Phi_n(\bar{s}_j, \bar{s}_i, a_n)|\bar{s}_i) &= s_i \mathrm{E}\left[f_n(s_j, x_j\zeta_0)\frac{1}{a_n^{r+1}} \nabla K\left(\frac{x_j\zeta_0 - x_i\zeta_0}{a_n}\right)\bigg| x_i\right] \\
&\quad - \mathrm{E}[f_n(s_j, x_j\zeta_0) D(s\,|\,x_j\zeta_0, \delta_0)] \\
&= s_i \frac{1}{a_n}\int_D \mathrm{E}[f_n(s, x_i\zeta_0 + a_n v)|x_j\zeta_0 = x_i\zeta_0 + a_n v, x_i] \\
&\quad \times g(x_i\zeta_0 + a_n v\,|\,\delta_0)\nabla K(v)\,\mathrm{d}v \\
&\quad - \mathrm{E}\{\mathrm{E}[f_n(s, x\zeta_0)|x\zeta_0] D(s\,|\,x\zeta_0, \delta_0)\} \\
&= 0.
\end{aligned}$$

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

A.5. Asymptotic properties

The asymptotic distributions of

$$\frac{1}{\sqrt{n}}\hat{X}_2'\hat{U}_n, \qquad \frac{1}{\sqrt{n}}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{U}_n, \qquad \frac{1}{\sqrt{n}}\frac{\partial \mathrm{E}_n(\hat{\beta}, \hat{\zeta})}{\partial \delta}\hat{\Lambda}_n^{-1}\hat{U}_n$$

can be derived from Propositions 1–6.

$$\begin{aligned}
\frac{1}{\sqrt{n}}\hat{X}_2'\hat{U}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^n t_n(x_i\hat{\zeta}, \hat{\xi}_n)(x_{(2)i} - \mathrm{E}_n(x_{(2)}|x_i\hat{\zeta}))'(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\hat{\zeta})) \\
&= C_{0,n} + \{C_{1,n} + C_{2,n} + C_{3,n}\}\sqrt{n}(\hat{\delta} - \delta_0) + C_{4,n}\sqrt{n}(\hat{\xi}_n - \xi_0),
\end{aligned}$$

where

$$C_{0,n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n t_n(x_i\zeta_0, \xi_0)(x_{(2)i} - \mathrm{E}_n(x_{(2)}|x_i\zeta_0))'(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\zeta_0)),$$

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

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

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

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

Proposition 1 implies that

$$C_{1,n} \xrightarrow{p} -\mathrm{E}\left(I_T(x\zeta_0)(x_{(2)} - \mathrm{E}(x_{(2)}|x\zeta_0))'\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\partial\delta'}\right),$$

and $C_{2,n}$ converges to zero in probability. As $\lim_{n\to\infty}(n/\ln n)\,h_n^{(1+4/r)} = \infty$, Proposition 2 implies that $C_{3,n}$ and $C_{4,n}$ converge to zero in probability. To see the latter, consider the single-index model for simplicity. From (A.3.2), since q has support on $[0,1]$ and is vanishing outside $[0,1]$, for large n (A.3.2) can be rewritten as

$$\frac{\partial t_n(x\zeta,\xi)}{\partial\delta'} = \frac{1}{h_n}q\left(\frac{x\zeta - \xi_p}{h_n}\right)x_{(2)} - \frac{1}{h_n}q\left(\frac{x\zeta - \xi_{(1-p)} + h_n}{h_n}\right)x_{(2)};$$

(A.3.3) can be written as

$$\frac{\partial t_n(x\zeta,\xi)}{\partial\xi_p} = -\frac{1}{h_n}q\left(\frac{x\zeta - \xi_p}{h_n}\right);$$

and (A.3.4) is

$$\frac{\partial t_n(x\zeta,\xi)}{\partial\xi_{(1-p)}} = \frac{1}{h_n}q\left(\frac{x\zeta - \xi_{(1-p)} + h_n}{h_n}\right).$$

