SEMIPARAMETRIC TWO STAGE ESTIMATION OF SAMPLE SELECTION MODELS SUBJECT TO TOBIT-TYPE SELECTION RULES

by

LUNG-FEI LEE

Discussion Paper No. 256, July 1990

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

Semiparametric Two Stage Estimation of Sample Selection Models Subject to Tobit-type Selection Rules

by Lung-Fei Lee

Department of Economics

University of Minnesota

Minneapolis, MN 55455

Date: July, 1990

Abstract

A semiparametric two stage estimation method is proposed for the estimation of sample selection models which are subject to Tobit-type selection rules. With randomization restrictions on the disturbances of the model, all the regression coefficients in the model are in general identifiable without exclusion restrictions. The proposed estimator is shown to be $\sqrt{n}$-consistent and asymptotically normal. Some Monte Carlo results to demonstrate its finite sample performance are also provided.

JEL classification number: 211

Sample selection, truncation, censoring, semiparametric estimation, randomization, kernel estimation, regression model

Correspondence Address:

Lung-Fei Lee, 1035 Management and Economics, Department of Economics, University of Minnesota, Minneapolis, MN 55455

Semiparametric Two Stage Estimation of Sample Selection Models Subject to Tobit-type Selection Rules

by

Lung-Fei Lee*

1. Introduction:

Microeconometric models of discrete choice, limited dependent variables and sample selection have found interesting applications in empirical studies with microeconomic data. Models with parametric distributions, however, may subject to distributional misspecification which might result in inconsistent estimates. Recent research efforts on estimation for such models have focused on semiparametric and distribution-free methods which relax parametric distribution assumptions. Various semiparametric methods have been proposed for the estimation of discrete choice models, censored or truncated regression models and sample selection models with discrete choice decision rules. Semiparametric estimation of sample selection models subject to Tobit-type selection rules has not been explicitly considered in the literature.

Tobit-type sample selection models differ from sample selection models with discrete choice rules in that the decision equations in such models are Tobit equations instead of discrete choice equations. An example is a model of female labor supply in Heckman [1974] where the market wages can be observed only for the individuals whose hours of work are positive. Consider a two equations model,

$$y_1 = x\alpha + u$$

$$y_2 = x\beta + v \tag{1.1}$$

where y_1 and y_2 can be observed only when $y_1 > 0$. This model provides much more information than the model with a discrete choice equation for y_1 in that positive values of y_1 can be observed instead of just its sign. With parametric distribution specification such as normal disturbances,

* I appreciate having financial support under NSF grant no. SES-8809939 and SES-9010516, and computing grant support from the Minnesota Supercomputer Institute for my research.

the parameter vector β in the outcome equation y_2 can be identified for models with either a discrete choice equation or a Tobit-type equation. For semiparametric models without parametric distribution assumptions, the identification issue for β can be quite different. For the model with a discrete choice equation for y_1, Chamberlain [1986] has shown that, under the assumption that (u, v) is independent with regressors in the model, identification of β requires exclusion restrictions on the regressors of the y_2 equation. Semiparametric methods for estimation of such a model have been suggested in Cosslett [1984], Robinson [1988] , Powell [1987], Ichimura and Lee [1988], Newey [1988] and Lee [1990]. For the model with a Tobit-type selection equation, observability of y_1 in a continuous range may provide enough restrictions for the identification of β. This article will propose a semiparametric method for the estimation of β. The estimation procedure is a simple two stage method. Given a consistent estimate of α, the bias of the observed outcome equation can be adjusted and β can be estimated by a regression procedure. Our procedure differs from the two stage estimation procedures in Heckman [1976], Cosslett [1984], Robinson [1988] and Powell [1987] in the way of constructing the bias adjustment term. The adjustment terms in their approaches are functions of the regressors in the model. The adjustment term in our approach is a function of the index $x\alpha$ and the observed values of y_1. Our adjustment term is designed for Tobit-type sample selection models. With general regularity conditions, our two stage estimator is $\sqrt{n}$-consistent asymptotically normal.

This article is organized as follows. Section 2 describes the estimation procedure. Regularity conditions for our model are listed in this section. Consistency of the estimator is discussed in Section 3. Asymptotic distribution of the estimator is described in Section 4. Some proofs of such asymptotic properties are provided in Appendix 2. Appendix 1 summarizes some relevant results for our analysis. Section 5 provides a consistent estimate of the covariance matrix of the estimator. Some Monte Carlo simulations are performed to investigate finite sample performance of the proposed estimator. The simulation results are reported in Section 6. Appendix 3 provides some corresponding first stage estimates of the decision equation for convenient reference.

2. A Two Stage Semiparametric Estimation Procedure

The model that will be considered in this article is a two equations model with a Tobit-type selection equation. Let x be the vector of regressors in the model. To be specific, x does not include any constant term [1]. The x_1 and x_2 are subvectors of x. The underlying latent equations in the model are

$$y_1 = x_1\alpha_o + u \tag{2.1}$$

and

$$y_2 = x_2\beta_o + v. \tag{2.2}$$

Values of y_1 and y_2 can be observed only when $y_1 > 0$. Equation (2.1) is a censored regression model if negative sign of y_1 can also be observed. It will be a truncated regression model if sample observations are available only for $y_1 > 0$. The disturbances u and v in (2.1) and (2.2) are assumed to be independent with x in the model. Sample observations are assumed to be random. Various semiparametric methods have been suggested for estimation of the Tobit model (2.1). If observations of y_1 are censored, α_o can be estimated by, for example, Powell's least absolute deviations method (Powell [1984]). For truncated sample, the method in Lee [1988] is applicable. Those estimators have been shown to be $\sqrt{n}$-consistent asymptotically normal. Estimation of β_o is the remaining issue.

Conditional on the observed sample y_2, the regression function of y_2 is

$$E(y_2|y_1 > 0, x) = x_2\beta_o + E(v|u > -x_1\alpha_o). \tag{2.3}$$

The two stage estimation method in Powell [1987] and the semiparametric nonlinear least squares method in Ichimura and Lee [1988] have used the "index property" that $E(v|u > -x_1\alpha_o)$ is a function of $x_1\alpha_o$ but not the "independence property" that u and v are independent with x in the latent structure. When $x_1 = x_2$ or x_2 is a subvector of x_1, index property alone does not provide enough restrictions for the identification of β_o (see Powell [1987] and Ichimura and Lee [1988]). For

[1] For our model, since no moment restrictions are imposed on the disturbances, constant terms in the equations are absorbed into the disturbances.

our model, let $f(v,u)$ be the joint density of (v,u) in (2.1) and (2.2) and let $f_u(\cdot)$ be the marginal density of u. Then

$$E(v|u > -x_1\alpha_o) = \frac{\int_{-\infty}^{\infty}\int_{-x_1\alpha_o}^{\infty} vf(v,u)dudv}{\int_{-x_1\alpha_o}^{\infty} f_u(t)dt}. \tag{2.4}$$

For the index formulation, the property that $f(v,u)$ is not a function of $x_1\alpha_o$ has not been imposed in estimation. Imposing this property of the model is the key for possible identification. Estimation of $f(v,u)$ with the observed samples of y_1 and y_2 should take into account the distribution of x since the density of (v,u) conditional on $y_1 > 0$ will depend on $x_1\alpha_o$. Let $h(z)$ be the marginal density of $z = x_1\alpha_o$. Equation (2.4) can be rewritten as

$$E(v|u > -x_1\alpha_o) = \frac{\int_{-\infty}^{\infty}\int_{x_1\alpha_o}^{\infty}\int_{-x_1\alpha_o}^{\infty} vf(v,u)h(z)dudzdv}{\int_{x_1\alpha_o}^{\infty}\int_{-x_1\alpha_o}^{\infty} f_u(t)h(z)dtdz} \tag{2.4'}$$

which can be estimated by some nonparametric functions as described below.

Let $\{(y_{1i}, y_{2i}, x_i) : y_{1i} > 0, i = 1, \cdots, n\}$ be the observed sample of (y_1, y_2, x) where $y_1 > 0$ [2]. Let $K(\cdot)$ be a kernel function on R^2 and $a_n > 0$ be a bandwidth parameter (Rao [1983]). For each (α, β) in the parameter space $\Theta_1 \times \Theta_2$, define

$$A_n(x_i,\alpha,\beta) = \int_{x_{1i}\alpha+\Delta_n}^{\infty}\int_{-x_{1i}\alpha}^{\infty} \frac{1}{(n-1)a_n^2}\sum_{j\neq i}^{n} v_j(\beta)K\left[\frac{u-u_j(\alpha)}{a_n}, \frac{z-x_{1j}\alpha}{a_n}\right]dudz \tag{2.5}$$

and

$$B_n(x_{1i},\alpha) = \int_{x_{1i}\alpha+\Delta_n}^{\infty}\int_{-x_{1i}\alpha}^{\infty} \frac{1}{(n-1)a_n^2}\sum_{j\neq i}^{n} K\left[\frac{u-u_j(\alpha)}{a_n}, \frac{z-x_{1j}\alpha}{a_n}\right]dudz \tag{2.6}$$

where $u_j(\alpha) = y_{1j} - x_{1j}\alpha$, $v_j(\beta) = y_{2j} - x_{2j}\beta$ and $\Delta_n > 0$ is a positive constant. As shown in the subsequent sections, at each point x_i, the ratio $\frac{A_n(x_i,\alpha,\beta)}{B_n(x_{1i},\alpha)}$ provides an estimate of the conditional expectation $E(y_2 - x_2\beta | y_1 - x_1\alpha > -x_{1i}\alpha, x_1\alpha > x_{1i}\alpha)$. Given a consistent estimate $\hat{\alpha}$ of α_o, our proposed estimation method is a semiparametric least squares procedure:

$$\min_{\beta} \frac{1}{n}\sum_{i=1}^{n} I_X(x_{1i})(y_{2i} - x_{2i}\beta - \frac{A_n(x_{1i},\hat{\alpha},\beta)}{B_n(x_{1i},\hat{\alpha})})^2 \tag{2.7}$$

[2] Our two stage estimation method does not use information of the event $y_1 < 0$ once a consistent estimate of α is given. Therefore the sample can simply consist of truncated observations.

where $I_X(x_{1i})$ is the indicator function of a set X constructed by trimming the regressors in x_1. The two stage estimator $\hat{\beta}$ from (2.7) has a closed form expression. Define a weight function:

$$W_n(x_{1i}, z_{1j}, \alpha) = \frac{\int_{x_{1i}\alpha+\Delta_n}^{\infty} \int_{-x_{1i}\alpha}^{\infty} K\left[\frac{u-u_j(\alpha)}{a_n}, \frac{z-x_{1j}\alpha}{a_n}\right] dudz}{\sum_{l\neq i}^{n} \int_{x_{1i}\alpha+\Delta_n}^{\infty} \int_{-x_{1i}\alpha}^{\infty} K\left[\frac{u-u_l(\alpha)}{a_n}, \frac{z-x_{1l}\alpha}{a_n}\right] dudz} \tag{2.8}$$

where $z_{1j} = (y_{1j},\ x_{1j})$. The two stage estimator $\hat{\beta}$ is

$$\begin{aligned}\hat{\beta} =& \{\sum_{i=1}^{n} I_X(x_{1i})[x_{2i} - \sum_{j\neq i}^{n} x_{2j} W_n(x_{1i}, z_{1j}, \hat{\alpha})]'[x_{2i} - \sum_{j\neq i}^{n} x_{2j} W_n(x_{1i}, z_{1j}, \hat{\alpha})]\}^{-1} \\ &\cdot \sum_{i=1}^{n} I_X(x_{1i})[x_{2i} - \sum_{j\neq i}^{n} x_{2j} W_n(x_{1i}, z_{1j}, \hat{\alpha})]'[y_{2i} - \sum_{j\neq i}^{n} y_{2j} W_n(x_{1i}, z_{1j}, \hat{\alpha})].\end{aligned} \tag{2.9}$$

This estimation procedure is similar to the two stage procedures in Robinson [1988] and Powell [1988]. However, the weight function $W_n(\cdot)$ is quite different from theirs. The estimator $\hat{\beta}$ can be shown to be consistent and asymptotically normally distributed under some regularity conditions if the banswidth parameter a_n and the trimming parameter Δ_n are chosen to converge to zero at certain rates as sample size n increases. These asymptotic properties can be derived by extending the analysis in Lee [1988] for the truncated regression model to this sample selection model.

To justify the statistical properties of our estimator, the following regularity conditions are assumed.

Assumption 1.

1. The disturbances u and v in the latent equations are independent with x.
2. The samples $(y_{1i}, y_{2i}, x_i), i = 1, \cdots, n$ where $y_{1i} > 0$ for all i are i.i.d.
3. The first four order moments of (y_{2i}, x_i) exist.
4. $\hat{\alpha}$ is a consistent estimate of α_o.
5. For each $\alpha \in \Theta_1$ where Θ_1 is a compact neighborhood of α_o, the index $x_1\alpha$ is a continuous random variable.
6. The set X is chosen to be a compact subset of the support S of x_1 such that $\max_{x_1\in X} x_1\alpha_o < \max_{x_1\in S} x_1\alpha_o$.
7. For each $x_{1i} \in X$, there exist with positive probability some x_1 in the set $\{x_1 | x_1\alpha_o > x_{1i}\alpha_o\}$ such that $P(y_1 > x_1\alpha_o - x_{1i}\alpha_o | x) > 0$.

Assumption 2.

1. For each $\alpha \in \Theta_1$, the density function $g(w|\alpha)$ of $(y_1 - x_1\alpha, x_1\alpha)$ conditional on $y_1 > 0$, and the conditional expectations $E(x|w,\alpha)$, $E(y_2|w,\alpha)$, $E(y_2x_1|w,\alpha)$ and $E(x_1x_2|w,\alpha)$ conditional on $(y_1 - x_1\alpha, x_1\alpha) = w$ are twice differentiable in w on its support $W = \{w | w_1 + w_2 > 0 \text{ where } w = (w_1, w_2)\}$.
2. The functions $g(w|\alpha)$, $E(x_2|w,\alpha)$ and $E(y_2|w,\alpha)$ are continuous in $\alpha \in \Theta_1$.

Assumption 3.

There exist Lebesque measurable functions $h_j(w)$, $j = 1, \cdots, 12$ with the following domination and integrability properties:

1. Domination properties: On some neighborhood $N_\delta(w)$ of w in W with radius $\delta > 0$ which does not depend on w,

(i) $\sup_{\alpha\in\Theta_1} \sup_{s\in N_\delta(w)} g(s|\alpha) \leq h_1(w)$.

(ii) $\sup_{\alpha\in\Theta_1} \sup_{s\in N_\delta(w)} \| E(x_2|s,\alpha)g(s|\alpha) \| \leq h_2(w)$.

(iii) $\sup_{\alpha\in\Theta_1} \sup_{s\in N_\delta(w)} \| E(y_2|s,\alpha)g(s|\alpha) \| \leq h_3(w)$.

(iv) $\sup_{\alpha\in\Theta_1} \sup_{s\in N_\delta(w)} \| \frac{\partial^2}{\partial s \partial s'} g(s|\alpha) \| \leq h_4(w)$.

(v) $\sup_{\alpha\in\Theta_1} \sup_{s\in N_\delta(w)} \| \frac{\partial^2}{\partial s \partial s'} [E(x_2|s,\alpha)g(s|\alpha)] \| \leq h_5(w)$.

(vi) $\sup_{\alpha\in\Theta_1} \sup_{s\in N_\delta(w)} \| \frac{\partial^2}{\partial s \partial s'} [E(y_2|s,\alpha)g(s|\alpha)] \| \leq h_6(w)$.

(vii) $\sup_{(\alpha,x_{1i})\in\Theta_1\times X} \sup_{s\in N_\delta(w)} E(\| x_{1i} - x_1 \|^2 |s,\alpha)g(s|\alpha) \leq h_7(w)$.

