

# Semiparametric two-stage estimation of sample selection models subject to Tobit-type selection rules

Lung-fei Lee\*

*University of Michigan, Ann Arbor, MI 48109-1220, USA*

Received September 1990, final version received October 1992

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 provided.

*Key words:* Sample selection; Truncation; Censoring; Semiparametric estimation. Kernel estimation; Randomization, Regression model

*JEL classification:* 211

## 1. Introduction

Econometric models of discrete choice, limited dependent variables, and sample selection have found interesting applications in empirical studies. Models with parametric distributions, however, may be subject to distributional misspecifications, which might result in inconsistent estimates. Recent research efforts on the estimation of such models have focused on semiparametric methods, which relax parametric distribution assumptions. Semiparametric methods have been proposed for the estimation of sample selection models with discrete choice decision rules. Semiparametric estimation of sample selection

*Correspondence to:* Lung-fei Lee, Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220, USA.

\*I am grateful to Professor Peter Schmidt and two anonymous referees for their comments and suggestions on earlier versions of this article. I appreciate having research support by National Science Foundation grant no. SES-9010516, and computing grant support from the Minnesota Supercomputer Institute for my research. This article was completed during the author's tenure at the University of Minnesota.

models subject to Tobit-type selection rules have 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 model of two equations:

$$y_1 = x\alpha + u, \quad y_2 = x\beta + v, \quad (1.1)$$

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 the positive values of  $y_1$  can be observed instead of just the sign of  $y_1$ . For the semiparametric model with a discrete choice equation for  $y_1$ , Chamberlain (1986) has shown that, under the assumption that  $(u, v)$  is independent of regressors in the model, the identification of  $\beta$  requires exclusion restrictions on the regressors of the  $y_2$  equation. Semiparametric methods for estimation of such a model have been suggested in Cosslett (1991), Robinson (1988), Powell (1987), Ichimura and Lee (1991), 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  $\beta$ . This article proposes a simple semiparametric two-stage method for the estimation of  $\beta$ . Given a consistent estimate of  $\alpha$ , the bias of the observed outcome equation can be adjusted, and  $\beta$  can be estimated by a regression procedure. Our procedure differs from the two-stage estimation procedures in Heckman (1976), Cosslett (1991), Robinson (1988), and Powell (1987) in the way of constructing the bias adjustment term. Our adjustment term is designed for Tobit-type sample selection models. Under general regularity conditions, our two-stage estimator is  $\sqrt{n}$ -consistent and asymptotically normal.

The 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. Section 5 provides a consistent estimate of the covariance matrix of the estimator. Several Monte Carlo simulations are performed to investigate the finite sample performance of the proposed estimator. The simulation results are reported in section 6. Appendix 1 summarizes some relevant results for our analysis. Some proofs of asymptotic properties of the estimator are provided in appendix 2.

## **2. A two-stage semiparametric estimation procedure**

The model considered has two equations. One of them is a Tobit-type selection equation. Let  $x$  be a  $k$ -dimensional vector of regressors in the model.

To be specific,  $x$  does not include a constant term.<sup>1</sup> The  $x_1$  and  $x_2$  are subvectors of  $x$  of dimensions  $k_1$  and  $k_2$ , respectively. The underlying latent equations in the model are

$$y_1 = x_1\alpha_0 + u \tag{2.1}$$

and

$$y_2 = x_2\beta_0 + v, \tag{2.2}$$

where  $\alpha_0$  and  $\beta_0$  are the true parameter column vectors of dimensions  $k_1$  and  $k_2$ , respectively. Values of  $y_1$  and  $y_2$  are defined on the whole real line, but they can be observed only when  $y_1 > 0$ . Eq. (2.1) is a censored regression model if  $\max\{0, y_1\}$  and the corresponding  $x$  are observable. It will be a truncated regression model if only the positive values of  $y_1$  and their corresponding  $x$  are available. In this article, the disturbances  $u$  and  $v$  in (2.1) and (2.2) are assumed to be independent of  $x$  in the model. If observations of  $y_1$  are censored,  $\alpha_0$  can be estimated by, for example, Powell's least absolute deviations method [Powell (1984)].<sup>2</sup> With a truncated sample, the method in Lee (1992) is applicable. Estimation of  $\beta_0$  is the remaining issue. In this article, we generalize the semiparametric estimation method in Lee (1992) for the truncated regression model to the estimation of the above sample selection model.

Conditional on  $y_2$  being observable and  $x$ , the regression function of  $y_2$  is

$$E(y_2 | y_1 > 0, x) = x_2\beta_0 + E(v | u > -x_1\alpha_0, x). \tag{2.3}$$

The two-stage estimation method in Powell (1987) and the semiparametric nonlinear least squares method in Ichimura and Lee (1991) have used the 'index property' that  $E(v | u > -x_1\alpha_0, x)$  is a function of  $x_1\alpha_0$ , i.e.  $E(v | u > -x_1\alpha_0, x) = E(v | u > -x_1\alpha_0, x_1\alpha_0)$ , but not the 'independence property' that  $u$  and  $v$  are independent of  $x$  in the latent structure. When  $x_1 = x_2$  or  $x_2$  is a subvector of  $x_1$ , the index property alone does not provide enough restrictions for identification of  $\beta_0$  [see Chamberlain (1986), Powell (1987), and Ichimura and Lee (1991)]. For our model, let  $f(v, u)$  be the joint density of  $(v, u)$ , let  $f_u(\cdot)$  be the marginal density of  $u$ , and let  $h(\cdot)$  be the density of  $x_1\alpha_0$ . Then

$$E(v | u > -x_1\alpha_0, x) = \frac{\int_{-\infty}^{\infty} \int_{-x_1\alpha_0}^{\infty} vf(v, u) du dv}{\int_{-x_1\alpha_0}^{\infty} f_u(t) dt}. \tag{2.4}$$

<sup>1</sup>For our model, since no moment restrictions are imposed on the disturbances, constant terms in the equations are absorbed into the disturbances

<sup>2</sup>Powell's approach assumes that  $u$  conditional on  $x$  has zero median, which is weaker than the independence assumption

For the index formulation, the property that  $f(v, u)$  is not a function of  $x_1\alpha_0$  has not been imposed on estimation. Imposing this property of the model is the key for identification. At first sight of (2.4), one might attempt to estimate  $f(v, u)$  and  $f_u(u)$  with observable residuals  $v$  and  $u$  by some nonparametric density estimates. Unfortunately, the problem is not so straightforward with truncated (or censored) data, because the joint density of  $(v, u)$  given  $y_1 > 0$  is  $f(v, u) \int_{-u}^{\infty} h(t) dt / D$ , where  $D = \int_{-\infty}^{\infty} (\int_{-z}^{\infty} f_u(t) dt) h(z) dz$  is the probability of the event  $y_1 > 0$ , and the density of  $u$  given  $y_1 > 0$  is  $f_u(u) \int_{-u}^{\infty} h(t) dt / D$ . Estimation of any relevant function of this model should take into account the distribution of  $x_1\alpha_0$ . Multiplying both the numerator and denominator of (2.4) by  $\int_{x_1\alpha_0}^{\infty} h(z) dz$ , (2.4) can be rewritten as

$$E(v | u > -x_1\alpha_0, x) = \frac{\int_{-\infty}^{\infty} \int_{x_1\alpha_0}^{\infty} \int_{-x_1\alpha_0}^{\infty} v f(v, u) h(z) du dz dv}{\int_{x_1\alpha_0}^{\infty} \int_{-x_1\alpha_0}^{\infty} f_u(t) h(z) dt dz} \tag{2.4'}$$