Hence $C_{3,n} = C_{3,n}^{(1)} - C_{3,n}^{(2)}$ where

$$C_{3,n}^{(1)} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}q\left(\frac{x_i\bar{\zeta} - \bar{\xi}_p}{h_n}\right)'(u_{1i} - \mathrm{E}_n(u_1\,|x_i\bar{\zeta}))(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\bar{\zeta}))'x_{(2)i},$$

and

$$C_{3,n}^{(2)} = -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}q\left(\frac{x_i\bar{\zeta} - \bar{\xi}_{(1-p)} + h_n}{h_n}\right)$$
$$\times(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\bar{\zeta}))(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\bar{\zeta}))'x_{(2)}i.$$

By the equivalent expressions for (A.3.3) and (A.3.4) above, $C_{4,n}$ has a similar expression. Proposition 2 is applicable to these terms. For example,

$$\begin{aligned}\operatorname*{plim}_{n\to\infty} C_{3,n}^{(1)} &= \mathrm{E}[(u_1 - \mathrm{E}(u_1\,|\,x\zeta_0))(x_{(2)} - \mathrm{E}(x_{(2)}|x\zeta_0))'x_{(2)}|x\zeta_0]p(x\zeta_0|\delta_0)|_{x\zeta_0=\xi_p(\delta_0)}\\ &= \mathrm{E}[(u_1 - \mathrm{E}(u_1|x\zeta_0))(x_{(2)} - \mathrm{E}(x_{(2)}|x\zeta_0))'x_{(2)}|x\zeta_0]p(x\zeta_0|\delta_0)|_{x\zeta_0=\xi_p(\delta_0)}\\ &= 0,\end{aligned}$$

because $\mathrm{E}(u_1 - \mathrm{E}(u_1 \mid x\zeta_0) \mid x) = 0$. Thus, both $C_{3,n}$ and $C_{4,n}$ converge to zero in probability.

Consider the term $C_{0,n}$. To simplify notations, let

$$A_{ni} = \frac{1}{(n-1)a_n^m} \sum_{j \neq i}^{n} x_{2j} K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n} \right),$$

$$B_{ni} = \frac{1}{(n-1)a_n^m} \sum_{j \neq i}^{n} K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n} \right),$$

$$C_{ni} = \frac{1}{(n-1)a_n^m} \sum_{j \neq i}^{n} u_{1j} K\left(\frac{x_i\zeta_0 - x_j\zeta_0}{a_n} \right).$$

Let $A_i = \mathrm{E}(x_{(2)i} \mid x_i\zeta_0) p(x_i\zeta_0 \mid \delta_0)$, $B_i = p(x_i\zeta_0 \mid \delta_0)$, and $C_i = \mathrm{E}(u_{1i} \mid x_i\zeta_0) \times p(x_i\zeta_0 \mid \delta_0)$ be their limits. By a Taylor expansion,

$$C_{0,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)(x_{(2)i} - \mathrm{E}(x_{(2)i} \mid x_i\zeta_0))'(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) + L_n + R_n,$$

where

$$\begin{aligned} L_n = & -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) \frac{1}{B_i}(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0))(A_{ni} - A_i)' \\ & -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)(x_{(2)i} - \mathrm{E}(x_{(2)i} \mid x_i\zeta_0))' \frac{1}{B_i}(C_{ni} - C_i) \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)[\mathrm{E}(x_{(2)i} \mid x_i\zeta_0)'(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \\ & + (x_{(2)i} - \mathrm{E}(x_{(2)i} \mid x_i\zeta_0))' \mathrm{E}(u_{1i} \mid x_i\zeta_0)] \frac{1}{B_i}(B_{ni} - B_i) \end{aligned}$$

and

$$\begin{aligned} R_n = & \frac{2}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) \frac{1}{\tilde{B}_{ni}^2}(A_{ni} - A_i)'(C_{ni} - C_i) \\ & + \frac{2}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)[(u_{1i}\tilde{B}_{ni} - \tilde{C}_{ni}) - \tilde{C}_{ni}] \frac{1}{\tilde{B}_{ni}^3}(A_{ni} - A_i)'(B_{ni} - B_i) \\ & + \frac{2}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)[-\tilde{A}'_{ni} + (x_{2i}\tilde{B}_{ni} - \tilde{A}_{ni})'] \frac{1}{\tilde{B}_{ni}^3}(B_{ni} - B_i)(C_{ni} - C_i) \\ & + \frac{2}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)[-\tilde{A}'_{ni}(u_{1i}\tilde{B}_{ni} - \tilde{C}_{ni}) + \tilde{A}'_{ni}\tilde{C}_{ni} - (x_{2i}\tilde{B}_{ni} - \tilde{A}_{ni})'\tilde{C}_{ni}] \\ & \times \frac{1}{\tilde{B}_{ni}^4}(B_{ni} - B_i)^2. \end{aligned}$$