(viii) $\sup_{(\alpha,x_{1i})\in\Theta_1\times X} \sup_{s\in N_\delta(w)} E(y_2^2 \| x_{1i} - x_1 \|^2 |s,\alpha)g(s|\alpha) \leq h_8(w)$.

(ix) $\sup_{(\alpha,x_{1i})\in\Theta_1\times X} \sup_{s\in N_\delta(w)} E(\| (x_{1i} - x_1)'x_2 \|^2 |s,\alpha)g(s|\alpha) \leq h_9(w)$.

(x) $\sup_{\alpha\in\Theta_1} \sup_{s\in N_\delta(w)} \| \frac{\partial^2}{\partial s \partial s'} [E(x_1|s,\alpha)g(s|\alpha)] \| \leq h_{10}(w)$.

(xi) $\sup_{(\alpha,x_{1i})\in\Theta_1\times X} \sup_{s\in N_\delta(w)} \| \frac{\partial^2}{\partial s \partial s'} [E(y_2(x_{1i} - x_1)|s,\alpha)g(s|\alpha)] \| \leq h_{11}(w)$.

(xii) $\sup_{(\alpha,x_{1i})\in\Theta_1\times X} \sup_{s\in N_\delta(w)} \| \frac{\partial^2}{\partial s \partial s'} [E((x_{1i} - x_1)'x_2|s,\alpha)g(s|\alpha)] \| \leq h_{12}(w)$.

2. Integrability properties:

(i) $\int_{-\infty}^{\infty} \int_{-z}^{\infty} h_j(u,z)dudz < \infty$, for $j = 1, \cdots 6$.

(ii) $\int_{-\infty}^{\infty} \sup_u h_j(u,z)dz < \infty$ and $\int_{-\infty}^{\infty} \sup_z h_j(u,z)du < \infty$ for $j = 7, \cdots, 12$.

Assumption 4.

1. The kernel function $K(w)$ on R^2 is bounded and has a bounded support.
2. $\int K(w)dw = 1$ and $\int wK(w)dw = 0$.
3. The bandwidth sequence $\{a_n\}$ with $a_n > 0$, converges to zero at a rate such that $\lim_{n\to\infty} \frac{na_n^3}{\ln n} = \infty$ and $\lim_{n\to\infty} na_n^4 = 0$.
4. The $\{\Delta_n\}$ is a positive sequence such that $\lim_{n\to\infty} \Delta_n = 0$ but $\lim_{n\to\infty} \frac{\Delta_n}{a_n} = \infty$.

Assumption 5.(Identification condition)

The matrix $E(I_X(x_{1i})[x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o)]'[x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o)])$ is nonsingular.

The conditions in Assumption 1 are some basic regularity conditions for our model. The regressors in x_{1i} are trimmed in Assumption 1(6) to guarantee that for each $x_{1i} \in X$ the probability $P(x_1\alpha_o > x_{1i}|x_{1i})$ is strictly positive [3]. Assumption 1(7) together with X assumes that $P(u > -x_{1i}\alpha_o)$ is also strictly bounded away from zero on X. This assumption is always satisfied if u is unbounded from above. Assumptions 2-4 are used to quarantee convergence of the nonparametric functions in (2.5) and (2.6) and their derivatives to some proper limit functions. The conditions in Assumption 3 permit interchange of order for limiting operators and integration operators by Lebesque dominated convergence theorem (LDC). The rate of convergence of the bandwidth parameter a_n controls rate of convergence of the nonparametric functions. The trimming parameter Δ_n is used to avoid complication in evaluation of biases of the nonparametric functions along the boundary of $y_1 > 0$ so that proper uniform rate of asymptotic biases can be established. For our proposed estimation method, since the rate of convergence of the bandwidth parameter a_n is not too slow, there is no need to select high order kernel functions with zero second or higher order moments and hence the kernel function $K(\cdot)$ can simply be a density function which in general can simplify computation of the estimate. These conditions with the identification condition in Assumption 5 are sufficient to prove that our two stage estimator is $\sqrt{n}$-consistent and

[3] To simplify notation, if there were no confusion in the text, we would drop the conditional argument x_{1i} in probability or expectation functions in the sequel.

is asymptotically equivalent to the sum of an asymptotic normal variable and a variable involving the estimate $\hat{\alpha}$ from the first stage estimation. To complete the asymptotic distribution of the two stage estimator, we have to be specific about the asymptotic properties of $\hat{\alpha}$. With proper first stage estimator $\hat{\alpha}$, our two stage estimator is asymptotically normal.

3. Consistency and Identification

Asymptotic properties of the two stage estimator $\hat{\beta}$ depend on properties of the nonparametric functions in (2.5) and (2.6). Denote

$$C_n(x_{1i},\alpha) = \int_{x_{1i}\alpha+\Delta_n}^{\infty}\int_{-x_{1i}\alpha}^{\infty} \frac{1}{(n-1)a_n^2}\sum_{j\neq i}^{n} y_{2j} K\left[\frac{u-u_j(\alpha)}{a_n},\frac{z-x_{1j}\alpha}{a_n}\right] dudz \tag{3.1}$$

and

$$D_n(x_{i1},\alpha) = \int_{x_{1i}\alpha+\Delta_n}^{\infty}\int_{-x_{1i}\alpha}^{\infty} \frac{1}{(n-1)a_n^2}\sum_{j\neq i}^{n} x_{2j} K\left[\frac{u-u_j(\alpha)}{a_n},\frac{z-x_{1j}\alpha}{a_n}\right] dudz. \tag{3.2}$$

The equation (2.9) becomes

$$\begin{aligned}\hat{\beta} = &\left\{\sum_{i=1}^{n} I_X(x_{1i})\left[x_{2i}-\frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})}\right]'\left[x_{2i}-\frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})}\right]\right\}^{-1}\\ &\cdot\sum_{i=1}^{n} I_X(x_{1i})\left[x_{2i}-\frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})}\right]'\left(y_{2i}-\frac{C_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})}\right).\end{aligned} \tag{3.3}$$

By a change of variables, the nonparametric functions $B_n(x_{1i},\alpha)$, $C_n(x_{1i},\alpha)$ and $D_n(x_{1i},\alpha)$ can be rewritten as

$$B_n(x_{1i},\alpha) = \frac{1}{n-1}\sum_{j\neq i}^{n}\int_{\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}}^{\infty}\int_{\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n}}^{\infty} K(u,z)dudz, \tag{3.4}$$

$$C_n(x_{1i},\alpha) = \frac{1}{n-1}\sum_{j\neq i}^{n} y_{2j}\int_{\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}}^{\infty}\int_{\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n}}^{\infty} K(u,z)dudz \tag{3.5}$$

and

$$D_n(x_{1i},\alpha) = \frac{1}{n-1}\sum_{j\neq i}^{n} x_{2j}\int_{\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}}^{\infty}\int_{\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n}}^{\infty} K(u,z)dudz. \tag{3.6}$$

Since the second moments of y_2 and x_2 are finite and $\int_{-\infty}^{\infty}|K(w)|dw < \infty$, the variances of the nonparametric functions in (3.4)-(3.6) have the following order,

$$\sup_{x_1,\alpha} var(B_n(x_1,\alpha)|x_1) = O(\frac{1}{n}), \tag{3.7}$$

$$\sup_{x_1,\alpha} var(C_n(x_1,\alpha)|x_1) = O(\frac{1}{n}) \tag{3.8}$$

and

$$\sup_{x_1,\alpha} var(D_n(x_1,\alpha)|x_1) = O(\frac{1}{n}). \tag{3.9}$$

Under Assumptions 3 and 4, Proposition 2 of Appendix 1 implies that

$$\sup_{x_1,\alpha} |E(B_n(x_1,\alpha)|x_1) - B(x_1,\alpha,\Delta_n)| = O(a_n^2), \tag{3.10}$$

$$\sup_{x_1,\alpha} |E(C_n(x_1,\alpha)|x_1) - C(x_1,\alpha,\Delta_n)| = O(a_n^2) \tag{3.11}$$

and

$$\sup_{x_1,\alpha} |E(D_n(x_1,\alpha)|x_1) - D(x_1,\alpha,\Delta_n)| = O(a_n^2) \tag{3.12}$$

where

$$B(x_1,\alpha,\Delta) = \int_{x_1\alpha+\Delta}^{\infty}\int_{-x_1\alpha}^{\infty} g(u,z|\alpha)dudz, \tag{3.13}$$

$$C(x_1,\alpha,\Delta) = \int_{x_1\alpha+\Delta}^{\infty}\int_{-x_1\alpha}^{\infty} E(y_2|u,z,\alpha)g(u,z|\alpha)dudz, \tag{3.14}$$

$$D(x_1,\alpha,\Delta) = \int_{x_1\alpha+\Delta}^{\infty}\int_{-x_1\alpha}^{\infty} E(x_2|u,z,\alpha)g(u,z|\alpha)dudz \tag{3.15}$$

and $E(\cdot|u,z,\alpha)$ is conditional expectation conditional on $(u,z) = (y - x_1\alpha, x_1\alpha)$ [4]. With LDC theorem, Assumption 2(2) and Assumption 3 imply that the function $B(x_1,\alpha,\Delta)$, $C(x_1,\alpha,\Delta)$ and $D(x_1,\alpha,\Delta)$ are uniformly continuous on $X \times \Theta_1 \times [0,\ 1]$. The uniform law of large numbers in Proposition 1 of Appendix 1 can be applied to (3.4)-(3.6) with $d = \bar{d} = 0$. Since $\hat{\alpha}$ is consistent, it follows that

$$\mathrm{plim}_{n\to\infty} \sup_{x_1\in X} \left| B_n(x_1,\hat{\alpha}) - \int_{x_1\alpha_o}^{\infty}\int_{-x_1\alpha_o}^{\infty} g(u,z|\alpha_o)dudz \right| = 0, \tag{3.16}$$

$$\mathrm{plim}_{n\to\infty} \sup_{x_1\in X} \left| C_n(x_1,\hat{\alpha}) - \int_{x_1\alpha_o}^{\infty}\int_{-x_1\alpha_o}^{\infty} E(y_2|u,z,\alpha_o)g(u,z|\alpha_o)dudz \right| = 0 \tag{3.17}$$

and

$$\mathrm{plim}_{n\to\infty} \sup_{x_1\in X} \left\| D_n(x_1,\hat{\alpha}) - \int_{x_1\alpha_o}^{\infty}\int_{-x_1\alpha_o}^{\infty} E(x_2|u,z,\alpha_o)g(u,z|\alpha_o)dudz \right\| = 0. \tag{3.18}$$

On X, Assumption 1(7) guarantees that the probability $\int_{x_1\alpha_o}^{\infty}\int_{-x_1\alpha_o}^{\infty} g(u,z|\alpha_o)dudz$ is uniformly bounded away from zero. As u and v are independent with x, $E(v|u > -x_{1i}\alpha_o, x_1\alpha_o > x_{1i}\alpha_o) =$

[4] As in all cases $u + z > 0$ in this article, this conditional expectation is indifferent to conditional expectation with or without the additional condition $y_1 > 0$.

$E(v|u > -x_{1i}\alpha_o)$ and $E(x_2|u > -x_{1i}\alpha_o, x_1\alpha_o > x_{1i}\alpha_o) = E(x_2|x_1\alpha_o > x_{1i}\alpha_o)$ for all x_{1i}. Therefore,

$$\text{plim}_{n\to\infty} \sup_{x_{1i}\in X} \left| \frac{C_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})} - E(y_2|u > -x_{1i}\alpha_o, x_1\alpha_o > x_{1i}\alpha_o) \right| = 0 \tag{3.19}$$

and

$$\text{plim}_{n\to\infty} \sup_{x_{1i}\in X} \left| \frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o) \right| = 0. \tag{3.20}$$

Since $A_n(x_{1i},\alpha,\beta) = C_n(x_{1i},\alpha) - D_n(x_{1i},\alpha)\beta$, (3.19) and (3.20) imply that

$$\text{plim}_{n\to\infty} \sup_{x_{1i}\in X} \left| \frac{A_n(x_{1i},\hat{\alpha},\beta_o)}{B_n(x_{1i},\hat{\alpha})} - E(v|u > -x_{1i}\alpha_o) \right| = 0. \tag{3.21}$$

The equation (3.3) implies that

$$\begin{aligned} \hat{\beta} - \beta_o = & \left\{ \sum_{i=1}^n I_X(x_{1i}) \left[x_{2i} - \frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})} \right]' \left[x_{2i} - \frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})} \right] \right\}^{-1} \\ & \cdot \sum_{i=1}^n I_X(x_{1i}) \left[x_{2i} - \frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})} \right]' \left(y_{2i} - x_{2i}\beta_o - \frac{A_n(x_{1i},\hat{\alpha},\beta_o)}{B_n(x_{1i},\hat{\alpha})} \right). \end{aligned} \tag{3.22}$$

Since the first two moments of (y_2, x_2) are finite, (3.19)-(3.21) and Kolmogorov's law of large numbers for i.i.d. random variables imply that

$$\frac{1}{n} \sum_{i=1}^n I_X(x_{1i}) \left[x_{2i} - \frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})} \right]' \left[x_{2i} - \frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})} \right] \xrightarrow{p} A \tag{3.23}$$

where $A = E(I_X(x_{1i})[x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o)]'[x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o)])$, and

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n I_X(x_{1i}) \left[x_{2i} - \frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})} \right]' \left(y_{2i} - x_{2i}\beta_o - \frac{A_n(x_{1i},\hat{\alpha},\beta_o)}{B_n(x_{1i},\hat{\alpha})} \right) \\ & \xrightarrow{p} E\{I_X(x_{1i})[x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o)]'[y_{2i} - x_{2i}\beta_o - E(v|u > -x_{1i}\alpha_o)]\} \\ & = 0. \end{aligned} \tag{3.24}$$

The consistency of $\hat{\beta}$ follows from (3.22)-(3.24) and the identification condition in Assumption 5.

The identification condition requires essentially that the random variables in $x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o)$ with $x_{1i} \in X$ are not linearly dependent. For the special case that x_2 is independent with x_1, this condition will reduce to the requirement that the variance matrix of x_2 is nonsingular. For models with a single regressor and $x_1 = x_2 = x$, the condition is simply $E[I_X(x_i)(x_i - E(x|x\alpha_o > x_i\alpha_o))^2] > 0$ which holds as $x_i < E(x|x > x_i)$ and $x_i > E(x|x < x_i)$ for all $x_i \in X$.