With a random sample  $(y_{1i}, y_{2i}, x_i), y_{1i} > 0, i = 1, \dots, n$ , consisting of independent observations drawn from a common population, (2.4') can be estimated nonparametrically. At  $x_i$ , the proportion of sample observations of  $(x_1\alpha_0, u)$ , which satisfy the conditions  $x_1\alpha_0 > x_{1i}\alpha_0$  and  $u > -x_{1i}\alpha_0$ , will provide a consistent estimate of the probability (conditional on  $x_i$ ) of the event  $u > -x_{1i}\alpha_0$  and  $x_1\alpha_0 > x_{1i}\alpha_0$ ; i.e.,

$$\frac{1}{n} \sum_{j=1}^n I(u_j > x_{1i}\alpha_0, x_{1j}\alpha_0 > x_{1i}\alpha_0) \xrightarrow{p} \int_{x_{1i}\alpha_0}^{\infty} \int_{x_{1i}\alpha_0}^{\infty} f_u(t) h(z) dt dz,$$

by the strong law of large numbers. Similarly,

$$\frac{1}{n} \sum_{j=1}^n v_j I(u_j > x_{1i}\alpha_0, x_{1j}\alpha_0 > x_{1i}\alpha_0) \xrightarrow{p} \int_{-\infty}^{\infty} \int_{x_{1i}\alpha_0}^{\infty} \int_{-x_{1i}\alpha_0}^{\infty} v f(v, u) h(z) du dz dv.$$

The ratio of the latter sample average over the former sample average provides an estimate of (2.4') at  $x_i$ . These sample averages are, however, not easy to work with due to their nonsmooth characteristic as functions of  $\alpha_0$ . Instead of frequency estimators, we consider smooth kernel density estimators. The intuition behind these formulations is based on the observation that the density of  $(v_i, u_i)$  conditional on  $y_{1i} > 0$  and  $x_{1i}\alpha_0$  is the same as the density of any  $(v_j, u_j)$  conditional on  $x_{1i}\alpha_0, u_j > -x_{1i}\alpha_0$  and  $x_{1j}\alpha_0 > x_{1i}\alpha_0$ .<sup>3</sup> In particular, the conditions  $x_1\alpha_0 > x_{1i}\alpha_0$  and  $u > -x_{1i}\alpha_0$  imply  $y_1 = x_1\alpha_0 + u > 0$ .

<sup>3</sup>See also Lee (1992) for the estimation of a truncated regression model and a geometric interpretation

Before we proceed further, let us clarify some of the notations and conventions that will be adopted throughout this article. The random variables  $y_1, y_2, x, u,$  and  $v$  are referred to the random variables in the latent eqs. (2.1) and (2.2). Sample observations are always referred to truncated observations.  $r_j = (y_{1j}, y_{2j}, x_j)$  is a row vector of the  $j$ th sample observations;  $r_{1j} = (y_{1j}, x_{1j})$  is a subvector of  $r_j$ . All the expectations and probability functions are taken with respect to the distribution of the latent random variables before truncation.  $P(S)$  denotes the probability of an event  $S$  under the distribution of the latent random variables. As a convention, with a realization  $x_i$  appearing in any expectation or probability function,  $x_i$  will be regarded as a parameter. For example,  $P(x_1\alpha_0 > x_{1i}\alpha_0)$  refers to the probability of the set  $\{x | x_1\alpha_0 > x_{1i}\alpha_0\}$ , and  $E(v | u > -x_{1i}\alpha_0, x_1\alpha_0 > x_{1i}\alpha_0)$  is the conditional expectation of  $v$  conditional on the set  $\{(u, x) | u > -x_{1i}\alpha_0, x_1\alpha_0 > x_{1i}\alpha_0\}$ .  $I_S(\cdot)$  denotes the indicator of a set  $S$ .  $\alpha$  and  $\beta$  are possible values of  $\alpha_0$  and  $\beta_0$ , respectively.  $\Theta_1 \times \Theta_2$  is the parameter space of  $(\alpha', \beta')$ . We define the variables  $u(\alpha) = y_1 - x_1\alpha$  and  $v(\beta) = y_2 - x_2\beta$ , and their realizations  $u_j(\alpha) = y_{1j} - x_{1j}\alpha$  and  $v_j(\beta) = y_{2j} - x_{2j}\beta$ .  $g(\cdot | x)$  denotes the density function of  $(u(\alpha), x_1\alpha)$ , and  $E(\cdot | \cdot; \alpha)$  is a conditional expectation conditional on  $(u(\alpha), x_1\alpha)$ . The  $u_j = u_j(\alpha_0)$ ,  $v_j = v_j(\beta_0)$ ,  $\varepsilon_{1j} = y_{1j} - x_{1j}\alpha_0 - E(u | u > -x_{1j}\alpha_0)$ , and  $\varepsilon_{2j} = y_{2j} - x_{2j}\beta_0 - E(v | u > -x_{1j}\alpha_0)$  are regression residuals. More notations will be introduced in the due course.

Let  $K(\cdot)$  be a kernel function on  $R^2$  and  $a_n > 0$  be a bandwidth parameter [Rao (1983)]. For any random variable  $s$ , define

$$C_n(s | x_{1i}, \alpha) = \int_{x_{1i}\alpha + \Delta_n}^{\infty} \int_{-x_{1i}\alpha}^{\infty} \frac{1}{(n-1)a_n^2} \sum_{j \neq i}^n s_j K \left[ \frac{u - u_j(\alpha)}{a_n}, \frac{z - x_{1j}\alpha}{a_n} \right] du dz, \tag{2.5}$$

where  $\Delta_n > 0$  is a trimming parameter, and

$$E_n(s | x_{1i}, \alpha) = \frac{C_n(s | x_{1i}, \alpha)}{C_n(1 | x_{1i}, \alpha)}. \tag{2.6}$$

As shown in the subsequent sections under some regularity conditions and the design that  $a_n$  and  $\Delta_n$  go to zero as  $n$  goes to infinity,  $E_n(v(\beta) | x_{1i}, \alpha)$  provides a consistent estimate of  $E(v(\beta) | u(\alpha) > -x_{1i}\alpha, x_1\alpha > x_{1i}\alpha)$ . Given a consistent estimate  $\hat{\alpha}$  of  $\alpha_0$ , our proposed estimation method is a semiparametric least squares procedure:

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

where the set  $X$  is constructed by trimming the regressors in  $x_1$ . The necessity of trimming the regressors will be explained in a later paragraph. The estimator  $\hat{\beta}$  from (2.7) has a closed form expression:

$$\hat{\beta} = \left\{ \sum_{i=1}^n I_X(x_{1i}) [x_{2i} - E_n(x_2 | x_{1i}, \hat{\alpha})]' [x_{2i} - E_n(x_2 | x_{1i}, \hat{\alpha})] \right\}^{-1} \times \sum_{i=1}^n I_X(x_{1i}) [x_{2i} - E_n(x_2 | x_{1i}, \hat{\alpha})]' [y_{2i} - E_n(y_2 | x_{1i}, \hat{\alpha})]. \tag{2.8}$$

This two-stage procedure is similar in a certain way to the two-stage procedures in Robinson (1988) and Powell (1987). Define a weight function:

$$W_n(x_{1i}, r_{1j}, \alpha) = \frac{\int_{x_{1i}\alpha + \Delta_n - x_{1i}\alpha}^{\infty} \int_{-x_{1i}\alpha}^{\infty} K \left[ \frac{u - u_j(\alpha)}{a_n}, \frac{z - x_{1j}\alpha}{a_n} \right] du dz}{\sum_{i \neq j} \int_{x_{1i}\alpha + \Delta_n - x_{1i}\alpha}^{\infty} \int_{-x_{1i}\alpha}^{\infty} K \left[ \frac{u - u_i(\alpha)}{a_n}, \frac{z - x_{1i}\alpha}{a_n} \right] du dz}. \tag{2.9}$$