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

$$\frac{1}{\sqrt{n}}\hat{X}_2'\hat{U}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_0,\xi_0)(x_{(2)i} - \mathrm{E}(x_{(2)i}\,|\,x_i\zeta_0))'(u_{1i} - \mathrm{E}(u_{1i}\,|\,x_i\zeta_0))$$

$$- \mathrm{E}\left(I_T(x\zeta_0)(x_{(2)} - \mathrm{E}(x_{(2)}\,|\,x\zeta_0))'\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\partial\delta'}\right)$$

$$\times\sqrt{n}(\hat{\delta} - \delta_0) + o_p(1), \qquad \text{(A.5.1)}$$

which is asymptotically normal $\mathrm{N}(0,\Delta)$, where Δ is defined in (4.11), under the property that $\sqrt{n}(\hat{\delta} - \delta_0)$ is asymptotically uncorrelated with $(u_1 - \mathrm{E}(u_1\,|\,x\zeta_0))$.

For the asymptotic distribution of the SG2LS, by a mean value theorem

$$\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{U}_n$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\hat{\zeta},\hat{\xi}_n)(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\hat{\zeta}))'\omega_{ni}^{-1}(\hat{\beta},\hat{\delta})(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\hat{\zeta}))$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_0,\xi_0)(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\zeta_0))'\omega_{ni}^{-1}(\beta_0,\delta_0)(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\zeta_0))$$

$$- \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\bar{\zeta},\bar{\xi})(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\bar{\zeta}))'\omega_{ni}^{-1}(\bar{\beta},\bar{\delta})\frac{\partial \mathrm{E}_n(u_1\,|\,x_i\bar{\zeta})}{\partial\delta'}\sqrt{n}(\hat{\delta} - \delta_0)$$

$$+ \frac{1}{n}\sum_{i=1}^{n}(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\bar{\zeta}))\frac{\partial}{\partial\delta'}[t_n(x_i\bar{\zeta},\bar{\xi})(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\bar{\zeta}))'\omega_{ni}^{-1}(\bar{\beta},\bar{\delta})]$$

$$\times\sqrt{n}(\hat{\delta} - \delta_0) + \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\bar{\zeta},\bar{\xi})(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\bar{\zeta}))(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\bar{\zeta}))'$$

$$\times\frac{\partial\omega_{ni}^{-1}(\bar{\beta},\bar{\delta})}{\partial\beta'}\sqrt{n}(\hat{\beta} - \beta_0) + \frac{1}{n}\sum_{i=1}^{n}(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\bar{\zeta}))(x_{(2)i} - \mathrm{E}_n(x_{(2)}\,|\,x_i\bar{\zeta}))'$$

$$\times\omega_{ni}^{-1}(\bar{\beta},\bar{\delta})\frac{\partial t_n(x_i\bar{\zeta},\bar{\xi})}{\partial\xi'}\sqrt{n}(\hat{\xi}_n - \xi_0)$$

$$= C_{0,n}^{(\omega)} - \mathrm{E}\left[I_T(x\zeta_0)(x_{(2)} - \mathrm{E}(x_{(2)}\,|\,x\zeta_0))'\omega^{-1}(x\zeta_0,\delta_0)\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\partial\delta'}\right]$$

$$\times\sqrt{n}(\hat{\delta} - \delta_0) + o_p(1),$$

where

$$C_{0,n}^{(\omega)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)(x_{(2)i} - \mathrm{E}_n(x_{(2)} \mid x_i\zeta_0))' \times \omega_{ni}^{-1}(\beta_0, \delta_0)(u_{1i} - \mathrm{E}_n(u_1 \mid x_i\zeta_0)).$$