4. Asymptotic Distribution

The asymptotic distribution of β_o can be derived from (3.22). Denote

$$L_n(\alpha,\beta_o) = \frac{1}{n}\sum_{i=1}^{n} I_X(x_{1i})\left[x_{2i} - \frac{D_n(x_{1i},\alpha)}{B_n(x_{1i},\alpha)}\right]'\left(y_{2i} - x_{2i}\beta_o - \frac{A_n(x_{1i},\alpha,\beta_o)}{B_n(x_{1i},\alpha)}\right). \tag{4.1}$$

By a Taylor expansion,

$$\sqrt{n}L_n(\hat{\alpha},\beta_o) = \sqrt{n}L_n(\alpha_o,\beta_o) + \frac{\partial L_n(\bar{\alpha},\beta_o)}{\partial\alpha'}\sqrt{n}(\hat{\alpha}-\alpha_o) \tag{4.2}$$

where $\bar{\alpha}$ lies between $\hat{\alpha}$ and α_o. The term $\frac{\partial L_n(\bar{\alpha},\beta_o)}{\partial\alpha'}$ depends on the first order derivatives of the nonparametric functions in (3.4)-(3.6). As shown in Appendix 2, with the rate of convergence for the bandwidth sequence $\{a_n\}$ in Assumption 4(3), $\frac{\partial A_n(x_1,\bar{\alpha},\beta_o)}{\partial\alpha}$, $\frac{\partial B_n(x_1,\bar{\alpha})}{\partial\alpha}$ and $\frac{\partial D_n(x_1,\bar{\alpha})}{\partial\alpha'}$ converge in probability uniformly in $x_1 \in X$ to some well defined limits and

$$\frac{\partial L_n(\bar{\alpha},\beta_o)}{\partial\alpha'} \xrightarrow{p} B \tag{4.3}$$

where

$$B = E(I_X(x_{1i})\tau(-x_{1i}\alpha_o)[x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o)]'[x_{1i} - E(x_1|x_1\alpha_o > x_{1i}\alpha_o)]) \tag{4.4}$$

and $\tau(t) = \frac{\partial}{\partial t}E(v|u > t,\alpha_o)$. The asymptotic distribution of $\sqrt{n}L_n(\alpha_o,\beta_o)$ can be analyzed with Propositions 5, 6 and 7 in Appendix 1. The details are in Appendix 2. It follows from Appendix 2 that

$$\sqrt{n}L_n(\alpha_o,\beta_o) \overset{D}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\Psi_1^{(1)}(z_i) + \Psi_1^{(2)}(z_i)) \tag{4.5}$$

where $\overset{D}{=}$ means that the statistics on both sides have the same limiting distribution,

$$\Psi_1^{(1)}(z_i) = I_X(x_{1i})(x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o))'\epsilon_{2i} \tag{4.6}$$

with $\epsilon_{2i} = y_{2i} - x_{2i}\beta_o - E(v|u > -x_{1i}\alpha_o)$, and

$$\begin{aligned}
&\Psi_1^{(2)}(z_i)\\
&= -E\Big\{E[I_X(x_{1j})(x_{2j} - E(x_2|x_1\alpha_o > x_{1j}\alpha_o))'|x_{1j}\alpha_o]\\
&\quad\cdot\frac{v_i - E(v|u > -x_{1j}\alpha_o)}{B(x_{1j},\alpha_o,0)}I(x_{1j}\alpha_o < x_{1i}\alpha_o)I(-x_{1j}\alpha_o < u_i)|x_i,y_i\Big\}\\
&= -\int_{-u_i}^{x_{1i}\alpha_o} E[I_X(x_{1j})(x_{2j} - E(x_2|x_1\alpha_o > x_{1j}\alpha_o))'|x_{1j}\alpha_o = z]\frac{v_i - E(v|u > -z)}{\int_z^\infty h(t)dt}h(z)dz.
\end{aligned} \tag{4.7}$$

The asymptotic distribution of $\sqrt{n}L_n(\hat{\alpha}, \beta_o)$ depends on the joint distribution of $\sqrt{n}L_n(\alpha_o, \beta_o)$ and $\sqrt{n}(\hat{\alpha} - \alpha_o)$. To complete the asymptotic distribution of the two stage estimator $\hat{\beta}$, one needs to be specific about the distribution of $\sqrt{n}(\hat{\alpha} - \alpha_o)$. As a specific example, consider the estimator $\hat{\alpha}$ derived from a semiparametric nonlinear least squares estimation procedure for the truncated regression model in Lee [1988]. Under the regularity conditions in that article, it was shown that

$$\begin{aligned}\sqrt{n}(\hat{\alpha} - \alpha_o) =& C^{-1}\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} I_X(x_{1i})(x_{1i} - E(x_1|x_1\alpha_o > x_{1i}\alpha_o + \Delta_n))\frac{\tau_1(x_{1i}\alpha_o)}{B(x_{1i}, \alpha_o, \Delta_n)} \\ &\cdot \int_{x_{1i}\alpha_o+\Delta_n}^{\infty}\int_{-x_{1i}\alpha_o}^{\infty}(y_{1i} - x_{1i}\alpha_o - u)\frac{1}{a_n^2}K\left(\frac{u - u_j}{a_n}, \frac{z - x_{1j}\alpha_o}{a_n}\right)dudz + o_p(1)\end{aligned} \tag{4.8}$$

where

$$\begin{aligned}\tau_1(z) &= \frac{\partial}{\partial z}\{z + E(u|u > -z, \alpha_o)\} \\ &= 1 - \frac{f_u(-z)}{\int_{-z}^{\infty} f_u(t)dt}(z + E(u|u > -z, \alpha_o))\end{aligned} \tag{4.9}$$

is the derivative of the regession function of y_1 conditional on x and $y_1 > 0$ with respect to the index $x_1\alpha_o$, and

$$C = E\left\{I_X(x_{1i})\tau_1^2(x_{1i}\alpha_o)[x_{1i} - E(x_1|x_1\alpha_o > x_{1i}\alpha_o)]'[x_{1i} - E(x_1|x_1\alpha_o > x_{1i}\alpha_o)]\right\}. \tag{4.10}$$

With this consistent estimate, Appendix 2 proves that the two stage estimator $\hat{\beta}$ is asymptotically normal:

$$\sqrt{n}(\hat{\beta} - \beta_o) \xrightarrow{D} N(0, \Omega) \tag{4.11}$$

where

$$\Omega = A^{-1}[I, \;\; BC^{-1}]\Sigma[I, \;\; BC^{-1}]'A^{-1}, \tag{4.12}$$

I is an identity matrix; A is the limit matrix in (3.23); B is the matrix in (4.4); C is defined in (4.10) and $\Sigma = E(\Psi(z_i)\Psi'(z_i))$ with

$$\begin{aligned}\Psi(z_i) =& I_X(x_{1i})\begin{pmatrix}(x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o))'\epsilon_{2i} \\ (x_{1i} - E(x_1|x_1\alpha_o > x_{1i}\alpha_o))'\tau_1(x_{1i}\alpha_o)\epsilon_{1i}\end{pmatrix} \\ &- \int_{-u_i}^{x_{1i}\alpha_o}\begin{pmatrix}E[I_X(x_{1j})(x_{2j} - E(x_2|x_1\alpha_o > x_{1j}\alpha_o))'|x_{1j}\alpha_o = z](v_i - E(v|u > -z)) \\ E[I_X(x_{1j})(x_{1j} - E(x_1|x_1\alpha_o > x_{1j}\alpha_o))'|x_{1j}\alpha_o = z](u_i - E(u|u > -z))\end{pmatrix}\frac{h(z)}{\int_z^{\infty} h(t)dt}dz.\end{aligned} \tag{4.13}$$

where $\epsilon_{1i} = y_{1i} - x_{1i}\alpha_o - E(u|u > -x_{1i}\alpha_o)$.

There are some interesting similarities between the asymptotic distribution of our two stage semiparametric estimate and the asymptotic distribution of a parametric two stage estimate of this model. If the functional form of $E(y_2 - x_2\beta|y_1 - x_1\alpha > -x_{1i}\alpha, x_1\alpha > x_{1i}\alpha)$ were known, a parametric two stage estimation method could be

$$\min_{\beta} \frac{1}{n}\sum_{i=1}^{n} I_X(x_{1i})(y_{2i} - x_{2i}\beta - E(y_2 - x_2\beta|u > -x_{1i}\hat{\alpha}, z > x_{1i}\hat{\alpha}))^2. \tag{4.14}$$

Let $\hat{\beta}_p$ be the two stage estimate of β from (4.14). Since $E(y_2 - x_2\beta|y_1 - x_1\alpha > -x_{1i}\alpha, x_1\alpha > x_{1i}\alpha) = \frac{C(x_{1i},\alpha)-D(x_{1i},\alpha)\beta}{B(x_{1i},\alpha)}$ where $B(x_{1i},\alpha) = B(x_{1i},\alpha,0)$, $C(x_{1i},\alpha) = C(x_{1i},\alpha,0)$ and $D(x_{1i},\alpha) = D(x_{1i},\alpha,0)$ defined in (3.13)-(3.15),

$$\begin{aligned}\hat{\beta}_p = &\left\{\sum_{i=1}^{n} I_X(x_{1i})\left[x_{2i} - \frac{D(x_{1i},\hat{\alpha})}{B(x_{1i},\hat{\alpha})}\right]'\left[x_{2i} - \frac{D(x_{1i},\hat{\alpha})}{B(x_{1i},\hat{\alpha})}\right]\right\}^{-1} \\ &\cdot\sum_{i=1}^{n} I_X(x_{1i})\left[x_{2i} - \frac{D(x_{1i},\hat{\alpha})}{B(x_{1i},\hat{\alpha})}\right]'\left(y_{2i} - \frac{C(x_{1i},\hat{\alpha})}{B(x_{1i},\hat{\alpha})}\right).\end{aligned} \tag{4.15}$$

which implies

$$\begin{aligned}\hat{\beta}_p - \beta_o = &\left\{\sum_{i=1}^{n} I_X(x_{1i})\left[x_{2i} - \frac{D(x_{1i},\hat{\alpha})}{B(x_{1i},\hat{\alpha})}\right]'\left[x_{2i} - \frac{D(x_{1i},\hat{\alpha})}{B(x_{1i},\hat{\alpha})}\right]\right\}^{-1} \\ &\cdot\sum_{i=1}^{n} I_X(x_{1i})\left[x_{2i} - \frac{D(x_{1i},\hat{\alpha})}{B(x_{1i},\hat{\alpha})}\right]'\left(y_{2i} - x_{2i}\beta_o - \frac{A(x_{1i},\hat{\alpha},\beta_o)}{B(x_{1i},\hat{\alpha})}\right)\end{aligned} \tag{4.16}$$

where $A(x_{1i},\alpha,\beta) = C(x_{1i},\alpha) - D(x_{1i},\alpha)\beta$. By a Taylor series expansion, it follows from (4.16) that

$$\sqrt{n}(\hat{\beta}_p - \beta_o) \overset{D}{=} A^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_{1i})[x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o)]'\epsilon + B\sqrt{n}(\hat{\alpha} - \alpha_o)\right\}. \tag{4.17}$$

For the semiparametric two stage estimator, (3.23), (4.2), (4.3) and (4.5) imply that

$$\sqrt{n}(\hat{\beta} - \beta_o) \overset{D}{=} A^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\Psi_1^{(1)}(z_i) + \Psi_1^{(2)}(z_i)) + B\sqrt{n}(\hat{\alpha} - \alpha_o)\right\}. \tag{4.18}$$

Comparing (4.17) with (4.18), the difference is that an extra term $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_1^{(2)}(z_i)$ appears in (4.18) for the semiparametric estimate. This extra term reflects the error introduced by replacing the conditional expectation $E(y_2 - x_2\beta|y_1 - x_1\alpha > -x_{1i}\alpha, x_1\alpha > x_{1i}\alpha)$ with the nonparametric estimate $\frac{A_n(x_{1i},\alpha,\beta)}{B_n(x_{1i},\alpha)}$ in the two stage estimation.

5. Covariance Estimation

The covaraince matrix of the limiting distribution of $\sqrt{n}(\hat{\beta}-\beta_o)$ is Ω in (4.12). From (3.23), the matrix A can be consistently estimated by

$$\frac{1}{n}\sum_{i=1}^{n} I_X(x_{1i})\left[x_{2i}-\frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})}\right]'\left[x_{2i}-\frac{D_n(x_{1i},\hat{\alpha})}{B_n(x_{1i},\hat{\alpha})}\right]. \tag{5.1}$$

Appendix 2 proves that $\frac{\partial L_n(\alpha,\beta)}{\partial\alpha'}$ converges in probability uniformly in (α,β) to a limit function which is continuous at (α_o,β_o). Hence, from (4.3), B can be consistently estimated by $\frac{\partial L_n(\hat{\alpha},\hat{\beta})}{\partial\alpha'}$. As suggested in Lee [1988], the matrix C can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n} I_X(x_{1i})\left[x_{1i}'+\frac{\partial E_{n,1}(x_{1i},\hat{\alpha})}{\partial\alpha}\right]\left[x_{1i}+\frac{\partial E_{n,1}(x_{1i},\hat{\alpha})}{\partial\alpha'}\right] \tag{5.2}$$

where

$$E_{n,1}(x_{1i},\hat{\alpha})=\frac{\sum_{j\neq i}^{n}\int_{x_{1i}\hat{\alpha}+\Delta_n}^{\infty}\int_{-x_{1i}\hat{\alpha}}^{\infty} uK\left(\frac{u-u_j(\hat{\alpha})}{a_n},\frac{z-x_{1j}\hat{\alpha}}{a_n}\right)dudz}{\sum_{j\neq i}^{n}\int_{x_{1i}\hat{\alpha}+\Delta_n}^{\infty}\int_{-x_{1i}\hat{\alpha}}^{\infty} K\left(\frac{u-u_j(\hat{\alpha})}{a_n},\frac{z-x_{1j}\hat{\alpha}}{a_n}\right)dudz} \tag{5.3}$$

is a nonparametric estimate of the conditional expectation $E(u|u>-x_i\alpha_o)$. It remains to consider the estimation of Σ. If the function $\Psi(z_i)$ in (4.13) could be evaluated, an estimate of Σ would be the sample covariance matrix of $\Psi(z_i)$. Eventhough $\Psi(z_i)$ can not be evaluated directly because the underlying distributions are unknown, it can be estimated by some sample functions. As in Appendix 2, $\Psi(z_i)$ is the limit function of $E(\Psi_n(z_i,z_j)|z_i)+E(\Psi_n(z_j,z_i)|z_i)$. This motivates the following estimate of Σ :

$$\hat{\Sigma}_n(\hat{\alpha},\hat{\beta})=\frac{1}{n^{(3)}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq i,j}^{n}[H_n(z_i,z_j,\hat{\alpha},\hat{\beta})+H_n(z_j,z_i,\hat{\alpha},\hat{\beta})][H_n(z_i,z_j,\hat{\alpha},\hat{\beta})+H_n(z_j,z_i,\hat{\alpha},\hat{\beta})]' \tag{5.4}$$

where $n^{(3)}=n(n-1)(n-2)$ and

$$H_n(z_i,z_j,\alpha,\beta)=(H_{n,1}'(z_i,z_j,\alpha,\beta)',\ H_{n,2}'(z_i,z_j,\alpha,\beta))', \tag{5.5}$$

$$\begin{aligned}H_{n,1}(z_i,z_j,\alpha,\beta)=I_X(x_{1i})&\left(x_{2i}-\frac{D_n(x_{1i},\alpha)}{B_n(x_{1i},\alpha)}\right)'\left\{\left(v_i(\beta)-\frac{A_n(x_{1i},\alpha,\beta)}{B_n(x_{1i},\alpha)}\right)\right.\\&\left.-\left(v_j(\beta)-\frac{A_n(x_{1j},\alpha,\beta)}{B_n(x_{1j},\alpha)}\right)\frac{1}{B_n(x_{1i},\alpha)}J_1(z_i,z_j,\alpha,\Delta_n)\right\},\end{aligned} \tag{5.6}$$

$$H_{n,2}(z_i, z_j, \alpha, \beta) = I_X(x_{1i}) \left(x_{1i}' + \frac{\partial E_n(x_{1i}, \alpha)}{\partial \alpha} \right) \frac{1}{B_n(x_{1i}, \alpha)} J_2(z_i, z_j, \alpha, \Delta_n) \tag{5.7}$$

with

$$J_1(z_i, z_j, \alpha, \Delta_n) = \int_{x_{1i}\alpha+\Delta_n}^{\infty} \int_{-x_{1i}\alpha}^{\infty} \frac{1}{a_n^2} K\left(\frac{u - u_j(\alpha)}{a_n}, \frac{z - x_{1j}\alpha}{a_n} \right) dudz \tag{5.8}$$

and

$$J_2(z_i, z_j, \alpha, \Delta_n) = \int_{x_{1i}\alpha+\Delta_n}^{\infty} \int_{-x_{1i}\alpha}^{\infty} (y_{1i} - x_{1i}\alpha - u) \frac{1}{a_n^2} K\left(\frac{u - u_j(\alpha)}{a_n}, \frac{z - x_{1j}\alpha}{a_n} \right) dudz. \tag{5.9}$$