The two-stage estimator  $\hat{\beta}$  can be rewritten as

$$\hat{\beta} = \left\{ \sum_{i=1}^n I_X(x_{1i}) \left[ x_{2i} - \sum_{j \neq i} x_{2j} W_n(x_{1i}, r_{1j}, \hat{\alpha}) \right] \times \left[ x_{2i} - \sum_{j \neq i} x_{2j} W_n(x_{1i}, r_{1j}, \hat{\alpha}) \right] \right\}^{-1} \times \sum_{i=1}^n I_X(x_{1i}) \left[ x_{2i} - \sum_{j \neq i} x_{2j} W_n(x_{1i}, r_{1j}, \hat{\alpha}) \right]' \times \left[ y_{2i} - \sum_{j \neq i} y_{2j} W_n(x_{1i}, r_{1j}, \hat{\alpha}) \right]. \tag{2.10}$$

The weighting functions in the two-stage procedures in Robinson (1988) and Powell (1987) are, however, quite different from ours. Our weighting function is specific to the sample selection model with a Tobit selection rule. Their weighting functions use index restrictions only. Our semiparametric estimation procedure can also be understood from an angle different from (2.7). At each  $x_i$ , (2.1) and (2.2) imply (2.3):  $E(y_{2i} | y_{1i} > 0, x_i) = x_{2i}\beta_0 + E(v | u > -x_{1i}\alpha_0, x_i)$ . In addition, they imply also that

$$E(y_2 | u > -x_{1i}\alpha_0, x_{1i}\alpha_0 > x_{1i}\alpha_0) = E(x_2 | x_{1i}\alpha_0 > x_{1i}\alpha_0)\beta_0 + E(v | u > -x_{1i}\alpha_0, x_{1i}\alpha_0 > x_{1i}\alpha_0). \tag{2.11}$$

As  $E(v|u > -x_{1i}\alpha_0, x_{1i}\alpha_0 > x_{1i}\alpha_0) = E(v|u > -x_{1i}\alpha_0, x_i)$  is an unknown function, it can be eliminated by taking the difference of (2.3) and (2.11) which gives

$$\begin{aligned} y_{2i} - E(y_2|u > -x_{1i}\alpha_0, x_{1i}\alpha_0 > x_{1i}\alpha_0) \\ = [x_{2i} - E(x_2|x_{1i}\alpha_0 > x_{1i}\alpha_0)]\beta_0 + \varepsilon_{2i}. \end{aligned} \tag{2.12}$$

With  $E(y_2|u > -x_{1i}\alpha_0, x_{1i}\alpha_0 > x_{1i}\alpha_0)$  and  $E(x_2|x_{1i}\alpha_0 > x_{1i}\alpha_0)$  replaced, respectively, by the nonparametric estimates  $E_n(y_2|x_{1i}, \hat{\alpha})$  and  $E_n(x_2|x_{1i}, \hat{\alpha})$ , a least squares procedure applied to (2.12) provides the estimate  $\hat{\beta}$  in (2.8). From this point of view, we see that the estimation procedure has explored the information in (2.12) in addition to (2.3).

The  $\hat{\beta}$  can be shown to be consistent and asymptotically normally distributed under some regularity conditions if  $a_n$  and  $\Delta_n$  are chosen to converge to zero at certain rates as the sample size  $n$  increases.

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 of  $x$ .
- (2) The samples  $(y_{1i}, y_{2i}, x_i), i = 1, \dots, n$ , where  $y_{1i} > 0$  for all  $i$ , are i.i.d.<sup>4</sup>
- (3) The first four order moments of  $(y_{2i}, x_i)$  exist.
- (4)  $\hat{\alpha}$  is a consistent estimate of  $\alpha_0$ .
- (5) For each  $\alpha \in \Theta_1$ , where  $\Theta_1$  is a compact neighborhood of  $\alpha_0$ , the index  $x_1\alpha$  is a continuous random variable.
- (6)  $X$  is chosen to be a compact subset of the support  $S$  of  $x_1$  such that  $\max_{x_i \in X} x_1\alpha_0 < \max_{x_i \in S} x_1\alpha_0$ .
- (7) For each  $x_{1i} \in X$ , there exists, with positive probability, some  $x_1$  in the set  $\{x_1 | x_1\alpha_0 > x_{1i}\alpha_0\}$  such that  $P(y_1 > x_1\alpha_0 - x_{1i}\alpha_0 | x) > 0$ .

*Assumption 2*

- (1) For each  $\alpha \in \Theta_1$ , the density function  $g(w|\alpha)$  of  $(y_1 - x_1\alpha, x_1\alpha)$ ; 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  $W = \{w | w_1 + w_2 > 0 \text{ where } w = (w_1, w_2)\}$ .
- (2)  $g(w|\alpha), E(x_2|w, \alpha)$  and  $E(y_2|w, \alpha)$  are continuous in  $\alpha \in \Theta_1$ .

<sup>4</sup>Our two-stage estimation method does not use information of the event  $y_1 < 0$  once a consistent estimate of  $\alpha_0$  is given. The sample can simply consist of truncated observations.

*Assumption 3*

There exist Lebesgue measurable functions  $h_j(w), j = 1, \dots, 4$ , with the following domination and integrability properties:

(1) Domination properties: Let  $t = 1, x_2$ , or  $y_2$ . In some neighborhood  $N_\delta(w)$  of  $w$  in  $W$  with radius  $\delta > 0$  that does not depend on  $w$ ,

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

$$(ii) \sup_{\alpha \in \Theta_1} \sup_{s \in N_\delta(w)} \left\| \frac{\partial^2}{\partial s \partial s'} [E(t | s, \alpha) g(s | \alpha)] \right\| \leq h_2(w),$$

$$(iii) \sup_{(a, x_{1t}) \in \Theta_1 \times X} \sup_{s \in N_\delta(w)} E(\|x_{1t} - x_1\|' t | s, \alpha) g(s | \alpha) \leq h_3(w),$$

$$(iv) \sup_{(a, x_{1t}) \in \Theta_1 \times X} \sup_{s \in N_\delta(w)} \left\| \frac{\partial^2}{\partial s \partial s'} [E((x_{1t} - x_1)' t | s, \alpha) g(s | \alpha)] \right\| \leq h_4(w),$$

(2) Integrability properties:

$$(i) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_j(u, z) du dz < \infty \quad \text{for } j = 1, 2.$$

$$(ii) \int_{-\infty}^{\infty} \sup_u h_j(u, z) dz < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} \sup_z h_j(u, z) du < \infty$$

for  $j = 3, 4$ .

*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  $\int w K(w) dw = 0$ .

(3) The bandwidth sequence  $\{a_n\}$ , with  $a_n > 0$ , converges to zero at a rate such that  $\lim_{n \rightarrow \infty} (na_n^3 / \ln n) = \infty$  and  $\lim_{n \rightarrow \infty} na_n^4 = 0$ .

(4) The  $\{\Delta_n\}$  is a positive sequence such that  $\lim_{n \rightarrow \infty} \Delta_n = 0$  but  $\lim_{n \rightarrow \infty} (\Delta_n / a_n) = \infty$ .