To simplify notations, let

$$H_{ni} = \frac{1}{(n-1)b_n^m} \sum_{j \neq i}^{n} H\left(\frac{x_i\zeta_0 - x_j\zeta_0}{b_n}\right),$$

$$R_{ni} = \frac{1}{(n-1)b_n^m} \sum_{j \neq i}^{n} u_{1j} H\left(\frac{x_i\zeta_0 - x_j\zeta_0}{b_n}\right),$$

$$S_{ni} = \frac{1}{(n-1)b_n^m} \sum_{j \neq i}^{n} u_{1j}^2 H\left(\frac{x_i\zeta_0 - x_j\zeta_0}{b_n}\right).$$

Let $H_i = B_i$, $R_i = C_i$, and $S_i = \mathrm{E}(u_{1i}^2 \mid x_i\zeta_0) p(x_i\zeta_0 \mid \delta_0)$ be their limits respectively. By Proposition 3, as $\lim_{n\to\infty} n a_n^{2m} = \infty$, $\lim_{n\to\infty} n a_n^{4s^*} = 0$, $\lim_{n\to\infty} n b_n^{2m} = \infty$, and $\lim_{n\to\infty} n b_n^{4h^*} = 0$,

$$C_{0,n}^{(\omega)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)(x_{(2)i} - \mathrm{E}(x_{(2)i} \mid x_i\zeta_0))' \times \omega^{-1}(x_i\zeta_0, \delta_0)(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) + L_n^{(\omega)} + o_p(1),$$

where

$$\begin{aligned} L_n^{(\omega)} = &- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)\omega^{-1}(x_i\zeta_0, \delta_0)(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \frac{1}{B_i}(A_{ni} - A_i)' \\ &- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)\omega^{-1}(x_i\zeta_0, \delta_0)(x_{(2)i} - \mathrm{E}(x_{(2)i} \mid x_i\zeta_0))' \frac{1}{B_i}(C_{ni} - C_i) \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)\omega^{-1}(x_i\zeta_0, \delta_0)[\mathrm{E}(x_{(2)i} \mid x_i\zeta_0)'(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \\ &+ (x_{(2)i} - \mathrm{E}(x_{(2)i} \mid x_i\zeta_0))'\mathrm{E}(u_{1i} \mid x_i\zeta_0)] \frac{1}{B_i}(B_{ni} - B_i) \\ &- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)\omega^{-2}(x_i\zeta_0, \delta_0)(x_{(2)i} - \mathrm{E}(x_{(2)i} \mid x_i\zeta_0))' \\ &\times (u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \frac{1}{H_i} \{(S_{ni} - S_i) - 2\mathrm{E}(u_{1i} \mid x_i\zeta_0)(R_{ni} - R_i) \\ &- [\mathrm{E}(u_{1i}^2 \mid x_i\zeta_0) - 2\mathrm{E}(u_{1i} \mid x_i\zeta_0)](H_{ni} - H_i)\}. \end{aligned}$$

Since $E[x_{(2)i} - E(x_{(2)i} | x_i\zeta_0)|x_i\zeta_0] = 0$, $E[u_{1i} - E(u_{1i} | x_i\zeta_0)|x_i] = 0$, and the other functions in $L_n^{(\omega)}$ are functions of $x_i\zeta_0$, 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,

$$\frac{1}{\sqrt{n}} \frac{\partial E_n(\hat{\beta}, \hat{\zeta})}{\partial \delta} \hat{\Lambda}_n^{-1} \hat{U}_n$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\hat{\zeta}, \hat{\xi}_n) \frac{\partial E_n(y - w\hat{\beta} | x_i\hat{\zeta})}{\partial \delta} \omega_{ni}^{-1}(\hat{\beta}, \hat{\delta})(u_{1i} - E_n(u_1 | x_i\hat{\zeta}))$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) \frac{\partial E_n(u_1 | x_i\zeta_0)}{\partial \delta} \omega_{ni}^{-1}(\beta_0, \delta_0)(u_{1i} - E_n(u_1 | x_i\zeta_0))$$