Define

$$\begin{aligned} \Sigma_n(\alpha, \beta) = \frac{1}{n^{(3)}} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq i,j}^{n} &[H(z_i, z_j, \alpha, \beta, a_n, \Delta_n) + H(z_j, z_i, \alpha, \beta, a_n, \Delta_n)] \\ &\cdot [H(z_i, z_j, \alpha, \beta, a_n, \Delta_n) + H(z_j, z_i, \alpha, \beta, a_n, \Delta_n)]' \end{aligned} \tag{5.10}$$

where

$$H(z_i, z_j, \alpha, \beta, a_n, \Delta_n) = (H_1'(z_i, z_j, \alpha, \beta, a_n, \Delta_n),\ H_2'(z_i, z_j, \alpha, \beta, a_n, \Delta_n))', \tag{5.11}$$

$$\begin{aligned} H_1(z_i, z_j, \alpha, \beta, a_n, \Delta_n) = I_X(x_{1i}) &\left(x_{2i} - \frac{D(x_{1i}, \alpha, \Delta_n)}{B(x_{1i}, \alpha, \Delta_n)} \right)' \left\{ \left(y_{2i} - x_{2i}\beta - \frac{A(x_{1i}, \alpha, \beta, \Delta_n)}{B(x_{1i}, \alpha, \Delta_n)} \right) \right. \\ &\left. - \left(y_{2j} - x_{2j}\beta - \frac{A(x_{1j}, \alpha, \beta, \Delta_n)}{B(x_{1j}, \alpha, \Delta_n)} \right) \frac{1}{B(x_{1i}, \alpha, \Delta_n)} J_1(z_i, z_j, \alpha, \Delta_n) \right\}, \end{aligned} \tag{5.12}$$

and

$$H_2(z_i, z_j, \alpha, \beta, a_n, \Delta_n) = I_X(x_{1i}) \left(x_{1i}' + \frac{\partial E(x_{1i}, \alpha)}{\partial \alpha} \right) \frac{1}{B(x_{1i}, \alpha, \Delta_n)} J_2(z_i, z_j, \alpha, \Delta_n) \tag{5.13}$$

where $E(x_{1i}, \alpha)$ is the limit function of $E_{n,1}(x_{1i}, \alpha)$. The consistency of the covariance estimator $\hat{\Sigma}_n(\hat{\alpha}, \hat{\beta})$ will follow by showing that $\Sigma_n(\hat{\alpha}, \hat{\beta})$ converges in probability to Σ and $\hat{\Sigma}_n(\alpha, \beta) - \Sigma_n(\alpha, \beta)$ converges in probability to zero uniformly in (α, β) in a neighborhood of (α_o, β_o).

By a change of variables in the integral,

$$J_1(z_i, z_j, \alpha, \Delta_n) = \int_{\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}}^{\infty} \int_{\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n}}^{\infty} K(u, z) dudz \tag{5.14}$$

and

$$J_2(z_i, z_j, \alpha, \Delta_n) = \int_{\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}}^{\infty} \int_{\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n}}^{\infty} (u_i(\alpha) - u_j(\alpha) - a_n u) K(u, z) dudz. \tag{5.15}$$

The uniform law of large numbers for U statistics in Proposition 1 of Appendix 1 can be applied to $\Sigma_n(\alpha,\beta)$ in (5.10) with $d = \bar{d} = 0$ and $\delta = 4$. Under the assumption that the first eight moments of y_1, y_2 and x exist, as n goes to infinity

$$\Sigma_n(\alpha,\beta) - E(\Sigma_n(\alpha,\beta)) \xrightarrow{p} 0$$

uniformly in some compact neighborhood of (α_o,β_o). For any sequence (α_n,β_n) converging to (α_o,β_o), with similar arguments for the proofs of (A.2.26), (A.2.28), (A.2.31) and (A.2.32),

$$\lim_{n\to\infty} E[H(z_i,z_j,\alpha_n,\beta_n,a_n,\Delta_n) + H(z_i,z_j,\alpha_n,\beta_n,a_n,\Delta_n)|z_i] = \Psi(z_i)$$

where $\Psi(z_i)$ is in (4.13). Since $\hat{\alpha}$ and $\hat{\beta}$ are consistent, $\Sigma_n(\hat{\alpha},\hat{\beta}) \xrightarrow{p} \Sigma$. Uniform convergence of $\hat{\Sigma}_n(\alpha,\beta)-\Sigma_n(\alpha,\beta)$ to zero in probability is apparent as all the nonparametric functions $B_n(x_{1i},\alpha)$, etc., in $H_n(\cdot)$ have converged in probability uniformly in (x_{1i},α,β) to their limit functions in $H(\cdot)$.

6. Monte Carlo Simulation

In this section, we report some results on Monte Carlo simulation for finite sample performance of our estimator.

Simulated data are generated from the following latent equations:

$$y_1 = \alpha_1 s_1 + \alpha_2 s_2 + \sigma_1 \epsilon_1 \tag{6.1}$$

and

$$y_2 = \beta_1 s_1 + \beta_2 s_2 + \sigma_2 \epsilon_2. \tag{6.2}$$

The true parameter vectors are $(\alpha_1,\ \alpha_2) = (1,\ -1)$ and $(\beta_1, \beta_2) = (1,\ 1)$. The regressors s_1 and s_2 are randomly drawn from a normal $N(0,\ 1)$ distribution and a uniform $U(-2,\ 2)$ distribution respectively. s_1 and s_2 are independent. Different experiments are constructed by varying distribution of the disturbances ϵ_1 and ϵ_2. The disturbances ϵ_1 are generated from three different distributions, namely, the standard normal distribution $N(0,\ 1)$(Normal); a mixed gamma and normal distribution(Gamma*Normal):

$$\sqrt{0.8}Gamma(0,\ 1) + \sqrt{0.2}N(0,\ 1)$$

and a mixed negative gamma and normal distribution(-Gamma*Normal):

$$-\{\sqrt{0.8}Gamma(0,\ 1) + \sqrt{0.2}N(0,\ 1)\},$$

where $Gamma(0,\ 1)$ is a standarized gamma random variate with zero mean and unit variance of which the density function is

$$f_G(\epsilon_1) = \frac{8}{3}(\epsilon + 2)^3 \exp[-2(\epsilon + 2)] \qquad \epsilon > -2$$

with mode at $-\frac{1}{2}$. The disturbance ϵ_2 is correlated with ϵ_1 as

$$\epsilon_2 = \sqrt{0.25}\epsilon_1 + \sqrt{0.75}\eta$$

where η is a $N(0,\ 1)$ random variable independent with ϵ_1. The correlation coefficient of ϵ_1 and ϵ_2 is 0.5. The variances of ϵ_1 and ϵ_2 are both unity. However, variance of equation disturbance can be controlled by selecting values for scale parameter. The scale parameters σ_1 and σ_2 are set to 1.5 which implies that the R^2 values for both latent equations (6.1) and (6.2) are 0.5. The correlation coefficient of the two equations' distrubances remains 0.5. For each simulated data point, the sample (y_1, y_2, s_1, s_2) is kept only when $y_1 > 0$. The sample sizes that will be considered are 30, 50, 100 and 200. With these designs, as the latent variable y_1 has zero mean, the sample observations of y_1 are results of 50% trunction.

The kernel function used for our estimation is the following Biweight kernel density function:

$$K(t) = \begin{cases} \frac{15}{16}(1-t^2)^2, & \text{for } |t| < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{6.3}$$

This density has a bounded support and is continuously differentiable. Any bandwidth sequence of the form

$$a_n = c/n^p \tag{6.4}$$

with $\frac{1}{4} < p < \frac{1}{3}$ and c being a constant factor independent of sample size, will satisfy the rate requirement that $na_n^3 \to \infty$ and $na_n^4 \to 0$. For our experiments, $p = 0.3$ is set. However, we will experiment with different values for the constant factor c so as to investigate the sensitivity of our estimator to the chosen bandwith parameter. The parameter Δ_n is set to $0.1a_n^{0.9}$. For our semiparametric estimation, the regressors are trimmed whenever $|x_1| > 1.9$ or $|x_2| > 1.8$ so as to satisfy Assumption 1 which implies that approximately 15% of the data will be trimmed.

For each case, 300 repetitions of data set with the same sample size are generated. All the summary statistics reported below for each case are based on three hundred estimates. First stage estimate of the truncated regression function is derived by the semiparametric nonlinear least square method introduced in Lee [1988] [5]. For each case, both the first stage and second stage semiparametric estimation methods use the same kernel function in (6.3) and the same corresponding bandwidth factor a_n. Some of the simulated results of first stage estimate have been reported in Lee [1988]. In this paragraph, we will report the simulation results of second stage

[5] See also Appendix 3 of this article.

estimate of the outcome equation. For convenient reference, the corresponding first stage estimates of the truncated regression equation are provided in Appendix 3.

Table 1 reports simulation results of the two stage semiparametric estimation of the outcome equation (6.2) with various sample sizes and distributions. The bandwidth factor c in (6.4) is fixed at value 1. The true parameters for β_1 and β_2 are both 1. The summary statistics reported in the table are the mean value(Mean), the standard deviation(SD) and the root mean square error(RMSE) [6]. The biases of the estimates can be derived by comparing the reported mean values and the true parameters. There are some small sample biases. The biases are larger for the estimates of β_1 than the corresponding estimates of β_2. For sample size 30, the largest bias is about 0.164 which occurs when the distribution is mixed gamma-normal. The biases tend to decrease as the sample size increases. For sample size 200, the largest bias is about 0.006. As expected, variances decrease as the sample size increases. Comparing the variances and the root mean squared errors across different distributions, the estimation procedure performs best for the model with negative mixed gamma-normal distribution followed by the model with normal distribution. The same pattern applies to the first stage estimate of the decision equation in Table 1a of Appendix 3. The negative gamma-normal distribution is skew to the left before truncation. On the other hand, the gamma-normal distribution is skew to the right. As the disturbances of the two equations in our model are positively correlated, the sample selection mechanism implies that the left tails of the disturbances are truncated. The better performance for the model with negative gamma-normal distribution may be related to the fact that such distribution has thinner upper tail and smaller variance after selection than the other two cases. Comparing the estimates of the decision equation in Table 1a and the estimates of the outcome equation in Table 1, the estimates of the outcome equation are relatively more precise which may be due to the fact that the decision equation is subject to more severe degree of truncation than the outcome equation.

Table 2 reports simulation results derived with six different bandwidth factors. The bandwidth

[6] The standard deviation is derived as the square root of bias-adjusted sample variance. The variance component in RMSE is not bias-adjusted.

factor c varies from 0.10 to 4.00. This range seems to be quite wide. The sample size in this experiment is fixed at 100. It is interesting to see that the two stage estimates do not seem to be very sensitive to the different values of bandwidth factor. The biases are small. The standard deviations and the RMSE are similar in magnitude. To go beyond these summary statistics, for each sample we compute standard deviation of the six estimates based on different bandwidth factors, the average standard deviations of the estimates of β_1 and β_2 are respectively 0.067 and 0.062 for the normal distribution case, 0.075 and 0.079 for the gamma-normal case, and 0.053 and 0.046 for the negative gamma-normal case. The estimates derived with various bandwidth factors are in general resonably similar. In addition to the above summary statistics, we have computed also the average residual sum of squares(RSS) for each bandwidth factor. It is interesting to note that the RSS on average tends to decrease as the bandwidth factor becomes larger even though the estimates of β_1 and β_2 do not change much. In practice, one may report all these estimates or select one of them by some intuitive criteria. One possibility is to select the best fitted model in terms of RSS. Another possibility is to take average values of the estimates. The row marked 'min' reports the performance of the estimates derived from the best fitted criterion. The row marked 'ave' reports the performance of the averaged estimates. These results are quite encouraging. The biases are resonably small. The variances and RMSE are even slightly less than most the variances and RMSE of the estimates based on fixed bandwidth factor. These two strategies seem good for our estimation procedure [7].

In Table 3, we compare our semiparametric two stage estimates with ordinary least square estimates(OLS) and some parametric estimates of the outcome equation (6.2). The OLS procedure ignores the sample selection bias and is known to be inconsistent. The OLS estimates of (6.2) are reported on the first colum block in Table 3. The OLS estimates of β_1 are biased downward and

[7] We would like to point out that we have not experimented with bandwidth factor values greater than 4.00. The reason is that eventhough the second stage estimates of the outcome equation are not quite sensitive to the values of bandwidth factors, the first stage estimates of the decision equation in Table 2a of Appendix 3 are relatively more sensitive to such factors. Both the SD and RMSE of the first stage estimates increase as value of bandwidth factor increases. Values of bandwidth factor around 0.5 to 2.0 seem to be the best ones for the first stage estimates.

the OLS estimates of β_2 are biased upward for all the sample sizes and the distributions that we considered. On average, the biases are about 23%, 30% and 20% respectively for the normal, gamma-normal and negative gamma-normal models. The biases persist as sample size increases. Comparing the OLS with our semiparametric two stage estimates in the last column block in Table 3, the OLS estimates have smaller SD for all the cases. However, for sample sizes 100 and 200, the biases dominate the SD which results in larger RMSE than the semiparametric estimates. In Table 3, we report also some parametric two stage estimates based on normal distributional formulation. The first stage estimate of α in (6.1) is derived from the following nonlinear least squares procedure:

$$\min_{\alpha,\sigma_1}\sum_{i=1}^{n}(y_{1i} - s\alpha - \sigma_1\frac{\phi(\frac{s\alpha}{\sigma_1})}{\Phi(\frac{s\alpha}{\sigma_1})})^2. \tag{6.5}$$

With the first stage estimates $\hat{\alpha}$ and $\hat{\sigma_1}$ from (6.5), the second stage estimate of β is derived from

$$\min_{\beta,\sigma_{12}}\sum_{i=1}^{n}(y_{2i} - s\beta - \sigma_{12}\frac{\phi(\frac{s\hat{\alpha}}{\hat{\sigma_1}})}{\Phi(\frac{s\hat{\alpha}}{\hat{\sigma_1}})})^2. \tag{6.6}$$

This parametric two stage procedure is similar to Heckman's procedure (Heckman [1976]). The parametric estimates of β are reported in Table 3 under the column block marked 'PN-OLS'. The estimates reported under the colum block marked 'PN-ROLS' differ from the estimates 'PN-OLS' in that the intercept term of (6.2) is assumed known and restricted to be zero. The parametric first stage estimates can be found in Table 3a of Appendix 3 [8]. For the normal distribution model, this parametric two stage procedure provides consistent estimates. For the mixed gamma-normal and negative mixed gamma-normal distributions, this procedure is in general inconsistent. The PN-OLS estimates do not perform well for our simulated models [9]. The variances and RMSE for these estimates are larger than our semiparametric estimates for all cases. The restricted PN-ROLS estimates perform much better. The PN-OLS apparently suffer from the problem of multicollinearity. By restricting the intercept term to be zero, this exclusion restriction reduces the

[8] In this first stage estimation, the intercept term is assumed known and restricted to zero.

[9] In Table 3, we have reported only estimates of the regression coefficients. The estimates of the intercept term and the coefficient of the sample selection adjustment term are even worse. Their variances are three or four times larger than the variances of the regression coefficient estimates.

severity of multicollinearity for the parametric two stage estimation. The PN-ROLS estimates have smaller variances than the semiparametric estimates in all cases. There are some evidences that for the misspecified distributions, the parametric two stage estimates are biased [10]. The biases of the estimates of β_2 of the mixed distributions models are larger than the biases of the semiparametric estimates for all the sample sizes considered. For sample size 200, the biases of the estimates of β_1 are also larger than the biases of the semiparametric estimates. These biases are however not very larger and hence in terms of RMSE the parametric PN-ROLS estimates still perform better for all the cases considered.