*Assumption 5 (Identification condition)*

The matrix  $E(I_X(x_{1t}) [x_{2t} - E(x_2 | x_1 \alpha_0 > x_1 \alpha_0)]' [x_{2t} - E(x_2 | x_1 \alpha_0 > x_1 \alpha_0)])$  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  $x_{1i} \in X$ ,  $P(x_1\alpha_0 > x_{1i}\alpha_0 | x_{1i})$  is strictly positive. Assumption 1(7) together with  $X$  implies that  $P(u > -x_1\alpha_0)$  is also strictly bounded away from zero on  $X$ . This assumption is always satisfied if  $u$  is unbounded from above. Without trimming of  $x_1$ , we have some technical difficulty to prove the uniform convergence of  $E_n(v(\beta) | x_{1i}, \alpha)$ , which is a ratio of  $C_n(v(\beta) | x_i, \alpha)$  over  $C_n(1 | x_{1i}, \alpha)$  to a well-defined limit. This is so, because at some boundary points of  $S$ , say  $\bar{x}_1$ , it is possible that the event  $x_1\alpha > \bar{x}_1\alpha$  will occur with zero probability. At those points,  $1/C_n(1 | \bar{x}_1, \alpha)$  will not converge. For the points close to the boundary, even their limits may exist, the limiting values might be very large.<sup>5</sup> Assumptions 2–4 are used to guarantee 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 the Lebesgue dominated convergence theorem (LDC). The rate of convergence of  $a_n$  controls the rate of convergence of the nonparametric functions. The trimming parameter  $\Delta_n$  is used to avoid complicated biases of the nonparametric functions along the boundary of  $y_1 > 0$  so that some uniform rate of asymptotic biases can be established. To clarify this issue a little bit more, consider the point  $(-x_{1i}\alpha_0, x_{1i}\alpha_0)$ . In any neighborhood of this point, some values  $(\bar{u}, \bar{z})$  of  $(u, x_1\alpha_0)$  with  $\bar{u} < -x_{1i}\alpha_0$  would not be observable because  $\bar{y}_1 = \bar{z} + \bar{u}$  could be negative. The kernel density estimate  $(1/(n-1)a_n^2) \times \sum_{j \neq i}^n K((-x_{1i}\alpha_0 - u_j)/a_n, (x_{1i}\alpha_0 - x_{1j}\alpha_0)/a_n)$  would not necessarily converge to the density of  $(u, x_1\alpha_0)$  at  $(-x_{1i}\alpha_0, x_{1i}\alpha_0)$  when the kernel puts some positive weight on every point in its neighborhood. This difficulty might prevent the conditional expectation of  $C_n(1 | x_{1i}, \alpha_0)$  (with  $\Delta_n = 0$ ), conditional on  $x_{1i}$ , to converge to  $\int_{x_{1i}\alpha_0}^{\infty} \int_{-x_{1i}\alpha_0}^{\infty} f_u(t_1)h(t_2)dt_1 dt_2$  with the rate of  $O(a_n^2)$  as required in Proposition 2 of the appendix. The  $\Delta_n$  is designed to overcome such irregularity.<sup>6</sup> For our proposed estimation method, since the rate of convergence of  $a_n$  is not too slow,  $K(\cdot)$  can simply be a density function. The regularity conditions of our assumptions are sufficient to prove that our two-stage estimator is  $\sqrt{n}$ -consistent and is asymptotically equivalent to the sum of an asymptotic normal

<sup>5</sup>Instead of trimming  $x_1$ , an alternative suggestion in Powell (1987) can be applied to our model. Powell's approach is to weight the squared residual at each point  $x_i$  by  $C_n^2(1 | x_{1i}, \hat{\alpha})$  so as to eliminate each denominator in the semiparametric least squares procedure. Following Powell's approach, the semiparametric least squares procedure would be a semiparametric weighted least squares procedure. However, such a weighting has nothing to do with optimal estimation as in the classical Aitken's estimator for linear regression models.

<sup>6</sup>An alternative approach, that might be useful to overcome this difficulty, is to select a kernel function  $K(t_1, t_2)$  with the property that it vanishes whenever  $t_1 > 0$ . Such a kernel for the density estimation of  $(u, x_1\alpha_0)$  at  $(-x_{1i}\alpha_0, x_{1i}\alpha_0)$  will put zero weight on any value  $(\bar{u}, \bar{z})$  with  $\bar{u} < -x_{1i}\alpha_0$ . However, such a kernel could not be a proper density function under the zero mean condition in Assumption 4(2). Kernel functions with some negative values would be needed.

variable and a variable involving  $\hat{\alpha}$ . With a  $\sqrt{n}$ -consistent and asymptotic normal  $\hat{\alpha}$ , our two-stage estimator is asymptotically normal.

### 3. Consistency and identification

Asymptotic properties of  $\hat{\beta}$  depend on properties of the nonparametric functions in (2.5) and (2.6). Let  $t = 1, y_2,$  or  $x_2$ . By a change of variables,

$$C_n(t | x_1, \alpha) = \frac{1}{n-1} \sum_{j \neq t}^n t_j \int_{(x_1, \alpha + \Delta_n - x_{1,j} \hat{\alpha})/a_n}^{\infty} \int_{(-x_1, \alpha - u_j(\hat{\alpha})/a_n)}^{\infty} K(u, z) du dz. \tag{3.1}$$

Since the second moments of  $y_2$  and  $x_2$  are finite and  $\int_{-\infty}^{\infty} |K(w)| dw < \infty$ ,

$$\sup_{x_1, \alpha} \text{var} (C_n(t | x_1, \alpha) | x_1) = O(1/n). \tag{3.2}$$

For any random variable  $s$  and a constant  $\Delta \geq 0$ , denote

$$C(s | x_1, \alpha, \Delta) = \int_{x_1 \alpha + \Delta}^{\infty} \int_{-x_1 \alpha}^{\infty} E(s | u, z, \alpha) g(u, z | \alpha) du dz. \tag{3.3}$$

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

$$\sup_{x_1, \alpha} |E(C_n(t | x_1, \alpha) | x_1) - C(t | x_1, \alpha, \Delta_n)| = O(a_n^2). \tag{3.4}$$

With the LDC theorem, Assumption 2(2) and Assumption 3 imply that  $C(t | x_1, \alpha, \Delta)$  with  $t = 1, y_2,$  or  $x_2$  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 (3.1) with  $d = \bar{d} = 0$ . Since  $\hat{\alpha}$  is consistent, it follows that

$$\text{plim}_{n \rightarrow \infty} \sup_{x_1 \in X} \left| C_n(t | x_1, \hat{\alpha}) - \int_{x_1 \alpha_0}^{\infty} \int_{-x_1 \alpha_0}^{\infty} E(t | u, z, \alpha_0) g(u, z | \alpha_0) du dz \right| = 0. \tag{3.5}$$

On  $X$ , Assumption 1(7) guarantees that the probability  $\int_{x_1 \alpha_0}^{\infty} \int_{-x_1 \alpha_0}^{\infty} g(u, z | \alpha_0) \times du dz$  is uniformly bounded away from zero. As  $u$  and  $v$  are independent of  $x$ ,  $E(v | u > -x_1, \alpha_0, x_1 \alpha_0 > x_1, \alpha_0) = E(v | u > -x_1, \alpha_0)$  and  $E(x_2 | u > -x_1, \alpha_0, x_1 \alpha_0 > x_1, \alpha_0) = E(x_2 | x_1 \alpha_0 > x_1, \alpha_0)$  for all  $x_{1t}$ . Therefore

$$\text{plim}_{n \rightarrow \infty} \sup_{x_{1t} \in X} |E_n(y_2 | x_{1t}, \hat{\alpha}) - E(y_2 | u > -x_{1t}, \alpha_0, x_1 \alpha_0 > x_{1t}, \alpha_0)| = 0 \tag{3.6}$$

and

$$\text{plim}_{n \rightarrow \infty} \sup_{x_{1i} \in X} |E_n(x_2|x_{1i}, \hat{\alpha}) - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0)| = 0. \quad (3.7)$$

Since  $C_n(v(\beta)|x_{1i}, \alpha) = C_n(y_2|x_{1i}, \alpha) - C_n(x_2|x_{1i}, \alpha)\beta$ , (3.6) and (3.7) imply that

$$\text{plim}_{n \rightarrow \infty} \sup_{x_{1i} \in X} |E_n(v|x_{1i}, \hat{\alpha}) - E(v|u > -x_{1i}\alpha_0)| = 0. \quad (3.8)$$

Eq. (2.8) implies that

$$\begin{aligned} \hat{\beta} - \beta_0 &= \left\{ \sum_{i=1}^n I_X(x_{1i}) [x_{2i} - E_n(x_2|x_{1i}, \hat{\alpha})]' [x_{2i} - E_n(x_2|x_{1i}, \hat{\alpha})] \right\}^{-1} \\ &\quad \times \sum_{i=1}^n I_X(x_{1i}) [x_{2i} - E_n(x_2|x_{1i}, \hat{\alpha})]' (v_i - E_n(v|x_{1i}, \hat{\alpha})). \end{aligned} \quad (3.9)$$