$$- \frac{1}{n} \sum_{i=1}^{n} t_n(x_i\bar{\zeta}, \bar{\xi}) \frac{\partial E_n(y - w\bar{\beta} | x_i\bar{\zeta})}{\partial \delta} \omega_{ni}^{-1}(\bar{\beta}, \bar{\delta}) \frac{\partial E_n(u_1 | x_i\bar{\zeta})}{\partial \delta'} \sqrt{n}(\hat{\delta} - \delta_0)$$

$$+ \frac{1}{n} \sum_{i=1}^{n} (u_{1i} - E_n(u_1 | x_i\bar{\zeta})) \frac{\partial}{\partial \delta'} \left[t(x_i\bar{\zeta}, \bar{\xi}) \frac{\partial E_n(y - w\bar{\beta} | x_i\bar{\zeta})}{\partial \delta} \omega_{ni}^{-1}(\bar{\beta}, \bar{\delta}) \right]$$

$$\times \sqrt{n}(\hat{\delta} - \delta_0) + \frac{1}{n} \sum_{i=1}^{n} t_n(x_i\bar{\zeta}, \bar{\xi})(u_{1i} - E_n(u_1 | x_i\bar{\zeta}))$$

$$\times \frac{\partial}{\partial \beta'} \left[\frac{\partial E_n(y - w\bar{\beta} | x_i\bar{\zeta})}{\partial \delta} \omega_{ni}^{-1}(\bar{\beta}, \bar{\delta}) \right] \sqrt{n}(\hat{\beta} - \beta_0)$$

$$+ \frac{1}{n} \sum_{i=1}^{n} (u_{1i} - E_n(u_1 | x_i\bar{\zeta})) \frac{\partial E_n(y - w\bar{\beta} | x_i\bar{\zeta})}{\partial \delta}$$

$$\times \omega_{ni}^{-1}(\bar{\beta}, \bar{\delta}) \frac{\partial t(x_i\bar{\zeta}, \bar{\xi})}{\partial \delta'} \sqrt{n}(\hat{\xi}_n - \xi_0)$$

$$= C_{0,n}^{(d)} - E\left[I_T(x\zeta_0) \frac{\partial E(u_1 | x\zeta_0, \delta_0)}{\partial \delta} \omega^{-1}(x\zeta_0, \delta_0) \frac{\partial E(u_1 | x\zeta_0, \delta_0)}{\partial \delta'} \right]$$

$$\times \sqrt{n}(\hat{\delta} - \delta_0) + o_p(1),$$

where

$$C_{0,n}^{(d)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) \frac{\partial E_n(u_1 | x_i\zeta_0, \delta_0)}{\partial \delta} \omega_n^{-1}(\beta_0, \delta_0)(u_{1i} - E_n(u_1 | x_i\zeta_0)).$$

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

$$C_{0,n}^{(d)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) \frac{\partial E(u_1 | x_i\zeta_0, \delta_0)}{\partial \delta} \omega^{-1}(x_i\zeta_0, \delta_0)(u_{1i} - E(u_{1i} | x_i\zeta_0))$$