As we have pointed out before, sample selection models are index models and can be estimated by the semiparametric methods introduced in Ichimura [1987] and Ichimura and Lee [1988]. The equations (6.1) and (6.2) in the above simulation are however not identifiable in the index formulation. Identification in index models will require appropriate restrictions on the outcome equation (6.2). For the sake of comparison without simulating another different data set, the true parameter β_2 in (6.2) is assumed to be known. Imposing the restriction that $\beta_2 = 1$, the parameter β_1 can be identified with the index formulation. The semiparametric two stage estimator of β_1 in the index model is

$$\begin{aligned}\tilde{\beta}_1 =&\{\sum_{i=1}^{n} I_X(s_i)[s_{1i} - \sum_{j\neq i}^{n} s_{1j}W_n^*(s_i, s_j, \hat{\alpha})]'[s_{1i} - \sum_{j\neq i}^{n} s_{1j}W_n^*(s_i, s_j, \hat{\alpha})]\}^{-1} \\ &\cdot \sum_{i=1}^{n} I_X(s_i)[s_{1i} - \sum_{j\neq i}^{n} s_{1j}W_n^*(s_i, s_j, \hat{\alpha})]'[y_{2i}^* - \sum_{j\neq i}^{n} y_{2j}^*W_n^*(s_i, s_j, \hat{\alpha})].\end{aligned} \tag{6.7}$$

where $y_2^* = y_2 - z_2$; the weight function $W_n^*(s_i, s_j, \alpha)$ is

$$W_n^*(s_i, s_j, \alpha) = \frac{K_4^*\left(\frac{s_i\alpha - s_j\alpha}{b_n}\right)}{\sum_{l\neq i}^{n} K_4^*\left(\frac{s_i\alpha - s_l\alpha}{b_n}\right)} \tag{6.8}$$

and the kernel function $K_4^*(\cdot)$ is

$$K_4^*(t) = 2K^*(t) - \frac{1}{\sqrt{2}}K^*(\frac{t}{\sqrt{2}}) \tag{6.9}$$

[10] The estimates of the coefficient of the sample selection bias adjustment term are also biased. While the true coefficient implied by our data generating processes is 0.75, the estimates of the mixed gamma-normal model and the negative mixed gamma-normal model are about 0.83 and 0.675.

with

$$K^*(t) = \begin{cases} \frac{35}{32}(1-t^2)^3, & \text{for } |t| < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{6.10}$$

and

$$b_n = \frac{c}{n^{1/5.5}}. \tag{6.11}$$

The function $K^*(t)$ is a proper density function which is twice continuously differentiable with bounded third order derivative. The kernel function K_4^* is a high order kernel with its first three moments being zero. The way of constructing such high order kernel function from density function is suggested in Bierens [1987]. This kernel function and its bandwidth rate (6.11) satisfy the regularity conditions in Ichimura and Lee [1988]. The second stage estimates based on the index formulation are presented in Table 4. Various values of the bandwidth factor c in (6.11) have also be tried. Imposing the restriction that $\beta_2 = 1$, the semiparametric estimates of β_1 from (2.9) are reported in the last row in Table 4. The simulations are based on sample size 200. The first stage estimates used are the corresponding ones on the bottom rows of Table 1a of Appendix 3. The semiparametric estimates based on the index formulation are relatively sensitive to the bandwidth factors. The magnitudes of the estimates as well as the variances decrease as the values of the bandwidth factor c in (6.11) increase from 1.0 to 20.0. The biases are the smallest around factor values of 5.0 or 7.5. The corresponding RMSE decease but eventually increase as bandwidth factor increases. Comparing these estimates with the semiparametric estimates from (2.9), the latter estimates have in general smaller biases and small RMSE. Taking into account the independence restriction, our proposed semiparametric estimates seem likely to be more efficient than the estimates which has utilized only the index resetriction [11].

[11] The asymptotic covariances of these two different two stage estimators are however not analytically comparable. Both estimators are consistent but not efficient as only some limited information in the sample have been utilized.

Table 1.

Results of 300 Repetitions with Various Sample Sizes

Design: Bandwidth factor c=1

	Normal	Gamma*Normal	-Gamma*Normal
N	Mean SD RMSE	Mean SD RMSE	Mean SD RMSE
30 β_1	.843 .746 .750	.834 .576 .590	.942 .544 .538
β_2	.938 .586 .579	1.037 .539 .531	.979 .548 .539
50 β_1	.919 .422 .426	.906 .516 .519	.928 .381 .384
β_2	.986 .427 .423	.996 .451 .446	.956 .358 .357
100 β_1	.992 .269 .268	.948 .294 .297	.970 .223 .224
β_2	.986 .248 .247	1.017 .275 .274	.983 .222 .222
200 β_1	1.005 .168 .168	.997 .224 .223	.994 .156 .156
β_2	.995 .159 .159	.996 .203 .203	1.004 .139 .139

Table 2.

Results with Various Banwidth Factors

Design: Sample Size 100

	Normal	Gamma*Normal	-Gamma*Normal
Factor	Mean SD RMSE RSS	Mean SD RMSE RSS	Mean SD RMSE RSS
.10 β_1	.973 .282 .282 173.865	.945 .296 .300 181.635	.959 .254 .256 170.393
β_2	.990 .231 .230	1.035 .256 .257	.974 .219 .219
.25 β_1	.989 .268 .267 175.482	.958 .293 .295 180.716	.955 .295 .297 174.063
β_2	.975 .285 .285	1.014 .279 .278	.974 .265 .265
.50 β_1	.997 .271 .270 172.786	.952 .283 .286 179.962	.956 .263 .265 169.427
β_2	.971 .273 .273	1.019 .261 .260	.988 .221 .220
1.00 β_1	.992 .269 .268 171.084	.948 .294 .297 178.625	.970 .223 .224 167.555
β_2	.986 .248 .247	1.017 .275 .274	.983 .222 .222
2.00 β_1	.994 .253 .252 168.173	.974 .300 .300 176.819	.978 .216 .216 164.056
β_2	.985 .254 .253	1.004 .279 .278	.993 .193 .192
4.00 β_1	1.023 .266 .266 167.573	1.010 .308 .307 176.423	.994 .217 .216 163.332
β_2	.959 .267 .269	.972 .300 .300	.978 .203 .203
min β_1	1.000 .242 .241 166.206	.965 .290 .291 174.743	.986 .213 .212 162.485
β_2	.991 .241 .240	1.019 .245 .245	.994 .193 .192
ave β_1	.995 .239 .238	.965 .272 .273	.969 .214 .215
β_2	.978 .232 .232	1.010 .249 .248	.982 .199 .199

Table 3.

OLS, Parametric Normal Two Stage and Semiparametric Estimates

N		OLS Mean	OLS SD	OLS RMSE	PN-OLS Mean	PN-OLS SD	PN-OLS RMSE	PN-ROLS Mean	PN-ROLS SD	PN-ROLS RMSE	Semiparametric E. Mean	Semiparametric E. SD	Semiparametric E. RMSE
					Normal								
30	β_1	.791	.317	.375	1.050	1.238	1.218	1.001	.328	.322	.843	.746	.750
	β_2	1.237	.261	.349	.991	1.342	1.319	1.009	.277	.272	.938	.586	.579
50	β_1	.787	.250	.327	.967	.826	.818	1.006	.262	.259	.919	.422	.426
	β_2	1.242	.195	.310	1.049	.839	.832	1.005	.214	.212	.986	.427	.423
100	β_1	.788	.163	.267	1.033	.488	.487	1.014	.183	.183	.992	.269	.268
	β_2	1.242	.137	.278	.984	.456	.454	1.001	.149	.148	.986	.248	.247
200	β_1	.779	.111	.247	1.002	.337	.336	1.009	.121	.121	1.005	.168	.168
	β_2	1.251	.098	.269	1.012	.327	.326	1.007	.102	.102	.995	.159	.159
					Gamma*Normal								
30	β_1	.697	.329	.443	1.231	3.562	3.510	.900	.390	.396	.834	.576	.590
	β_2	1.316	.272	.414	1.362	2.814	2.790	1.108	.306	.320	1.037	.539	.531
50	β_1	.706	.229	.371	.963	1.758	1.741	.934	.256	.262	.906	.516	.519
	β_2	1.322	.202	.379	1.047	1.267	1.225	1.092	.220	.236	.996	.451	.446
100	β_1	.720	.158	.321	1.024	.543	.541	.953	.179	.184	.948	.294	.297
	β_2	1.312	.141	.342	1.009	.582	.579	1.079	.156	.174	1.017	.275	.274
200	β_1	.724	.110	.297	1.033	.347	.348	.958	.124	.131	.997	.224	.223
	β_2	1.302	.098	.317	.980	.365	.365	1.063	.111	.127	.996	.203	.203
					-Gamma*Normal								
30	β_1	.796	.301	.359	.988	1.004	.987	1.017	.310	.305	.942	.544	.538
	β_2	1.210	.286	.351	1.008	1.040	1.023	.964	.279	.277	.979	.548	.539
50	β_1	.810	.223	.291	.933	.827	.821	1.026	.227	.226	.928	.381	.384
	β_2	1.189	.206	.278	1.043	.746	.740	.947	.196	.201	.956	.358	.357
100	β_1	.800	.158	.254	.965	.421	.420	1.020	.154	.155	.970	.223	.224
	β_2	1.199	.138	.242	1.011	.433	.431	.954	.136	.143	.983	.222	.222
200	β_1	.807	.118	.226	.969	.257	.258	1.027	.113	.116	.994	.156	.156
	β_2	1.203	.102	.227	1.021	.282	.282	.959	.097	.105	1.004	.139	.139

Table 4.

Index Model and Semiparametric Estimation

Design: sample size 200

	Normal				Gamma*Normal				-Gamma*Normal			
Factor	Mean	SD	RMSE	RSS	Mean	SD	RMSE	RSS	Mean	SD	RMSE	RSS
					Index model							
1.0 β_1	1.028	.451	.451	386.789	1.123	1.499	1.500	578.010	1.033	.437	.437	372.099
2.5 β_1	1.035	.272	.274	355.892	1.031	.348	.349	364.874	1.004	.235	.234	339.289
5.0 β_1	1.002	.330	.329	345.618	1.012	.348	.347	358.918	.949	.486	.487	351.707
7.5 β_1	.983	.249	.249	340.055	.985	.312	.312	356.015	.985	.258	.258	333.816
10.0 β_1	.962	.199	.202	338.433	.937	.272	.279	356.600	.960	.194	.198	327.917
15.0 β_1	.899	.168	.196	340.395	.874	.196	.233	358.676	.912	.157	.180	328.812
20.0 β_1	.859	.152	.207	341.973	.827	.156	.233	360.451	.881	.150	.191	329.832
min. β_1	1.005	.236	.235	335.174	1.013	.335	.334	351.045	.984	.192	.192	325.182
					Tobit selection model							
1.0 β_1	1.001	.168	.168	340.513	.986	.214	.214	357.208	.993	.153	.153	329.468

Appendix 1:

In this appendix, several propositions that are useful for our analysis are collected here for convenient reference. The proofs of these propositions have been established in our previous works.

Proposition 1. (A Uniform Law of Large Numbers) *Let $\{y_i\}$ be a sequence of i.i.d. random vectors and $y_{i_1}, \cdots, y_{i_l}$ be l distinct observations. Suppose that the measurable function $g(y_{i_1}, \cdots, y_{i_l}, a_n, \alpha)$ can be represented in the form,*

$$g(y_{i_1}, \cdots, y_{i_l}, a_n, \alpha) = \frac{1}{a_n^d} t(y_{i_1}, \cdots, y_{i_l}) h \left[y_{i_1}, \cdots, y_{i_l}, \frac{s(y_{i_1}, \cdots, y_{i_l}, \alpha)}{a_n} \right]$$

where $a_n = O(\frac{1}{n^p})$, $p > 0$, $d \geq 0$, $\alpha \in B$ and $s(y_{i_1}, \cdots, y_{i_l}, \alpha)$ is a finite dimensional vector value function, and the following conditions are satisfied:

1. *B is a compact subset of a finite dimensional Euclidean space.*
2. *The function $t(y_1, \cdots, y_l)$ is bounded by a finite order (say, order δ) polynomial of $y_1, \cdots, y_l$.*
3. *The first $\delta \cdot r$ moments of y exist, where $r \geq 2$.*
4. *The function $h(\cdot)$ is a bounded function.*
5. *$E\left\{ t^2(y_1, \cdots, y_l) h^2 \left[y_1, \cdots, y_l, \frac{s(y_1, \cdots, y_l, \alpha)}{a_n} \right] \right\} = O(a_n^{\bar{d}})$ uniformly in $\alpha \in B$, where $\bar{d} \leq d$.*
6. *The functions $h(y_1, \cdots, y_l, a_n, s)$ and $s(y_1, \cdots, y_l, \alpha)$ satisfy the bounded Lipschizian condition of order 1 with respect to α and s, uniformly in $y_1, \cdots, y_l$.*

   If $\lim_{n \to \infty} \frac{n}{\ln n} a_n^{2(1+\frac{\delta}{r})d - \bar{d}} = \infty$, then

$$\frac{1}{n^{(l)}} \sum_{n,l} \{ g(y_{i_1}, \cdots, y_{i_l}, a_n, \alpha) - E(g(y_1, \cdots, y_l, a_n, \alpha)) \} \xrightarrow{p} 0,$$

uniformly in $\alpha \in B$, where $n^{(l)} = n(n-1) \cdots (n-l+1)$ and the sum $\sum_{n,l}$ is taken over all l-tuples of distinct integers not exceeding n.

Furthermore, in addition to the above conditions, if

7. *$E(g(y_{i_1}, \cdots, y_{i_l}, a_n, \alpha))$ converges to a limit function $g^*(\alpha)$ uniformly in $\alpha \in B$, then*

$$\frac{1}{n^{(l)}} \sum_{n,l} g(y_{i_1}, \cdots, y_{i_l}, a_n, \alpha) - g^*(\alpha) \xrightarrow{p} 0,$$

uniformly in $\alpha \in B$.

Proof: This is a uniform law of large numbers with kernel function which generalizes a uniform law in Ichimura [1987] to cover unbounded random variables and U-statistics. The proof of this law can be found in Ichimura and Lee [1988].

Proposition 2. *Let $K(w)$ be a kernel function with zero mean and a bounded support. Let $g(w|\alpha)$ denote the density function of $(u(\alpha), x\alpha)$ with support $W = \{w|w_1 + w_2 > 0, \text{where } w = (w_1, w_2)\}$.*

Suppose that there exists a measurable function $h(v)$ such that

$$\sup_{\alpha} \sup_{w \in N_\delta(v)} \| \frac{\partial^2}{\partial w \partial w'} [E(c(y,x)|(u(\alpha), x\alpha) = w) g(w|\alpha)] \| \leq h(v)$$

for some neighborhood $N_\delta(v)$ of v in W with radius $\delta > 0$ (δ is independent with v), and

$$\int_{-\infty}^{\infty} \int_{-z}^{\infty} h(u,z) du dz < \infty,$$

then

$$\sup_{(x_i,\alpha)} | \int_{x_i\alpha + \Delta_n}^{\infty} \int_{-x_i\alpha}^{\infty} E\Big[c(y,x) \frac{1}{a_n^2} K\left(\frac{u - u(\alpha)}{a_n}, \frac{z - x\alpha}{a_n}\right)\Big] du dz$$
$$- \int_{x_i\alpha + \Delta_n}^{\infty} \int_{-x_i\alpha}^{\infty} E(c(y,x)|u(\alpha) = u, x\alpha = z) g(u,z|\alpha) du dz| = O(a_n^2),$$

where $\{\Delta_n\}$ is a sequence of positive numbers such that $\lim_{n\to\infty} \Delta_n = 0$ and $\lim_{n\to\infty} \frac{\Delta_n}{a_n} = \infty$.