Since the first two moments of  $(y_2, x_2)$  are finite, (3.6)–(3.8) and Kolmogorov’s law of large numbers imply that

$$\frac{1}{n} \sum_{i=1}^n I_X(x_{1i}) [x_{2i} - E_n(x_2|x_{1i}, \hat{\alpha})]' [x_{2i} - E_n(x_2|x_{1i}, \hat{\alpha})] \xrightarrow{p} A, \quad (3.10)$$

where  $A = E(I_X(x_{1i}) [x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0)]' [x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0)])$ , and

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n I_X(x_{1i}) [x_{2i} - E_n(x_2|x_{1i}, \hat{\alpha})]' (v_i - E_n(v|x_{1i}, \hat{\alpha})) \\ &\xrightarrow{p} E\{I_X(x_{1i}) [x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0)]' \varepsilon_{2i}\} = 0. \end{aligned} \quad (3.11)$$

The consistency of  $\hat{\beta}$  follows from (3.9)–(3.11) and the identification condition in Assumption 5.

The identification condition requires that the components of the random vector  $x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0)$ , with  $x_{1i} \in X$ , are not linearly dependent a.e. This identification condition is apparent as the estimation procedure is applied to the estimation of (2.12). For the special case that  $x_2$  is independent of  $x_1$ , this condition will be reduced 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  $E[I_X(x_i)(x_i - E(x|x\alpha_0 > x_i\alpha_0))^2] > 0$ , which holds as  $x_i < E(x|x > x_i)$  and  $x_i > E(x|x < x_i)$  for all  $x_i \in X$ . The identification of  $\beta$  is based on

information contained in both the bias-adjusted eqs. (2.3) and (2.11). Eq. (2.3) alone may not be enough for identification of  $\beta$  when  $x_1 = x_2 = x$ . For example, if  $E(v|u)$  is a linear function of  $u$  and  $E(u|u > -x\alpha_0)$  is a linear function of  $x\alpha_0$ ,  $E(v|u > -x\alpha, x)$  will be a linear function of  $x\alpha$  and  $\beta$  in (2.3) will not be identifiable due to perfect multicollinearity.<sup>7</sup> Eq. (2.11) provides additional identification restriction because even if  $E(v|u > -x_1\alpha_0)$  were linear in  $x_1\alpha_0$ ,  $E(x|x\alpha_0 > x_1\alpha_0)$  would, in general, not be linear in  $x_1\alpha_0$ .

**4. Asymptotic distribution**

The asymptotic distribution of  $\hat{\beta}$  can be derived from (3.9). Denote

$$L_n(\alpha, \beta_0) = \frac{1}{n} \sum_{i=1}^n I_X(x_{1i}) [x_{2i} - E_n(x_2|x_{1i}, \alpha)]' (v_i - E_n(v|x_{1i}, \alpha)). \tag{4.1}$$

By a Taylor expansion,

$$\sqrt{n}L_n(\hat{\alpha}, \beta_0) = \sqrt{n}L_n(\alpha_0, \beta_0) + \frac{\partial L_n(\bar{\alpha}, \beta_0)}{\partial \alpha'} \sqrt{n}(\hat{\alpha} - \alpha_0), \tag{4.2}$$

where  $\bar{\alpha}$  lies between  $\hat{\alpha}$  and  $\alpha_0$ .  $\partial L_n(\bar{\alpha}, \beta_0)/\partial \alpha'$  depends on the derivatives of the nonparametric function in (3.5). As shown in appendix 2, with the rate of convergence for the bandwidth sequence  $\{a_n\}$  in Assumption 4(3),  $\partial C_n(t|x_1, \bar{\alpha})/\partial \alpha$ , where  $t = 1, y_2, x_2$ , converges in probability uniformly in  $x_1 \in X$  to some well-defined limits, and

$$\frac{\partial L_n(\bar{\alpha}, \beta_0)}{\partial \alpha'} \xrightarrow{p} B, \tag{4.3}$$

where

$$B = E(I_X(x_{1i})\tau(-x_{1i}\alpha_0)[x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0)]' \times [x_{1i} - E(x_1|x_1\alpha_0 > x_{1i}\alpha_0)]), \tag{4.4}$$

and  $\tau(t) = (\partial/\partial t)E(v|u > t, \alpha_0)$ . The asymptotic distribution of  $\sqrt{n}L_n(\alpha_0, \beta_0)$  can be analyzed with Propositions 5, 6, and 7 in appendix 1. The details are in appendix 2.

<sup>7</sup>When  $u$  is either a uniform variate or an exponential variate,  $E(u|u > -x\alpha_0)$  will be a linear function of  $x\alpha_0$ . The uniform variate case is known in Olsen (1980) Professor Peter Schmidt has pointed out the exponential variate case to me.

It follows from appendix 2 that

$$\sqrt{n}L_n(\alpha_0, \beta_0) \stackrel{D}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Psi_1^{(1)}(r_i) + \Psi_1^{(2)}(r_i)), \tag{4.5}$$

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

$$\Psi_1^{(1)}(r_i) = I_X(x_{1i})(x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0))' \varepsilon_{2i} \tag{4.6}$$

and

$$\begin{aligned} \Psi_1^{(2)}(r_i) &= - E \left\{ E [I_X(x_{1j})(x_{2j} - E(x_2|x_1\alpha_0 > x_{1j}\alpha_0))' | x_{1j}\alpha_0] \right. \\ &\quad \times \frac{v_i - E(v|u > -x_{1j}\alpha_0)}{B(x_{1j}, \alpha_0, 0)} I(x_{1j}\alpha_0 < x_{1i}\alpha_0) \\ &\quad \left. \times I(-x_{1j}\alpha_0 < u_i) \middle| x_i, y_i \right\} \tag{4.7} \\ &= - \int_{-u_i}^{x_{1i}\alpha_0} E [I_X(x_{1j})(x_{2j} - E(x_2|x_1\alpha_0 > x_{1j}\alpha_0))' | x_{1j}\alpha_0 = z] \\ &\quad \times \frac{v_i - E(v|u > -z)}{\int_z^x h(t) dt} h(z) dz. \end{aligned}$$

The asymptotic distribution of  $\sqrt{n}L_n(\hat{\alpha}, \beta_0)$  depends on the joint distribution of  $\sqrt{n}L_n(\alpha_0, \beta_0)$  and  $\sqrt{n}(\hat{\alpha} - \alpha_0)$ . To complete the asymptotic distribution  $\hat{\beta}$ , one needs to be specific about the distribution of  $\sqrt{n}(\hat{\alpha} - \alpha_0)$ . As a specific example, consider the estimator  $\hat{\alpha}$  in Lee (1992). Under the regularity conditions in that article, it was shown that

$$\begin{aligned} \sqrt{n}(\hat{\alpha} - \alpha_0) &= C^{-1} \frac{1}{\sqrt{n(n-1)}} \\ &\quad \times \sum_{i=1}^n \sum_{j \neq i}^n I_X(x_{1i})(x_{1i} - E(x_1|x_1\alpha_0 > x_{1i}\alpha_0 + \Delta_n)) \\ &\quad \times \frac{\tau_1(x_{1i}, \alpha_0)}{C(1|x_{1i}, \alpha_0, \Delta_n)} \int_{x_{1i}\alpha_0 + \Delta_n}^{\infty} \int_{-x_{1i}\alpha_0}^{\infty} (u_i - u) \\ &\quad \times \frac{1}{a_n^2} K \left( \frac{u - u_j}{a_n}, \frac{z - x_{1j}\alpha_0}{a_n} \right) du dz + o_p(1), \tag{4.8} \end{aligned}$$

where

$$\begin{aligned} \tau_1(z) &= \frac{\partial}{\partial z} \{z + E(u|u > -z, \alpha_0)\} \\ &= 1 - \frac{f_u(-z)}{\int_{-z}^{\infty} f_u(t) dt} (z + E(u|u > -z, \alpha_0)) \end{aligned} \tag{4.9}$$

is the derivative of the regression function of  $y_1$  conditional on  $x$  and  $y_1 > 0$  with respect to  $x_1\alpha_0$ , and

$$\begin{aligned} C &= E\{I_X(x_{1i}) \tau_1^2(x_{1i}, \alpha_0) [x_{1i} - E(x_1|x_1\alpha_0 > x_{1i}\alpha_0)]'\} \\ &\quad \times [x_{1i} - E(x_1|x_1\alpha_0 > x_{1i}\alpha_0)]. \end{aligned} \tag{4.10}$$