$$+ L_{0,n}^{(d)} + o_p(1),$$

where

$$
\begin{aligned}
L_{0,n}^{(d)} = & -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) \frac{\partial \mathrm{E}(u_1 \mid x_i\zeta_0, \delta_0)}{\partial \delta} \omega^{-1}(x_i\zeta_0, \delta_0) \frac{1}{B_i}(A_{ni} - A_i) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)\omega^{-1}(x_i\zeta_0, \delta_0) \Bigg[\frac{\partial \mathrm{E}(u_1 \mid x_i\zeta_0, \delta_0)}{\partial \delta} A_i \\
& - (u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \left\{ \frac{\partial C_i}{\partial \delta} - 2\mathrm{E}(u_{1i} \mid x_i\zeta_0) \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_0, \xi_0)\omega^{-1}(x_i\zeta_0, \delta_0)(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \\
& \times \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_0, \xi_0)\omega^{-1}(x_i\zeta_0, \delta_0) \\
& \times (u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0))\mathrm{E}(u_{1i} \mid x_i\zeta_0) \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_0, \xi_0)\omega^{-1}(x_i\zeta_0, \delta_0)(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \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_0, \xi_0)\omega^{-2}(x_i\zeta_0, \delta_0)(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \frac{\partial \mathrm{E}(u_1 \mid x_i\zeta_0, \delta_0)}{\partial \delta} \\
& \times \mathrm{E}(u_{1i} \mid x_i\zeta_0) \frac{1}{H_i}(R_{ni} - R_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) \frac{\partial \mathrm{E}(u_1 \mid x_i\zeta_0, \delta_0)}{\partial \delta} \\
& \times \omega^{-2}(x_i\zeta_0, \delta_0)(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \frac{1}{H_i}(S_{ni} - S_i) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)\omega^{-2}(x_i\zeta_0, \delta_0)(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \frac{\partial \mathrm{E}(u_1 \mid x_i\zeta_0, \delta_0)}{\partial \delta} \\
& \times \left(S_i - 2\frac{R_i^2}{H_i} \right) \frac{1}{H_i^2}(H_{ni} - H_i).
\end{aligned}
$$

Since

$$
\frac{\partial \mathrm{E}(u_1 \mid x\zeta_0, \delta_0)}{\partial \delta_k} = (x - \mathrm{E}(x \mid x\zeta_0))' \frac{\partial \zeta(\delta_0)}{\partial \delta_k} \nabla \mathrm{E}(u_1 \mid x\zeta_0),
$$

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 after some matrix manipulation and simplification that

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\Sigma}^{-1}\hat{U}_n \\
&= \frac{1}{\sqrt{n}}\hat{X}_2'\hat{\Lambda}_n^{-1}\hat{U}_n - \hat{X}_2'\hat{\Lambda}_n^{-1}\frac{\partial \mathrm{E}_n(\hat{\beta},\hat{\zeta})}{\partial\delta'}\left(\hat{V}_{n,\delta}^{-1} + \frac{\partial \mathrm{E}_n'(\hat{\beta},\hat{\zeta})}{\partial\delta}\hat{\Lambda}_n^{-1}\frac{\partial \mathrm{E}_n(\hat{\beta},\hat{\zeta})}{\partial\delta'}\right)^{-1} \\
&\quad\times\frac{1}{\sqrt{n}}\frac{\partial \mathrm{E}_n'(\hat{\beta},\hat{\zeta})}{\partial\delta}\hat{\Lambda}_n^{-1}\hat{U}_n \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_0,\xi_0)(x_{(2)i} - \mathrm{E}(x_{(2)i}\,|\,x_i\zeta_0))'\omega^{-1}(x_i\zeta_0,\delta_0)(u_{1i} - \mathrm{E}(u_{1i}\,|\,x_i\zeta_0)) \\
&\quad - \mathrm{E}\left[I_T(x\zeta_0)(x_{(2)} - \mathrm{E}(x_{(2)}\,|\,x\zeta_0))'\omega^{-1}(x\zeta_0,\delta_0)\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\partial\delta'}\right] \\
&\quad\times\left(V_\delta^{-1} + \mathrm{E}\left[I_T(x\zeta_0)\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\partial\delta'}\omega^{-1}(x\zeta_0,\delta_0)\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\partial\delta'}\right]\right)^{-1} \\
&\quad\times\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_0,\xi_0)\frac{\partial \mathrm{E}(u_1\,|\,x_i\zeta_0,\delta_0)}{\partial\delta}\omega^{-1}(x_i\zeta_0,\delta_0)(u_{li} - \mathrm{E}(u_{1i}\,|\,x_i\zeta_0))\right. \\
&\quad\left. + V_\delta^{-1}\sqrt{n}(\hat{\delta}-\delta_0)\right\} + \mathrm{o}_p(1) \\
&\xrightarrow{D} \mathrm{N}(0,\Gamma),
\end{aligned}
\tag{A.5.2}
$$

where Γ is in (5.12).