Proof: This is Lemma 1 in Lee [1988].

Proposition 3. *Let $K(w)$ be a kernel function with zero mean and a bounded support, and $g(w|\alpha)$ denote the density function of $(u(\alpha), x\alpha)$ as in Proposition 2. Suppose that there exists a Lebesque measurable function $h(v)$ such that*

$$\sup_{\alpha} \sup_{w \in N_\delta(v)} \| \frac{\partial^2}{\partial w \partial w'} [E(c(y,x)|(u(\alpha), x\alpha) = w) g(w|\alpha)] \| \leq h(v)$$

for some neighborhood $N_\delta(v)$ of v in W with radius $\delta > 0$ (δ is independent with v).

(i) If $\int_{-\infty}^{\infty} \sup_z h(u,z) du < \infty$, then

$$\sup_{(x_i,\alpha)} | \int_{-x_{1i}\alpha}^{\infty} E\Big[c(y,x) \frac{1}{a_n^2} K\left(\frac{u - u(\alpha)}{a_n}, \frac{x_{1i}\alpha + \Delta_n - x\alpha}{a_n}\right)\Big] du$$
$$- \int_{-x_{1i}\alpha}^{\infty} E\Big[c(y,x)|u(\alpha) = u, x\alpha = x_{1i}\alpha + \Delta_n\Big] g(u, x_{1i}\alpha + \Delta_n|\alpha) du| = O(a_n^2).$$

(ii) If $\int_{-\infty}^{\infty} \sup_u h(u,z)dz < \infty$, *then*

$$\sup_{(x_i,\alpha)} \Big| \int_{x_{1i}\alpha+\Delta_n}^{\infty} E\Big[c(y,x)\frac{1}{a_n^2}K\left(\frac{-x_{1i}\alpha - u(\alpha)}{a_n}, \frac{z - x\alpha}{a_n}\right)\Big]dz$$
$$- \int_{x_{1i}\alpha+\Delta_n}^{\infty} E\Big[c(y,x)|u(\alpha) = -x_{1i}\alpha, x\alpha = z\Big]g(-x_{1i}\alpha, z|\alpha)dz\Big| = 0(a_n^2).$$

Proof: This is Lemma 2 in Lee [1988].

Proposition 4. *Let $F(w)$ be a bounded function with a bounded support and $g(w|\alpha)$ denote the density function of $(u(\alpha), x\alpha)$ as in Proposition 2. Suppose that there exists a Lebesque measurable function $h(v)$ such that*

$$\sup_{\alpha} \sup_{w \in N_\delta(v)} E[|c(y,x)|\Big|(u(\alpha), x\alpha) = w]g(w|\alpha) \le h(v)$$

for some neighborhood $N_\delta(v)$ of v in W with radius $\delta > 0$ (δ is independent with v).

(i) If $\int_{-\infty}^{\infty} \sup_u h(u,z)dz < \infty$, *then*

$$\sup_{x_i,\alpha} E\big(|c(y,x)|\{\int_{\frac{x_{1i}\alpha+\Delta_n - x\alpha}{a_n}}^{\infty} F[\frac{-x_{1i}\alpha - u(\alpha)}{a_n}, z]dz\}^2\big) = O(a_n).$$

(ii) If $\int_{-\infty}^{\infty} \sup_z h(u,z)du < \infty$, *then*

$$\sup_{x_i,\alpha} E\big(|c(y,x)|\{\int_{\frac{-x_{1i}\alpha - u(\alpha)}{a_n}}^{\infty} F[u, \frac{x_{1i}\alpha + \Delta_n - x\alpha}{a_n}]du\}^2\big) = O(a_n).$$

Proof: This proposition is Lemma 3 in Lee [1988].

Proposition 5. *Let $f_n(z_1, \cdots, z_{i-1}, z_{i+1}, \cdots, z_n; z_i)$ and $g_n(z_1, \cdots, z_{i-1}, z_{i+1}, \cdots, z_n; z_i)$ be two sequences of random functions of an i.i.d. sample $\{z_i\}$ and $h(z_i)$, $\bar{f}_n(z_i)$ and $\bar{g}_n(z_i)$ are measuralbe functions of z_i. Suppose that*

1. $E|h(z)| < \infty$
2. $\sup_{z_i} |E(f_n(z_1, \cdots, z_n; z_i)|z_i) - \bar{f}_n(z_i)| = O(a_n^{s_1})$,
3. $\sup_{z_i} |E(g_n(z_1, \cdots, z_n; z_i)|z_i) - \bar{g}_n(z_i)| = O(a_n^{s_2})$,
4. $\sup_{z_i} var(f_n(z_1, \cdots, z_n; z_i)|z_i) = O(\frac{1}{na_n^{r_1}})$, *and*
5. $\sup_{z_i} var(g_n(z_1, \cdots, z_n; z_i)|z_i)| = O(\frac{1}{na_n^{r_2}})$.

If $2s_1 > r_2 \geq 0$, $2s_2 > r_1 \geq 0$, $\lim_{n\to\infty} na_n^{r_1+r_2} = \infty$ *and* $\lim_{n\to\infty} na_n^{2(s_1+s_2)} = 0$, *then*

$$\text{plim}_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|h(z_i)|\cdot|f_n(z_1,\cdots,z_n;z_i)-\bar{f}_n(z_i)|\cdot|g_n(z_1,\cdots,z_n;z_i)-\bar{g}_n(z_i)| = 0.$$

Proof: This is Lemma 6 in Lee [1988] with slight generalization. The result follows from Markov inequality and Cauchy inequality.

Proposition 6. *Let z_i be a sequence of i.i.d. random variable and $\Phi_n(z_1, z_2, a_n)$ be a sequence of measurable functions with bandwidth sequence a_n where $a_n > 0$. Suppose that*

(1) $E(\Phi_n(z_1,z_2,a_n)) = O(a_n^s)$ *and* $var(\Phi_n(z_1,z_2,a_n)) = O(\frac{1}{a_n^r})$,

(2) there exists square integrable functions $h_j(z), j = 1,2$ such that $|E(\Phi_n(z_1,z_2,a_n)|z_1)| \leq h_1(z_1)$, $|E(\Phi_n(z_2,z_1,a_n)|z_1)| \leq h_2(z_1)$, *and*

(3) $\lim_{n\to\infty} E(\Phi_n(z_1,z_2,a_n)|z_j) = 0$, *a.e.*, $j = 1,2$.

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

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

Proof: This is Proposition 6 in Lee [1989].

Proposition 7. *Let z_i be a sequence of i.i.d. random variable and $\Phi_n(z_1, z_2, a_n)$ be a sequence of measurable functions with bandwidth sequence a_n where $a_n > 0$. Suppose that*

(1) $E(\Phi_n(z_1,z_2,a_n)) = O(a_n^s)$ *and* $var(\Phi_n(z_1,z_2,a_n)) = O(\frac{1}{a_n^r})$,

(2) $\lim_{n\to\infty} E(\Phi_n(z_1,z_2,a_n)|z_1) = f_1(z_1)$ *and* $\lim_{n\to\infty} E(\Phi_n(z_1,z_2,a_n)|z_2) = f_2(z_2)$, *a.e., for some measurable functions $f_1(z)$ and $f_2(z)$, and*

(3)

$$\begin{aligned}\lim_{n\to\infty} & E\{[E(\Phi_n(z_1,z_2,a_n)|z_1) + E(\Phi_n(z_2,z_1,a_n)|z_1)] \\ & \cdot[E(\Phi_n(z_1,z_2,a_n)|z_1) + E(\Phi_n(z_2,z_1,a_n)|z_1)]'\} \\ &= E\{[f_1(z_1)+f_2(z_1)][f_1(z_1)+f_2(z_1)]'\} \\ &= \Sigma\end{aligned}$$

where Σ is a finite matrix.

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

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

Proof: This is Lemma 8 in Lee [1988].

Appendix 2: Proofs of Asymptotic Distribution

From (4.1)

$$\begin{aligned}
&\frac{\partial L_n(\alpha,\beta)}{\partial\alpha'} \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_X(x_{1i})\frac{1}{B_n(x_{1i},\alpha)}\left[x_{2i}-\frac{D_n(x_{1i},\alpha)}{B_n(x_{1i},\alpha)}\right]' \\
&\quad\cdot\left[\frac{\partial C_n(x_{1i},\alpha)}{\partial\alpha'}-\beta'\frac{\partial D_n'(x_{1i},\alpha)}{\partial\alpha'}-\frac{A_n(x_{1i},\alpha,\beta)}{B_n(x_{1i},\alpha)}\frac{\partial B_n(x_{1i},\alpha)}{\partial\alpha'}\right] \\
&\quad-\frac{1}{n}\sum_{i=1}^{n} I_X(x_{1i})\left(y_{2i}-x_{2i}\beta-\frac{A_n(x_{1i},\alpha,\beta)}{B_n(x_{1i},\alpha)}\right)\frac{1}{B_n(x_{1i},\alpha)} \\
&\quad\cdot\left[\frac{\partial D_n'(x_{1i},\alpha)}{\partial\alpha'}-\frac{D_n'(x_{1i},\alpha)}{B_n(x_{1i},\alpha)}\frac{\partial B_n(x_{1i},\alpha)}{\partial\alpha'}\right]
\end{aligned}\tag{A.2.1}$$

where

$$\begin{aligned}
&\frac{\partial B_n(x_{1i},\alpha)}{\partial\alpha} \\
&= \frac{1}{n-1}\sum_{j\neq i}^{n}(x_{1i}-x_{1j})'\int_{\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}}^{\infty}\frac{1}{a_n}K\left(\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n},z\right)dz \\
&\quad-\frac{1}{n-1}\sum_{j\neq i}^{n}(x_{1i}-x_{1j})'\int_{\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n}}^{\infty}\frac{1}{a_n}K\left(u,\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}\right)du \\
&= \int_{x_{1i}\alpha+\Delta_n}^{\infty}\frac{1}{n-1}\sum_{j\neq i}^{n}(x_{1i}-x_{1j})'\frac{1}{a_n^2}K\left(\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n},\frac{z-x_{1j}\alpha}{a_n}\right)dz \\
&\quad-\int_{-x_{1i}\alpha}^{\infty}\frac{1}{n-1}\sum_{j\neq i}^{n}(x_{1i}-x_{1j})'\frac{1}{a_n^2}K\left(\frac{u-u_j(\alpha)}{a_n},\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}\right)du,
\end{aligned}\tag{A.2.2}$$

$$\begin{aligned}
&\frac{\partial C_n(x_{1i},\alpha)}{\partial\alpha} \\
&= \frac{1}{n-1}\sum_{j\neq i}^{n}y_{2j}(x_{1i}-x_{1j})'\int_{\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}}^{\infty}\frac{1}{a_n}K\left(\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n},z\right)dz \\
&\quad-\frac{1}{n-1}\sum_{j\neq i}^{n}y_{2j}(x_{1i}-x_{1j})'\int_{\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n}}^{\infty}\frac{1}{a_n}K\left(u,\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}\right)du \\
&= \int_{x_{1i}\alpha+\Delta_n}^{\infty}\frac{1}{n-1}\sum_{j\neq i}^{n}y_{2j}(x_{1i}-x_{1j})'\frac{1}{a_n^2}K\left(\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n},\frac{z-x_{1j}\alpha}{a_n}\right)dz \\
&\quad-\int_{-x_{1i}\alpha}^{\infty}\frac{1}{n-1}\sum_{j\neq i}^{n}y_{2j}(x_{1i}-x_{1j})'\frac{1}{a_n^2}K\left(\frac{u-u_j(\alpha)}{a_n},\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}\right)du
\end{aligned}\tag{A.2.3}$$

and

$$
\begin{aligned}
&\frac{\partial D_n(x_{1i},\alpha)}{\partial \alpha} \\
&= \frac{1}{n-1}\sum_{j\neq i}^{n}(x_{1i}-x_{1j})'x_{2j}\int_{\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}}^{\infty}\frac{1}{a_n}K\left(\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n},z\right)dz \\
&\quad-\frac{1}{n-1}\sum_{j\neq i}^{n}(x_{1i}-x_{1j})'x_{2j}\int_{\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n}}^{\infty}\frac{1}{a_n}K\left(u,\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}\right)du \qquad (A.2.4)\\
&= \int_{x_{1i}\alpha+\Delta_n}^{\infty}\frac{1}{n-1}\sum_{j\neq i}^{n}(x_{1i}-x_{1j})'x_{2j}\frac{1}{a_n^2}K\left(\frac{-x_{1i}\alpha-u_j(\alpha)}{a_n},\frac{z-x_{1j}\alpha}{a_n}\right)dz \\
&\quad-\int_{-x_{1i}\alpha}^{\infty}\frac{1}{n-1}\sum_{j\neq i}^{n}(x_{1i}-x_{1j})'x_{2j}\frac{1}{a_n^2}K\left(\frac{u-u_j(\alpha)}{a_n},\frac{x_{1i}\alpha+\Delta_n-x_{1j}\alpha}{a_n}\right)du.
\end{aligned}
$$

Under the conditions 1(vii)-1(ix) and 2(ii) of Assumption 3, Proposition 4 of Appendix 1 implies that $var(\frac{\partial B_n(x_{1i},\alpha)}{\partial\alpha}|x_{1i})$, $var(\frac{\partial C_n(x_{1i},\alpha)}{\partial\alpha}|x_{1i})$ and $var(\frac{\partial D_n(x_{1i},\alpha)}{\partial\alpha}|x_{1i})$ have order $O(\frac{1}{na_n})$ uniformly in $x_{1i}\in X$. Under Assumption 3, Proposition 3 implies that

$$
\begin{aligned}
\sup_{(\alpha,x_{1i})\in\Theta_1\times X}\Big\|&E(\frac{\partial B_n(x_{1i},\alpha)}{\partial\alpha}|x_{1i})-\{\int_{x_{1i}\alpha+\Delta_n}^{\infty}[x_{1i}-E(x_1|-x_{1i}\alpha,z)]'g(-x_{1i}\alpha,z|\alpha)dz \\
&-\int_{-x_{1i}\alpha}^{\infty}[x_{1i}-E(x_1|u,x_{1i}\alpha+\Delta_n)]'g(u,x_{1i}\alpha+\Delta_n|\alpha)du\}\Big\| = O(a_n^2), \qquad (A.2.5)
\end{aligned}
$$

$$
\begin{aligned}
\sup_{(\alpha,x_{1i})\in\Theta_1\times X}\Big\|&E(\frac{\partial C_n(x_{1i},\alpha)}{\partial\alpha}|x_{1i})-\{\int_{x_{1i}\alpha+\Delta_n}^{\infty}E(y_2(x_{1i}-x_1)'|-x_{1i}\alpha,z)g(-x_{1i}\alpha,z|\alpha)dz \\
&-\int_{-x_{1i}\alpha}^{\infty}E(y_2(x_{1i}-x_1)'|u,x_{1i}\alpha+\Delta_n)g(u,x_{1i}\alpha+\Delta_n|\alpha)du\}\Big\| = O(a_n^2) \qquad (A.2.6)
\end{aligned}
$$

and

$$
\begin{aligned}
\sup_{(\alpha,x_{1i})\in\Theta_1\times X}\Big\|&E(\frac{\partial D_n(x_{1i},\alpha)}{\partial\alpha}|x_{1i})-\{\int_{x_{1i}\alpha+\Delta_n}^{\infty}E((x_{1i}-x_1)'x_2|-x_{1i}\alpha,z)g(-x_{1i}\alpha,z|\alpha)dz \\
&-\int_{-x_{1i}\alpha}^{\infty}E((x_{1i}-x_1)'x_2|u,x_{1i}\alpha+\Delta_n)g(u,x_{1i}\alpha+\Delta_n|\alpha)du\}\Big\| = O(a_n^2). \qquad (A.2.7)
\end{aligned}
$$