With this  $\hat{\alpha}$ , appendix 2 proves that  $\hat{\beta}$  is asymptotically normal:

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, \Omega), \tag{4.11}$$

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 (3.10),  $B$  is the matrix in (4.4),  $C$  is defined in (4.10), and  $\Sigma = E(\Psi(r_i)\Psi'(r_i))$  with

$$\begin{aligned} \Psi(r_i) &= \\ I_X(x_{1i}) &\begin{pmatrix} (x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0))' \varepsilon_{2i} \\ (x_{1i} - E(x_1|x_1\alpha_0 > x_{1i}\alpha_0))' \tau_1(x_{1i}, \alpha_0) \varepsilon_{1i} \end{pmatrix} \\ &- \int_{-u_i}^{x_{1i}, \alpha_0} \left( E[I_X(x_{1j})(x_{2j} - E(x_2|x_1\alpha_0 > x_{1j}\alpha_0))' | x_{1j}\alpha_0 = z] (v_i - E(v|u > -z)) \right) \\ &\quad \left( E[I_X(x_{1j})(x_{1j} - E(x_1|x_1\alpha_0 > x_{1j}\alpha_0))' | x_{1j}\alpha_0 = z] (u_i - E(u|u > -z)) \right) \\ &\times \frac{h(z)}{\int_z^{\infty} h(t) dt} dz. \end{aligned} \tag{4.13}$$

There are some interesting similarities between the asymptotic distribution of our two-stage semiparametric estimator and the asymptotic distribution of a parametric two-stage estimator of this model. If the functional form of

$E(v(\beta)|u(\alpha) > -x_1\alpha, x_1\alpha > x_1\alpha)$  were known, a parametric two-stage estimator  $\hat{\beta}_p$  could be derived from

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

For any random variable  $s$  and a constant  $\Delta \geq 0$ , denote  $E_{\infty}(s|x_1, \alpha, \Delta) = C(s|x_1, \alpha, \Delta)/C(1|x_1, \alpha, \Delta)$  and  $E_{\infty}(s|x_1, \alpha) = E_{\infty}(s|x_1, \alpha, 0)$  for simplicity. Since  $E(v(\beta)|u(\alpha) > -x_1\alpha, x_1\alpha > x_1\alpha) = E_{\infty}(v(\beta)|x_{1i}, \alpha)$ ,

$$\begin{aligned} \hat{\beta}_p &= \left\{ \sum_{i=1}^n I_X(x_{1i})[x_{2i} - E_{\infty}(x_2|x_{1i}, \hat{\alpha})]' [x_{2i} - E_{\infty}(x_2|x_{1i}, \hat{\alpha})] \right\}^{-1} \\ &\quad \times \sum_{i=1}^n I_X(x_{1i})[x_{2i} - E_{\infty}(x_2|x_{1i}, \hat{\alpha})]' (y_{2i} - E_{\infty}(y_2|x_{1i}, \hat{\alpha})) \end{aligned} \tag{4.15}$$

implies

$$\begin{aligned} \hat{\beta}_p - \beta_0 &= \left\{ \sum_{i=1}^n I_X(x_{1i})[x_{2i} - E_{\infty}(x_2|x_{1i}, \hat{\alpha})]' [x_{2i} - E_{\infty}(x_2|x_{1i}, \hat{\alpha})] \right\}^{-1} \\ &\quad \times \sum_{i=1}^n I_X(x_{1i})[x_{2i} - E_{\infty}(x_2|x_{1i}, \hat{\alpha})]' (v_i - E_{\infty}(v|x_{1i}, \hat{\alpha})). \end{aligned} \tag{4.16}$$

By a Taylor series expansion, (4.16) implies that

$$\begin{aligned} \sqrt{n}(\hat{\beta}_p - \beta_0) &\stackrel{D}{=} A^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n I_X(x_{1i})[x_{2i} - E(x_2|x_1\alpha_0 > x_1\alpha_0)]' \varepsilon_{2i} \right. \\ &\quad \left. + B\sqrt{n}(\hat{\alpha} - \alpha_0) \right\}. \end{aligned} \tag{4.17}$$

For the semiparametric estimator  $\hat{\beta}$ ,

$$\sqrt{n}(\hat{\beta} - \beta_0) \stackrel{D}{=} A^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Psi_1^{(1)}(r_i) + \Psi_1^{(2)}(r_i)) + B\sqrt{n}(\hat{\alpha} - \alpha_0) \right\}, \tag{4.18}$$

from (3.9) and (4.1)–(4.5). Comparing (4.17) with (4.18), the difference is that an extra term, namely  $(1/\sqrt{n}) \sum_{i=1}^n \Psi_1^{(2)}(r_i)$ , appears in (4.18) for the semiparametric estimate. This extra term reflects the error introduced by replacing  $E(v(\beta)|u(\alpha) > -x_{1i}\alpha, x_{1i}\alpha > x_{1i}\hat{\alpha})$  with the nonparametric estimate  $E_n(v(\beta)|x_{1i}, \alpha)$  in the second-stage estimation.

**5. Covariance estimation**

The covariance matrix of the limiting distribution of  $\sqrt{n}(\hat{\beta} - \beta_0)$  is  $\Omega$  in (4.12). From (3.10),  $A$  can be consistently estimated by

$$\frac{1}{n} \sum_{i=1}^n I_X(x_{1i}) [x_{2i} - E_n(x_2 | x_{1i}, \hat{\alpha})]' [x_{2i} - E_n(x_2 | x_{1i}, \hat{\alpha})]. \tag{5.1}$$

The  $B$  can be consistently estimated by  $\partial L_n(\hat{\alpha}, \hat{\beta})/\partial \alpha'$ . As suggested in Lee (1992),  $C$  in (4.10) 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

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

is a nonparametric estimate of  $E(u|u > -x_i\alpha_0)$ . If  $\Psi(r_i)$  in (4.13) could be evaluated, an estimate of  $\Sigma$  would be the sample covariance matrix of  $\Psi(r_i)$ . As in appendix 2,  $\Psi(r_i)$  is the limit function of  $E(\Psi_n(r_i, r_j)|r_i) + E(\Psi_n(r_j, r_i)|r_i)$ , where  $\Psi_n(r_i, r_j) = (\Psi'_{n,1}(r_i, r_j), \Psi'_{n,2}(r_i, r_j))'$  in (A.17) and (A.23). This motivates the following estimate of  $\Sigma$ :

$$\begin{aligned} \hat{\Sigma}_n(\hat{\alpha}, \hat{\beta}) &= \frac{1}{n^{(3)}} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq i, j}^n [H_n(r_i, r_j, \hat{\alpha}, \hat{\beta}) + H_n(r_j, r_i, \hat{\alpha}, \hat{\beta})] \\ &\quad \times [H_n(r_i, r_k, \hat{\alpha}, \hat{\beta}) + H_n(r_k, r_i, \hat{\alpha}, \hat{\beta})]', \end{aligned} \tag{5.4}$$