For the instrumental variable estimator in (5.18),

$$
\begin{aligned}
\frac{1}{\sqrt{n}}P'\hat{U}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_0,\xi_0)p_i'(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\zeta_0)) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n} t_n(x_i\bar{\zeta},\bar{\xi})p_i'\frac{\partial \mathrm{E}_n(u_1\,|\,x_i\bar{\zeta})}{\partial\delta'}\sqrt{n}(\hat{\delta}-\delta_0) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} p_i'(u_{1i} - \mathrm{E}_n(u_1\,|x_i\bar{\zeta}))\frac{\partial t_n(x_i\bar{\zeta},\bar{\xi})}{\partial\delta'}\sqrt{n}(\hat{\delta}-\delta_0) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} p_i'(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\bar{\zeta}))\frac{\partial t_n(x_i\bar{\zeta},\bar{\xi})}{\partial\xi'}\sqrt{n}(\hat{\xi}_n-\xi_0) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_0,\xi_0)p_i'(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\zeta_0)) \\
&\quad - \mathrm{E}\left[I_T(x\zeta_0)p'\frac{\partial \mathrm{E}(u_1\,|\,x\zeta_0,\delta_0)}{\partial\delta'}\right]\sqrt{n}(\hat{\delta}-\delta_0) + \mathrm{o}_p(1).
\end{aligned}
\tag{A.5.3}
$$

Since

$$\frac{\partial \mathrm{E}(u_1 \mid x\zeta_0, \delta_0)}{\partial \delta_k} = (x - \mathrm{E}(x \mid x\zeta_0))' \frac{\partial \zeta(\delta_0)}{\partial \delta_k} \nabla \mathrm{E}(u_1 \mid x\zeta_0),$$

it follows that

$$\mathrm{E}\left[\mathrm{E}(p' \mid x\zeta_0) \frac{\partial \mathrm{E}(u_1 \mid x\zeta_0, \delta_0)}{\partial \delta'} \right] = 0,$$

and therefore,

$$\mathrm{E}\left[I_T(x\zeta_0) p' \frac{\partial \mathrm{E}(u_1 | x\zeta_0, \delta_0)}{\partial \delta'} \right] = \mathrm{E}\left[I_T(x\zeta_0)(p - \mathrm{E}(p|x\zeta_0))' \frac{\partial \mathrm{E}(u_1 | x\zeta_0, \delta_0)}{\partial \delta'} \right].$$

On the other hand,

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) p_i'(u_{1i} - \mathrm{E}_n(u_1 \mid x_i\zeta_0)) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) p_i'(u_{1i} - \mathrm{E}(u_{1i} \mid x_i\zeta_0)) \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) p_i' \frac{1}{B_i}(C_{ni} - C_i) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) p_i' \mathrm{E}(u_{1i} \mid x_i\zeta_0) \frac{1}{B_i}(B_{ni} - B_i) + o_p(1).
\end{aligned}$$

With a high-order kernel of orders s^* such that $\lim_{n\to\infty} a_n^{2s^*} = 0$,[11] Proposition 5 implies

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0) p_i' \frac{1}{B_i}(C_{ni} - C_i) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ I_T(x_i\zeta_0) u_{1i} \mathrm{E}(p_i' | x_i\zeta_0) - \mathrm{E}[I_T(x_i\zeta_0) \mathrm{E}(u_{1i} | x_i\zeta_0) \mathrm{E}(p_i' | x_i\zeta_0)] \} + o_p(1)
\end{aligned}$$

[11] This rate requirement implies $\lim_{n\to\infty} n a_n^{4s^*} = 0$ 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 $p_i - \mathrm{E}_n(p \mid x_i\hat{\zeta})$ 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 $p_i - \mathrm{E}_n(p \mid x_i\hat{\zeta})$ as an instrumental variable instead of p_i.