The limit functions in (A.2.5)-(A.2.7) are uniformly continuous in (x_{1i},α,Δ_n) on $X\times\Theta_1\times[0,\ 1]$ by Assumptions 2 and 3. With $d=1$, $\bar{d}=1$, $\delta=2$ and $r=2$, Proposition 1 implies that as $\lim_{n\to\infty}\frac{n}{\ln n}a_n^3=\infty$,

$$
\begin{aligned}
\frac{\partial B_n(x_{1i},\bar{\alpha})}{\partial\alpha}\xrightarrow{p}&\int_{x_{1i}\alpha_o}^{\infty}[x_{1i}-E(x_1|z,\alpha_o)]'g(-x_{1i}\alpha_o,z|\alpha_o)dz \\
&-[x_{1i}-E(x_1|x_{1i}\alpha_o)]'\int_{-x_{1i}\alpha_o}^{\infty}g(u,x_{1i}\alpha_o|\alpha_o)du, \qquad (A.2.8)
\end{aligned}
$$

$$\frac{\partial C_n(x_{1i},\bar{\alpha})}{\partial \alpha} \xrightarrow{p} \int_{x_{1i}\alpha_o}^{\infty} E(y_2(x_{1i}-x_1)'|-x_{1i}\alpha_o,z)g(-x_{1i}\alpha_o,z|\alpha_o)dz - \int_{-x_{1i}\alpha_o}^{\infty} E(y_2(x_{1i}-x_1)'|u,x_{1i}\alpha_o)g(u,x_{1i}\alpha_o|\alpha_o)du \tag{A.2.9}$$

and

$$\frac{\partial D_n(x_{1i},\bar{\alpha})}{\partial \alpha} \xrightarrow{p} \int_{x_{1i}\alpha_o}^{\infty} E((x_{1i}-x_1)'x_2|z,\alpha_o)g(-x_{1i}\alpha_o,z|\alpha_o)dz - E((x_{1i}-x_1)'x_2|x_{1i}\alpha_o)\int_{-x_{1i}\alpha_o}^{\infty} g(u,x_{1i}\alpha_o|\alpha_o)du \tag{A.2.10}$$

uniformly in $x_{1i}\in X$. With (3.16), (3.21) and (A.2.8)-(A.2.10),

$$\left(\frac{\partial C_n(x_{1i},\bar{\alpha})}{\partial \alpha'} - \beta_o{}'\frac{\partial D_n'(x_{1i},\bar{\alpha})}{\partial \alpha'} - \frac{A_n(x_{1i},\bar{\alpha},\beta_o)}{B_n(x_{1i},\bar{\alpha})}\frac{\partial B_n(x_{1i},\bar{\alpha})}{\partial \alpha'}\right)/B_n(x_{1i},\bar{\alpha}) \xrightarrow{p} G(x_{1i},\alpha_o) \tag{A.2.11}$$

uniformly in $x_{1i}\in X$, where

$$\begin{aligned} &G(x_{1i},\alpha_o)\\ =&\frac{1}{\int_{x_{1i}\alpha_o}^{\infty}\int_{-x_{1i}\alpha_o}^{\infty} g(u,z|\alpha_o)dudz}\Big\{\int_{x_{1i}\alpha_o}^{\infty} E(v(x_{1i}-x_1)|-x_{1i}\alpha_o,z)g(-x_{1i}\alpha_o,z|\alpha_o)dz\\ &-\int_{-x_{1i}\alpha_o}^{\infty} E(v(x_{1i}-x_1)|u,x_{1i}\alpha_o)g(u,x_{1i}\alpha_o|\alpha_o)du\\ &-E(v|u>-x_{1i}\alpha_o)\int_{x_{1i}\alpha_o}^{\infty}(x_{1i}-E(x_1|z,\alpha_o))g(-x_{1i}\alpha_o,z|\alpha_o)dz\\ &+E(v|u>-x_{1i}\alpha_o)(x_{1i}-E(x_1|x_{1i}\alpha_o))\int_{-x_{1i}\alpha_o}^{\infty} g(u,x_{1i}\alpha_o|\alpha_o)du\Big\}. \end{aligned} \tag{A.2.12}$$

The conditional density $g(u,z|\alpha_o)$ is proportional to the product of the marginal density $f_u(\cdot)$ of u and the marginal density $h(z)$ of $x_1\alpha_o$, i.e.,

$$g(u,z|\alpha_o) = \frac{f_u(u)h(z)}{D} \tag{A.2.13}$$

where $D=\int_{-\infty}^{\infty}\int_{-z}^{\infty} f_u(t)h(z)dtdz$ is the probability of the event $x_1\alpha_o+u>0$. Using (A.2.13) and the property that (u,v) is independent with x in the latent model, the function $G(x_{1i},\alpha_o)$ can be simplified to

$$G(x_{1i},\alpha_o) = \big(E(v|u=-x_{1i}\alpha_o) - E(v|u>-x_{1i}\alpha_o)\big)\lambda(-x_{1i}\alpha_o)[x_{1i}-E(x_1|x_1\alpha_o>x_{1i}\alpha_o)] \tag{A.2.14}$$

where $\lambda(z) = f_u(z)/\int_z^\infty f_u(t)dt$ is the hazard function of u. The expectation of v conditional on $u > z$ where z is a constant argument is

$$E(v|u > z) = \int_z^\infty \int_{-\infty}^\infty vf(v,u)dvdu / \int_z^\infty f_u(t)dt.$$

It follows that

$$\begin{aligned}\frac{\partial E(v|u>z)}{\partial z} &= -\frac{\int_{-\infty}^\infty vf(v,z)dv}{\int_z^\infty f_u(t)dt} + E(v|u>z)\lambda(z) \\ &= -(E(v|u=z) - E(v|u>z))\lambda(z).\end{aligned}$$

Therefore,

$$G(x_{1i},\alpha_o) = -\tau(-x_{1i}\alpha_o)[x_{1i} - E(x_1|x_1\alpha_o > x_{1i}\alpha_o)] \tag{A.2.15}$$

where $\tau(z)$ denotes $\frac{\partial E(v|u>z)}{\partial z}$. With the uniform convergence of the nonparametric functions and their derivatives, the first term on the right hand side of (A.2.1) will converge in probability to $E(I_X(x_{1i})[x_{2i} - E(x_2|x_1\alpha_o > x_{1i}\alpha_o)]\tau(-x_{1i}\alpha_o)[x_{1i} - E(x_1|x_1\alpha_o > x_{1i}\alpha_o)]$. Similarly, the second term will converge in probability to a limit which is apparently zero. This establishes (4.3).

The asymptotic distribution of $\sqrt{n}L_n(\alpha_o,\beta_o)$ can be analyzed as follows. Define

$$L_{n,i}(\alpha_o,\beta_o) = I_X(x_{1i})\left[x_{2i} - \frac{D_n(x_{1i},\alpha_o)}{B_n(x_{1i},\alpha_o)}\right]'\left(y_{2i} - x_{2i}\beta_o - \frac{A_n(x_{1i},\alpha_o,\beta_o)}{B_n(x_{1i},\alpha_o)}\right). \tag{A.2.16}$$

By a Taylor series expansion up to the second order,

$$L_{n,i}(\alpha_o,\beta_o) = S_n(y_{2i},x_i,\alpha_o,\beta_o) + R_n(y_{2i},x_i,\alpha_o,\beta_o)$$

where

$$\begin{aligned}
&S_n(y_{2i},x_i,\alpha_o,\beta_o) \\
&= I_X(x_{1i})\Big\{\left(x_{2i} - \frac{D(x_{1i},\alpha_o,\Delta_n)}{B(x_{1i},\alpha_o,\Delta_n)}\right)'\left(y_{2i} - x_{2i}\beta_o - \frac{A(x_{1i},\alpha_o,\beta_o,\Delta_n)}{B(x_{1i},\alpha_o,\Delta_n)}\right) \\
&\quad + \Big[\frac{D'(x_{1i},\alpha_o,\Delta_n)}{B^2(x_{1i},\alpha_o,\Delta_n)}\left(y_{2i} - x_{2i}\beta_o - \frac{A(x_{1i},\alpha_o,\beta_o,\Delta_n)}{B(x_{1i},\alpha_o,\Delta_n)}\right) \\
&\qquad + \left(x_{2i} - \frac{D(x_{1i},\alpha_o,\Delta_n)}{B(x_{1i},\alpha_o,\Delta_n)}\right)'\frac{A(x_{1i},\alpha_o,\beta_o,\Delta_n)}{B^2(x_{1i},\alpha_o,\Delta_n)}\Big]\cdot(B_n(x_{1i},\alpha_o) - B(x_{1i},\alpha_o,\Delta_n)) \\
&\quad - \left(x_{2i} - \frac{D(x_{1i},\alpha_o,\Delta_n)}{B(x_{1i},\alpha_o,\Delta_n)}\right)'\frac{1}{B(x_{1i},\alpha_o,\Delta_n)}(A_n(x_{1i},\alpha_o,\beta_o) - A(x_{1i},\alpha_o,\beta_o,\Delta_n)) \\
&\quad - \left(y_{2i} - x_{2i}\beta_o - \frac{A(x_{1i},\alpha_o,\beta_o,\Delta_n)}{B(x_{1i},\alpha_o,\Delta_n)}\right)\frac{1}{B(x_{1i},\alpha_o,\Delta_n)}(D_n(x_{1i},\alpha_o) - D(x_{1i},\alpha_o,\Delta_n))'\Big\}
\end{aligned} \tag{A.2.17}$$

with $A(x_{1i},\alpha_o,\beta_o,\Delta_n) = C(x_{1i},\alpha_o,\Delta_n) - D(x_{1i},\alpha_o,\Delta_n)\beta_o$, and

$$
\begin{aligned}
&R_n(y_{2i},x_i,\alpha_o,\beta_o) \\
&= I_X(x_{1i})\Big\{\Big[\frac{\tilde{A}_n(x_{1i})}{\tilde{B}_n^4(x_{1i})}\tilde{D}'_n(x_{1i}) - \frac{\tilde{D}'_n(x_{1i})}{\tilde{B}_n^3(x_{1i})}\left(y_{2i} - x_{2i}\beta_o - \frac{\tilde{A}_n(x_{1i})}{\tilde{B}_n(x_{1i})}\right) \\
&\qquad - \left(x_{2i} - \frac{\tilde{D}_n(x_{1i})}{\tilde{B}_n(x_{1i})}\right)' \frac{\tilde{A}_n(x_{1i})}{\tilde{B}_n^4(x_{1i})}\Big] \cdot (B_n(x_{1i},\alpha_o) - B(x_{1i},\alpha_o,\Delta_n))^2 \\
&\quad + \left[\frac{1}{\tilde{B}_n^2(x_{1i})}\left(x_{2i} - \frac{\tilde{D}_n(x_{1i})}{\tilde{B}_n(x_{1i})}\right)' - \frac{\tilde{D}'_n(x_{1i})}{\tilde{B}_n^3(x_{1i})}\right](B_n(x_{1i},\alpha_o) - B(x_{1i},\alpha_o,\Delta_n)) \\
&\quad \cdot (A_n(x_{1i},\alpha_o,\beta_o) - A(x_{1i},\alpha_o,\beta_o,\Delta_n)) \\
&\quad + \left[\frac{1}{\tilde{B}_n^2(x_{1i})}\left(y_{2i} - x_{2i}\beta_o - \frac{\tilde{A}_n(x_{1i})}{\tilde{B}_n(x_{1i})}\right) - \frac{\tilde{A}_n(x_{1i})}{\tilde{B}_n^3(x_{1i})}\right](B_n(x_{1i},\alpha_o) - B(x_{1i},\alpha_o,\Delta_n)) \\
&\quad \cdot (D_n(x_{1i},\alpha_o) - D(x_{1i},\alpha_o,\Delta_n))' \\
&\quad + \frac{1}{\tilde{B}_n^2(x_{1i})}(A_n(x_{1i},\alpha_o) - A(x_{1i},\alpha_o,\Delta_n))(D_n(x_{1i},\alpha_o) - D(x_{1i},\alpha_o,\Delta_n))'
\end{aligned}
\tag{A.2.18}
$$

with $\tilde{B}_n(x_{1i})$ lying between $B_n(x_{1i},\alpha_o,\Delta_n)$ and $B(x_{1i},\alpha_o,\Delta_n)$, etc. It follows that

$$
\sqrt{n}L_n(\alpha_o,\beta_o) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(y_{2i},x_i,\alpha_o,\beta_o) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_n(y_{2i},x_i,\alpha_o,\beta_o). \tag{A.2.19}
$$

The remainder $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_n(y_{2i},x_i,\alpha_o,\beta_o)$ can be shown to converge in probability to zero. Consider, for example, the last component in the remainder:

$$
\begin{aligned}
&\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_{1i})\frac{1}{\tilde{B}_n^2(x_{1i})}(A_n(x_{1i},\alpha_o) - A(x_{1i},\alpha_o,\Delta_n))(D_n(x_{1i},\alpha_o) - D(x_{1i},\alpha_o,\Delta_n))'\| \\
&\le \sup_{x_{1i}\in X}|\frac{1}{\tilde{B}_n^2(x_{1i})}|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_{1i})|A_n(x_{1i},\alpha_o) - A(x_{1i},\alpha_o,\Delta_n)| \cdot \|D_n(x_{1i},\alpha_o) - D(x_{1i},\alpha_o,\Delta_n)\| \\
&= O_p(1)\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_{1i})|A_n(x_{1i},\alpha_o) - A(x_{1i},\alpha_o,\Delta_n)| \cdot \|D_n(x_{1i},\alpha_o) - D(x_{1i},\alpha_o,\Delta_n)\|.
\end{aligned}
\tag{A.2.20}
$$

With the order of biases in (3.10)-(3.12) and the order of variances in (3.7)-(3.9), Proposition 5 of Appendix 1 implies that as $\lim_{n\to\infty} na_n^8 = 0$, the term on the right hand side of (A.2.20) will converge in probability to zero. Similarly, the other components in the remainder will also converge to zero in probability. Some of the terms in $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(y_{2i},x_i,\alpha_o,\beta_o)$ will also converge to zero in probability. Since v is independent with x in the latent model, $\frac{A(x_{1i},\alpha_o,\beta_o,\Delta_n)}{B(x_{1i},\alpha_o,\Delta_n)} = E(v|u > -x_{1i}\alpha_o)$.

It follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_{1i})\frac{D'(x_{1i},\alpha_o,\Delta_n)}{B^2(x_{1i},\alpha_o,\Delta_n)}\left(y_{2i}-x_{2i}\beta_o-\frac{A(x_{1i},\alpha_o,\beta_o,\Delta_n)}{B(x_{1i},\alpha_o,\Delta_n)}\right)$$
$$\cdot(B_n(x_{1i},\alpha_o)-B(x_{1i},\alpha_o,\Delta_n)) \qquad (A.2.21)$$
$$=\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Phi_{n,1}(z_i,z_j)$$

where $z_i=(x_i,y_i)$, $z_j=(x_j,y_j)$,

$$\Phi_{n,1}(z_i,z_j)$$
$$=I_X(x_{1i})\frac{D'(x_{1i},\alpha_o,\Delta_n)}{B^2(x_{1i},\alpha_o,\Delta_n)}\epsilon_{2i}\left(\int_{\frac{x_{1i}\alpha_o+\Delta_n-x_{1j}\alpha_o}{a_n}}^{\infty}\int_{\frac{-x_{1i}\alpha_o-u_j}{a_n}}^{\infty}K(u,z)dudz-B(x_{1i},\alpha_o,\Delta_n)\right) \qquad (A.2.22)$$

and $\epsilon_{2i}=y_{2i}-x_{2i}\beta_o-E(v|u>-x_{1i}\alpha_o)$. Proposition 6 of Appendix 1 can be applied to (A.2.21). The variance of $\Phi_{n,1}(z_i,z_j)$ has apparently order O(1). Since $E(\Phi_{n,1}(x_i,y_i,z_j)|x_i,z_j)=0$, $E(\Phi_{n,1}(z_i,z_j))=0$. The remaining conditions in Proposition 6 are apparently satisfied for $\Phi_{n,1}(\cdot)$. Therefore, (A.2.21) converges to zero in probability. Similarly, Proposition 6 implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_{1i})\frac{1}{B(x_{1i},\alpha_o,\Delta_n)}(D_n(x_{1i},\alpha_o)-D(x_{1i},\alpha_o,\Delta_n))'\epsilon_{2i}\xrightarrow{p}0.$$

Hence, (A.2.19) reduces to

$$\sqrt{n}L_n(\alpha_o,\beta_o)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_{1i})(x_{2i}-E(x_2|x_1\alpha_o>x_{1i}\alpha_o+\Delta_n))'$$
$$\cdot\{\epsilon_{2i}+\frac{1}{B(x_{1i},\alpha_o,\Delta_n)}(E(v|u>x_{1i}\alpha_o)B_n(x_{1i},\alpha_o)-A_n(x_{1i},\alpha_o))\}+o_p(1)$$
$$=\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} I_X(x_{1i})(x_{2i}-E(x_2|x_1\alpha_o>x_{1i}\alpha_o+\Delta_n))'$$
$$\cdot\left\{\epsilon_{2i}-\frac{(v_j-E(v|u>-x_{1i}\alpha_o))}{B(x_{1i},\alpha_o,\Delta_n)}\int_{x_{1i}\alpha_o+\Delta_n}^{\infty}\int_{-x_{1i}\alpha_o}^{\infty}\frac{1}{a_n^2}K\left(\frac{u-u_j}{a_n},\frac{z-x_{1j}\alpha_o}{a_n}\right)dudz\right\}+o_p(1). \qquad (A.2.23)$$

The asymptotic distribution of (A.2.23) can be derived by the central limit theorem for U-statistics in Proposition 7 of Appendix 1. Denote

$$\Psi_{n,1}(z_i,z_j)=I_X(x_{1i})(x_{2i}-E(x_2|x_1\alpha_o>x_{1i}\alpha_o+\Delta_n))'$$
$$\cdot\left(\epsilon_{2i}-\frac{(v_j-E(v|u>-x_{1i}\alpha_o))}{B(x_{1i},\alpha_o,\Delta_n)}\int_{x_{1i}\alpha_o+\Delta_n}^{\infty}\int_{-x_{1i}\alpha_o}^{\infty}\frac{1}{a_n^2}K\left(\frac{u-u_j}{a_n},\frac{z-x_{1j}\alpha_o}{a_n}\right)dudz\right). \qquad (A.2.24)$$

Equations (3.11) and (3.12) imply that

$$\sup_{x_{1i}} \Big| E(v_j \int_{x_{1i}\alpha_o+\Delta_n}^{\infty} \int_{-x_{1i}\alpha_o}^{\infty} \frac{1}{a_n^2} K\left(\frac{u-u_j}{a_n}, \frac{z-x_{1j}\alpha_o}{a_n}\right) dudz | x_{1i}) - E(v|u > -x_{1i}\alpha_o) B(x_{1i}, \alpha_o, \Delta_n) \Big| = O(a_n^2). \tag{A.2.25}$$

It follows from (3.10) and (A.2.25) that $E(\Psi_{n,1}(x_i, y_i, x_j, y_j)) = O(a_n^2)$ and

$$\lim_{n\to\infty} E(\Psi_{n,1}(z_i, z_j)|z_i) = \Psi_1^{(1)}(z_i) \tag{A.2.26}$$

where $\Psi_1^{(1)}(z_i)$ is defined in (4.6). On the other hand,

$$\begin{aligned} &E(\Psi_{n,1}(z_j, z_i)|z_i) \\ &= -E\Big\{ I_X(x_{1j})(x_{2j} - E(x_2|x_1\alpha_o > x_{1j}\alpha_o + \Delta_n))' \frac{v_i - E(v|u > -x_{1j}\alpha_o)}{B(x_{1j}, \alpha_o, \Delta_n)} \\ &\quad \cdot \int_{\frac{x_{1j}\alpha_o+\Delta_n-x_{1i}\alpha_o}{a_n}}^{\infty} \int_{\frac{-x_{1j}\alpha_o-u_i}{a_n}}^{\infty} K(u,z) dudz | x_i, y_i \Big\}. \end{aligned} \tag{A.2.27}$$

The nonparametric probability function $\int_{\frac{x_{1j}\alpha_o+\Delta_n-x_{1i}\alpha_o}{a_n}}^{\infty} \int_{\frac{-x_{1j}\alpha_o-u_i}{a_n}}^{\infty} K(u,z)dudz$ converges to 1 if $x_{1j}\alpha_o < x_{1i}\alpha_o$ and $-x_{1j}\alpha_o < u_i$; to $\int_{-\infty}^{\infty}\int_0^{\infty} K(u,z)dudz$ if $x_{1j}\alpha_o < x_{1i}\alpha_o$ and $u_i = -x_{1j}\alpha_o$; and to 0, otherwise. The event $u_i = -x_{1j}\alpha_o$ occurs with zero probability. By LDC theorem,

$$\lim_{n\to\infty} E(\Psi_{n,1}(z_j, z_i)|z_i) = \Psi_1^{(2)}(z_i) \tag{A.2.28}$$

where $\Psi_1^{(2)}(z_i)$ is defined in (4.7). The limiting distribution of $\sqrt{n}L_n(\alpha_o, \beta_o)$ follows from Proposition 7. This establishes (4.5).

With the estimate $\hat{\alpha}$ in (4.8), it follows from (4.2) and (A.2.23) that

$$\sqrt{n}L_n(\hat{\alpha}, \beta_o) = \left[I, \ \frac{\partial L_n(\bar{\alpha}, \beta_o)}{\partial \alpha'} C^{-1} \right] \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \Psi_n(z_i, z_j) + o_p(1) \tag{A.2.29}$$

where $\Psi_n(z_i, z_j) = (\Psi'_{n,1}(z_i, z_j), \ \Psi'_{n,2}(z_i, z_j))'$ and

$$\begin{aligned} \Psi_{n,2}(z_i, z_j) =& I_X(x_{1i})(x_{1i} - E(x_1|x_1\alpha_o > x_{1i}\alpha_o + \Delta_n))' \frac{\tau_1(x_{1i}\alpha_o)}{B(x_{1i}, \alpha_o, \Delta_n)} \\ &\cdot \int_{x_{1i}\alpha_o+\Delta_n}^{\infty} \int_{-x_{1i}\alpha_o}^{\infty} (u_i - u) \frac{1}{a_n^2} K\left(\frac{u-u_j}{a_n}, \frac{z-x_{1j}\alpha_o}{a_n}\right) dudz. \end{aligned} \tag{A.2.30}$$

The CLT in Proposition 7 of Appendix 1 can be applied to derive the asymptotic distribution of (A.2.29). For $\Psi_{n,2}(z_i, z_j)$, it has been shown in Lee [1988] that $E(\Psi_{n,2}(z_i, z_j)) = O(a_n^2)$, $var(\Psi_{n,2}(z_i, z_j)) = O(1)$,

$$\lim_{n\to\infty} E(\Psi_{n,2}(z_i, z_j)|z_i) = I_X(x_{1i})(x_{1i} - E(x_1|x_1\alpha_o > x_{1i}\alpha_o))'\tau_1(x_{1i}\alpha_o)\epsilon_{1i} \tag{A.2.31}$$

where $\epsilon_{1i} = y_{1i} - x_{1i}\alpha_o - E(u|u > -x_{1i}\alpha_o)$, and

$$\lim_{n\to\infty} E(\Psi_{n,2}(z_j, z_i)|z_i)$$
$$= -\int_{-u_i}^{x_{1i}\alpha_o} E\big[I_X(x_{1j})(x_{1j} - E(x_1|x_1\alpha_o > x_{1j}\alpha_o))'|x_{1j}\alpha_o = z\big]\frac{u_i - E(u|u > -z)}{\int_z^\infty h(t)dt}h(z)dz. \tag{A.2.32}$$

Proposition 7 implies that as $\lim_{n\to\infty} na_n^4 = 0$,

$$\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Psi_n(z_i, z_j) \xrightarrow{D} N(0, \Sigma) \tag{A.2.33}$$

where $\Sigma = E(\Psi(z_i)\Psi'(z_i))$ as defined in (4.13). These establish the asymptotic distribution of the two stage estimator $\hat{\beta}$ in (4.11).

Appendix 3: First Stage Estimation

In this appendix, we provide some first stage semiparametric estimates of the decision equation (6.1). The first stage semiparametric estimates in Table 1a and Table 2a are derived from the following semiparametric nonlinear least squares estimation (SNLS) of truncated regression model introduced in Lee [1988]:

$$\min_{\alpha} \frac{1}{n} \sum_{i=1}^{n} I_X(x_{1i})\big(y_{1i} - x_{1i}\alpha - \frac{A_n(x_{1i},\alpha)}{B_n(x_{1i},\alpha)}\big)^2$$

where

$$A_n(x_{1i},\alpha) = \frac{1}{(n-1)a_n^2} \sum_{j\neq i}^{n} \int_{x_{1i}\alpha+\Delta_n}^{\infty} \int_{-x_{1i}\alpha}^{\infty} uK\left[\frac{u-(y_{1j}-x_{1j}\alpha)}{a_n}, \frac{z-x_{1j}\alpha}{a_n}\right] dudz$$

and

$$B_n(x_{1i},\alpha) = \frac{1}{(n-1)a_n^2} \sum_{j\neq i}^{n} \int_{x_{1i}\alpha+\Delta_n}^{\infty} \int_{-x_{1i}\alpha}^{\infty} K\left[\frac{u-(y_{1j}-x_{1j}\alpha)}{a_n}, \frac{z-x_{1j}\alpha}{a_n}\right] dudz.$$

Some of the results have been presented in Lee [1988]. The parametric estimates in Table 3a are derived from the nonlinear least square estimation procedure with normal distributional specification in (6.5).

Table 1a.

Results of 300 Repetitions with Various Sample Sizes

Design: Bandwidth factor $c = 1$

	Normal	Gamma*Normal	-Gamma*Normal
N	Mean SD RMSE	Mean SD RMSE	Mean SD RMSE
30 α_1	.960 .650 .640	.855 .777 .778	1.018 .509 .501
α_2	-.953 .636 .627	-.877 .758 .755	-.982 .474 .466
50 α_1	.934 .525 .524	.954 .605 .601	.963 .388 .386
α_2	-.947 .434 .433	-.947 .580 .577	-1.022 .407 .404
100 α_1	.975 .381 .380	.949 .464 .464	.987 .274 .273
α_2	-.989 .340 .338	-.975 .446 .444	-.982 .266 .265
200 α_1	1.006 .241 .240	1.016 .389 .388	.995 .187 .187
α_2	-.997 .263 .262	-1.024 .359 .359	-.985 .178 .178

Table 2a.
SNLS with Various Banwidth Factors

Design: sample size=100

Factor	Normal Mean SD RMSE	Gamma*Normal Mean SD RMSE	-Gamma*Normal Mean SD RMSE
.10 α_1	.964 .305 .306	.937 .358 .362	.981 .229 .229
α_2	-.957 .290 .292	-.928 .338 .344	-.975 .213 .213
.25 α_1	.984 .325 .324	.963 .400 .400	.992 .235 .234
α_2	-.968 .332 .332	-.961 .389 .389	-.977 .239 .240
.50 α_1	.993 .338 .336	.968 .431 .430	.989 .256 .255
α_2	-.980 .343 .342	-.975 .421 .420	-.976 .243 .243
1.00 α_1	.975 .381 .380	.949 .464 .464	.987 .274 .273
α_2	-.989 .340 .338	-.975 .446 .444	-.982 .266 .265
2.00 α_1	.979 .402 .401	.967 .508 .507	.984 .286 .285
α_2	-1.017 .393 .391	-1.021 .478 .476	-1.007 .273 .272
4.00 α_1	1.078 .440 .445	1.090 .537 .542	1.057 .313 .317
α_2	-1.113 .436 .448	-1.132 .550 .563	-1.089 .329 .339
Min. α_1	1.006 .435 .433	1.016 .525 .523	1.025 .305 .304
α_2	-1.054 .414 .415	-1.057 .534 .534	-1.034 .302 .302
Ave. α_1	.996 .327 .325	.979 .407 .406	.998 .244 .243
α_2	-1.004 .328 .326	-.999 .389 .387	-1.001 .243 .242

Table 3a.
Parametric Normal Nonlinear Least Squares Estimation with Various Sample Sizes

N	Normal Mean SD RMSE	Gamma*Normal Mean SD RMSE	-Gamma*Normal Mean SD RMSE
30 α_1	.965 .426 .420	.870 .534 .541	1.061 .343 .343
α_2	-.974 .357 .352	-.800 .489 .521	-1.061 .299 .300
50 α_1	.972 .327 .325	.927 .370 .373	1.054 .248 .251
α_2	-.971 .282 .281	-.862 .317 .343	-1.065 .228 .235
100 α_1	.990 .230 .229	.945 .246 .251	1.047 .174 .179
α_2	-.982 .194 .194	-.889 .219 .245	-1.060 .161 .171
200 α_1	1.004 .150 .150	.944 .181 .189	1.054 .125 .136
α_2	-.991 .140 .140	-.904 .153 .180	-1.065 .112 .129

References

1. Bierens, H.J.(1985), "Kernel Estimators of Regression Functions". In T.W. Bewley (ed.), *Advances in Econometrics Fifth World Congress*, vol.1, pp.99-144. New York: Cambridge University Press.
2. Chamberlain, G.(1986), "Asymptotic Efficiency in Semi-parametric Models with Censoring", *Journal of Econometrics* 32, pp. 189-218.
3. Cosslett, S.(1984), "Distribution-free Estimation of A Regression Model with Sample Selectivity", forthcoming in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, edited by W.A. Barnett, J. Powell and G. Tauchen.
4. Heckman, J.(1974), "Shadow Prices, Market Wages, and Labor Supply", *Econometrica* 42, pp. 679-694.
5. Heckman, J.(1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models", *Annals of Economic and Social Measurement* 5: 475-492.
6. Ichimura, H.(1987), "Estimation of Single Index Models", Ph.D. Thesis, Department of Economics, M.I.T.
7. Ichimura, H. and L.F. Lee (1988), "Semiparametric Estimation of Multiple Index Models: Single Equation Estimation", forthcoming in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, edited by W.A. Barnett, J. Powell and G.Tauchen.
8. Lee, L.F.(1988), "Semiparametric Nonlinear Least Square Estimation of Truncated Regression Models", Discussion Paper No. 254, Center for Economic Research, University of Minnesota.
9. Lee, L.F.(1989),"Semiparametric Maximum Profile Likelihood Estimation of Polytomous and Sequential Choice Models", Discussion Paper No. 253, Center for Economic Research, U. of Minnesota.
10. Lee, L.F.(1990), "Efficient Semiparametric Scoring Estimation of Sample Selection Models", Discussion Paper No. 255, Center of Economics Research, University of Minnesota.
11. Newey, W.K.(1988), "Two Step Estimation of Sample Selection Models", manuscript, Department of Economics, Princeton University.
12. Powell,J.L.(1984),"Least Absolute Deviations Estimation for the Censored Regression Model", *Journal of Econometrics* 25: 303-325.
13. Powell, J.L.(1987), "Semiparametric Estimation of Bivariate Latent Variable Models", Discussion paper no. 8704, Social Systems Research Institute, U. of Wisconsin.
14. Rao, B.L.S.P.(1983), *Nonparametric Functional Estimation*, New York: Academic Press.
15. Robinson, P.M.(1988), "Root-N-Consistent Semiparametric Regression", *Econometrica* 56, pp. 931-954.