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

$$H_n(r_i, r_j, \alpha, \beta) = (H'_{n,1}(r_i, r_j, \alpha, \beta), H'_{n,2}(r_i, r_j, \alpha, \beta))', \tag{5.5}$$



$$\begin{aligned}
 H_{n,1}(r_i, r_j, \alpha, \beta) &= I_X(x_{1i})(x_{2i} - E_n(x_2|x_{1i}, \hat{\alpha}))' \left\{ (v_i(\beta) - E_n(v(\beta)|x_{1i}, \alpha)) \right. \\
 &\quad \left. - (v_j(\beta) - E_n(v(\beta)|x_{1j}, \alpha)) \frac{1}{C_n(1|x_{1i}, \alpha)} \right. \\
 &\quad \left. \times J_1(r_i, r_j, \alpha, \Delta_n) \right\}, \tag{5.6}
 \end{aligned}$$

$$\begin{aligned}
 H_{n,2}(r_i, r_j, \alpha, \beta) &= I_X(x_{1i}) \left( x'_{1i} + \frac{\partial E_{n,1}(x_{1i}, \alpha)}{\partial \alpha} \right) \frac{1}{C_n(1|x_{1i}, \alpha)} \\
 &\quad \times J_2(r_i, r_j, \alpha, \Delta_n), \tag{5.7}
 \end{aligned}$$

$$J_1(r_i, r_j, \alpha, \Delta_n) = \int_{\lambda_{i,x} + \Delta_n}^{\alpha} \int_{\lambda_{i,x}}^{\alpha} \frac{1}{a_n^2} K\left(\frac{u - u_j(\alpha)}{a_n}, \frac{z - x_{1j}\alpha}{a_n}\right) du dz, \tag{5.8}$$

$$\begin{aligned}
 J_2(r_i, r_j, \alpha, \Delta_n) &= \int_{\lambda_{i,x} + \Delta_n}^{\alpha} \int_{\lambda_{i,x}}^{\alpha} (y_{1i} - x_{1i}\alpha - u) \frac{1}{a_n^2} \\
 &\quad \times K\left(\frac{u - u_j(\alpha)}{a_n}, \frac{z - x_{1j}\alpha}{a_n}\right) du dz. \tag{5.9}
 \end{aligned}$$

Define

$$\begin{aligned}
 \Sigma_n(\alpha, \beta) &= \frac{1}{n^{(3)}} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq i,j}^n [H(r_i, r_j, \alpha, \beta, a_n, \Delta_n) + H(r_j, r_i, \alpha, \beta, a_n, \Delta_n)] \\
 &\quad \times [H(r_i, r_k, \alpha, \beta, a_n, \Delta_n) + H(r_k, r_i, \alpha, \beta, a_n, \Delta_n)], \tag{5.10}
 \end{aligned}$$

where

$$H(r_i, r_j, \alpha, \beta, a_n, \Delta_n) = (H'_1(r_i, r_j, \alpha, \beta, a_n, \Delta_n), H'_2(r_i, r_j, \alpha, \beta, a_n, \Delta_n))', \tag{5.11}$$

$$\begin{aligned}
 H_1(r_i, r_j, \alpha, \beta, a_n, \Delta_n) &= I_X(x_{1i})(x_{2i} - E_\alpha(x_2|x_{1i}, \alpha, \Delta_n))' \\
 &\quad \times \left\{ (v_i(\beta) - E_\alpha(v(\beta)|x_{1i}, \alpha, \Delta_n)) \right. \\
 &\quad \left. - (v_j(\beta) - E_\alpha(v(\beta)|x_{1j}, \alpha, \Delta_n)) \right. \\
 &\quad \left. \times \frac{1}{C(1|x_{1i}, \alpha, \Delta_n)} J_1(r_i, r_j, \alpha, \Delta_n) \right\}, \tag{5.12}
 \end{aligned}$$

$$\begin{aligned}
 H_2(r_i, r_j, \alpha, \beta, a_n, \Delta_n) &= I_X(x_{1i}) \left( x'_{1i} + \frac{\partial E(x_{1i}, \alpha)}{\partial \alpha} \right) \\
 &\quad \times \frac{1}{C(1|x_{1i}, \alpha, \Delta_n)} J_2(r_i, r_j, \alpha, \Delta_n), \tag{5.13}
 \end{aligned}$$

and  $E(x_{1i}, \alpha)$  is the limit function of  $E_{n,1}(x_{1i}, \alpha)$ . The consistency of  $\hat{\Sigma}_n(\hat{\alpha}, \hat{\beta})$  will follow by showing that  $\Sigma_n(\hat{\alpha}, \hat{\beta})$  converges in probability to  $\Sigma$  and  $\hat{\Sigma}_n(\alpha, \beta) - \Sigma_n(\alpha, \beta)$  converges in probability to zero uniformly in  $(\alpha, \beta)$  in a neighborhood of  $(\alpha_0, \beta_0)$ .

By a change of variables in the integral,

$$J_1(r_i, r_j, \alpha, \Delta_n) = \int_{(x_{1i}\alpha + \Delta_n - x_{1j}\alpha)/a_n}^{\infty} \int_{(-x_{1j}\alpha - u_j(\alpha))/a_n}^{\infty} K(u, z) \, du \, dz \tag{5.14}$$

and

$$\begin{aligned}
 J_2(r_i, r_j, \alpha, \Delta_n) &= \int_{(x_{1i}\alpha + \Delta_n - x_{1j}\alpha)/a_n}^{\infty} \int_{(-x_{1j}\alpha - u_j(\alpha))/a_n}^{\infty} (u_i(\alpha) - u_j(\alpha) - a_n u) \\
 &\quad \times K(u, z) \, du \, dz. \tag{5.15}
 \end{aligned}$$

The uniform law of large number for  $U$  statistics in Proposition 1 can be applied to  $\Sigma_n(\alpha, \beta)$  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  $(\alpha_0, \beta_0)$ . For any sequence  $(\alpha_n, \beta_n)$  converging to  $(\alpha_0, \beta_0)$ , with similar arguments for the proofs of (A.19), (A.21) (A.24), and (A.25),

$$\lim_{n \rightarrow \infty} E[H(r_i, r_j, \alpha_n, \beta_n, a_n, \Delta_n) + H(r_i, r_j, \alpha_n, \beta_n, a_n, \Delta_n)|r_i] = \Psi(r_i).$$

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 in  $H_n(\cdot)$  have converged in probability uniformly in  $(x_{1i}, \alpha, \beta)$  to their limit functions in  $H(\cdot)$ .

### 6. Monte Carlo simulation

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

Simulated data are generated from the following latent equations:

$$y_1 = \alpha_1 s_1 + \alpha_2 s_2 + \sigma_1 \bar{u} \quad (6.1)$$

and

$$y_2 = \beta_1 s_1 + \beta_2 s_2 + \sigma_2 \bar{v}. \quad (6.2)$$

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 distributions of  $\bar{u}$  and  $\bar{v}$ . Data on  $\bar{u}$  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 standardized gamma random variate with zero mean and unit variance of which the density function is  $f_G(\varepsilon) = \frac{8}{3}(\varepsilon + 2)^3 \exp[-2(\varepsilon + 2)]$ ,  $\varepsilon > -2$ , with its mode at  $-\frac{1}{2}$ . The disturbance  $\bar{v}$  is generated from  $\bar{v} = \sqrt{0.25}\bar{u} + \sqrt{0.75}\eta$ , where  $\eta$  is a  $N(0, 1)$  random variable independent of  $\bar{u}$ . The distribution of  $\bar{v}$  is the convolution of the distributions of  $\sqrt{0.25}\bar{u}$  and  $\sqrt{0.75}\eta$ . The correlation coefficient of  $\bar{u}$  and  $\bar{v}$  is 0.5. The variances of  $\bar{u}$  and  $\bar{v}$  are both unity. However, variances of equation disturbances can be controlled by selecting values for scale parameters. The scale parameters  $\sigma_1$  and  $\sigma_2$  are set to 1.5 which imply that the  $R^2$  values for both latent eqs. (6.1) and (6.2) are 0.5. The correlation coefficient of the two equations' disturbances remains to be 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 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% truncation.