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)p_i' \mathrm{E}(u_{1i}\,|\,x_i\zeta_0)\frac{1}{B_i}(B_{ni} - B_i)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \{I_T(x_i\zeta_0)\,\mathrm{E}(p_i'\,|\,x_i\zeta_0)\mathrm{E}(u_{1i}\,|\,x_i\zeta_0)$$

$$- \mathrm{E}[I_T(x_i\zeta_0)\mathrm{E}(p_i'\,|\,x_i\zeta_0)\mathrm{E}(u_{1i}\,|\,x_i\zeta_0)]\} + \mathrm{o}_p(1).$$

Therefore,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(x_i\zeta_0, \xi_0)\,p_i'(u_{1i} - \mathrm{E}_n(u_1\,|\,x_i\zeta_0))$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_T(x_i\zeta_0)(p_i - \mathrm{E}(p_i\,|\,x_i\zeta_0))'(u_{1i} - \mathrm{E}(u_{1i}\,|\,x_i\zeta_0)) + \mathrm{o}_p(1). \quad \text{(A.5.4)}$$

References

Amemiya, T., 1978, The estimation of a simultaneous equation generalized probit model, Econometrica 46, 1193–1206.

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.

Bierens, H.J., 1985, Kernel estimators of regression functions, in: T.W. Bewley, ed., Advances in econometrics, Fifth world congress, Vol. 1 (Cambridge University Press, New York, NY) 99–144.

Blundell, R.W. and R.J. Smith, 1989, Estimation in a class of simultaneous equation limited dependent variable models, Review of Economic Studies 56, 37–58.

Chamberlain, G., 1986, Asymptotic efficiency in semi-parametric models with censoring, Journal of Econometrics 32, 189–218.

Cosslett, S., 1987, Efficiency bounds for distribution-free estimators of the censored regression models, Econometrica 55, 559–585.

Heckman, J.J., 1979, Sample bias as specification error, Econometrica 47, 153–162.

Ichimura, H., 1987, Estimation of single index models, Ph.D. thesis (Department of Economics, MIT, Cambridge, MA).

Ichimura, H. and L.F. Lee, 1991, Semiparametric estimation of multiple index models: Single equation estimation, Ch. 1 in: W.A. Barnett, J. Powell, and G. Tauchen, eds., Nonparametric and semiparametric methods in econometrics and statistics (Cambridge University Press, New York, NY).

Klein, R.W. and R.H. Spady, 1987, An efficient semiparametric estimator for discrete choice models, Economics discussion paper no. 67 (Bellcore, Morristown, NJ).

LeCam, L., 1960, Locally asymptotically normal families of distributions, University of California Publications in Statistics 3, 37–98.

Lee, L.F., 1981, Simultaneous equations models with discrete endogenous variables, Ch. 9 in: C.F. Manski and D. McFadden, eds., Structural analysis of discrete data and econometric applications (MIT Press, Cambridge, MA).

Lee, L.F., 1990, Efficient semiparametric scoring estimation of sample selection models, Discussion paper no. 255 (Center for Economic Research, University of Minnesota, Minneapolis, MN).

Lee, L.F., 1992, Semiparametric nonlinear least-squares estimation of truncated regression models, Econometric Theory 8, 52–94.
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.
Manski, C.F., 1984, Adaptive estimation of non-linear regression models, Econometric Reviews 3, 145–194.
Nadaraja, E., 1964, On regression estimator, Theory of Probability and Its Applications 9, 157–159.
Newey, W.K., 1987, Efficient estimation of limited dependent variable models with endogenous explanatory variables, Journal of Econometrics 36, 231–250.
Olsen, R.J., 1980, A least squares correction for selectivity bias, Econometrica 48, 1815–1820.
Powell, J.L., 1987, Semiparametric estimation of bivariate latent variable models, Discussion paper no. 8704 (Social Systems Research Institute, University of Wisconsin, Madison, WI).
Powell, J.L., J.H. Stock, and T.M. Stoker, 1989, Semiparametric estimation of index coefficients, Econometrica 57, 1403–1430.
Robinson, P.M., 1988, Root-*n*-consistent semiparametric regression, Econometrica 56, 931–954.
Serfling, R.J., 1980, Approximation theorems of mathematical statistics (Wiley, New York, NY).
Silverman, B.W., 1986, Density estimation for statistics and data analysis (Chapman and Hall, New York, NY).
Stoker, T.M., 1986, Consistent estimation of scaled coefficients, Econometrica 54, 1461–1481.
Watson, G.S., 1964, Smooth regression analysis, Sankha Series A 26, 101–116.