The bivariate kernel function used for our estimation is the product of two univariate biweight kernel density functions, i.e.,  $K(t_1, t_2) = K_1(t_1)K_1(t_2)$ , where

$$K_1(t) = \begin{cases} \frac{15}{16}(1-t^2)^2 & \text{for } |t| < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (6.3)$$

This density has a bounded support and is continuously differentiable. In addition to its smooth character, this kernel  $K$  is chosen for its computational efficiency.<sup>8</sup>

<sup>8</sup>The nonparametric function in (3.1) involves double integrals. When the bivariate kernel function is a product of two univariate kernel functions, the double integrals become the product of two univariate integrals. With  $K_1$  in (6.3), the univariate integral has a simple closed form expression which is a polynomial function. This bivariate kernel function is not the unique choice. Many other kernel functions will also be useful. It is known in the density estimation literature [see, e.g., Silverman (1986)] that different kernel functions have only some minor differences in terms of efficiency in density estimation.

Table 1.  
Results of 300 replications with various sample sizes, bandwidth factor  $c = 1$

N		Normal			Gamma*Normal			- Gamma*Normal		
		Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
30	$\beta_1$	0.843	0.746	0.761	0.834	0.576	0.599	0.942	0.544	0.546
	$\beta_2$	0.938	0.586	0.588	1.037	0.539	0.539	0.979	0.548	0.547
50	$\beta_1$	0.919	0.422	0.429	0.906	0.516	0.524	0.928	0.381	0.387
	$\beta_2$	0.986	0.427	0.427	0.996	0.451	0.450	0.956	0.358	0.360
100	$\beta_1$	0.992	0.269	0.268	0.948	0.294	0.297	0.970	0.223	0.224
	$\beta_2$	0.986	0.248	0.247	1.017	0.275	0.274	0.983	0.222	0.222
200	$\beta_1$	1.005	0.168	0.168	0.997	0.224	0.223	0.994	0.156	0.156
	$\beta_2$	0.995	0.159	0.159	0.996	0.203	0.203	1.004	0.139	0.139

Any bandwidth sequence of the form

$$a_n = c/n^p, \quad (6.4)$$

with  $\frac{1}{4} < p < \frac{1}{3}$  and  $c$  being a constant factor independent of the sample size, will satisfy the rate requirement that  $na_n^3 \rightarrow \infty$  and  $na_n^4 \rightarrow 0$ . For our experiments,  $p = 0.3$  is set. However, we experiment with different values of  $c$  to investigate the sensitivity of our estimator to the chosen bandwidth parameter. The parameter  $\Delta_n$  is set to  $0.1a_n^{0.99}$ . For our semiparametric estimation, the regressors are trimmed whenever  $|x_1| > 1.9$  or  $|x_2| > 1.8$  to satisfy Assumption 1, which implies that approximately 15% of the data will be trimmed.

For each case, 300 data sets with the same sample size are generated. All the summary statistics reported below for each case are based on 300 estimates. First-stage estimates of the truncated regression function are in Lee (1992). Both the first-stage and the second-stage semiparametric methods use the same kernel function and the same bandwidth  $a_n$ .

Table 1 reports simulation results of the two-stage semiparametric estimation of (6.2) with various sample sizes and distributions. The  $c$  in (6.4) is 1. The true  $\beta_1$  and  $\beta_2$  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).<sup>10</sup> The biases of the estimates can be derived by comparing their mean values with the true parameters. There are some small sample biases. For the sample size 30, the largest bias is about 0.164. The biases tend to decrease as the

<sup>9</sup>In Lee (1992) for the estimation of  $\alpha$  in the truncated regression model, we have experimented with different values of  $d$  in  $\Delta_n = da_n^{0.9}$  in a Monte Carlo study. The semiparametric estimates of  $\alpha$  are not sensitive to the values of  $\Delta_n$ . For details, see table 1 in Lee (1992).

<sup>10</sup>The standard deviation is derived as the square root of the bias-adjusted sample variance. The variance component in RMSE is not bias-adjusted

sample size increases. For the sample size 200, the largest bias is about 0.006. 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 the mixed negative gamma-normal distribution followed by the model with the normal distribution. 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 the negative gamma-normal distribution may be related to the fact that such a distribution has a thinner upper tail and a smaller variance after selection than the other two cases.

Table 2 reports simulation results derived with six different bandwidth factors. The  $c$  varies from 0.10 to 4.00. This range seems to be quite wide. The sample size is 100. The two-stage estimates do not seem to be very sensitive to the different bandwidths. The biases are small. The standard deviations and the RMSEs are similar in magnitude. 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 the residual sum of squares (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 reasonably small. The variances and RMSEs are even slightly less than most of the variances and RMSEs of the estimates based on fixed bandwidth factors. These two strategies seem good for our estimation procedure.

In table 3, we compare our semiparametric two-stage estimates with ordinary least square (OLS) estimates and some parametric estimates of (6.2). The OLS procedure ignores the sample selection bias and is inconsistent. For all the sample sizes and the distributions considered, the OLS estimates of  $\beta_1$  are biased downward and the OLS estimates of  $\beta_2$  are biased upward. 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 the sample size increases. Comparing the OLS estimates with our semiparametric two-stage estimates, the OLS estimates have smaller SDs for all the cases. However, for sample sizes 100 and 200, the biases dominate the SDs, which results in larger RMSEs than the RMSEs of the semiparametric estimates.<sup>11</sup>

<sup>11</sup>The biases of the OLS estimates would be severer, as the sample selection bias becomes harsher, i.e., the correlation of  $u$  and  $v$  becomes larger. This can be seen from the Monte Carlo simulation results for the truncated regression model in Lee (1992), which corresponds essentially to the case that  $u$  and  $v$  are perfectly correlated. For the latter case, the semiparametric estimates (the first-stage estimates) can even have smaller RMSEs than the OLS estimates [see table 3 in Lee (1992)] for the sample size 50

Table 2  
Results with various bandwidth factors; sample size 100.

Factor	Normal					Gamma*Normal					- Gamma*Normal				
	Mean	SD	RMSE	RSS		Mean	SD	RMSE	RSS		Mean	SD	RMSE	RSS	
0.10	$\beta_1$	0.973	0.282	0.282	173.865	0.945	0.296	0.300	181.635		0.959	0.254	0.256	170.393	
	$\beta_2$	0.990	0.231	0.230		1.035	0.256	0.257			0.974	0.219	0.219		
0.25	$\beta_1$	0.989	0.268	0.267	175.482	0.958	0.293	0.295	180.716		0.955	0.295	0.297	174.063	
	$\beta_2$	0.975	0.285	0.285		1.014	0.279	0.278			0.974	0.265	0.265		
0.50	$\beta_1$	0.997	0.271	0.270	172.786	0.952	0.283	0.286	179.962		0.956	0.263	0.265	169.427	
	$\beta_2$	0.971	0.273	0.273		1.019	0.261	0.260			0.988	0.221	0.220		
1.00	$\beta_1$	0.992	0.269	0.268	171.084	0.948	0.294	0.297	178.625		0.970	0.223	0.224	167.555	
	$\beta_2$	0.986	0.248	0.247		1.017	0.275	0.274			0.983	0.222	0.222		
2.00	$\beta_1$	0.994	0.253	0.252	168.173	0.974	0.300	0.300	176.819		0.978	0.216	0.216	164.056	
	$\beta_2$	0.985	0.254	0.253		1.004	0.279	0.278			0.993	0.193	0.192		
4.00	$\beta_1$	1.023	0.266	0.266	167.573	1.010	0.308	0.307	176.423		0.994	0.217	0.216	163.332	
	$\beta_2$	0.959	0.267	0.269		0.972	0.300	0.300			0.978	0.203	0.203		
Min	$\beta_1$	1.000	0.242	0.241	166.206	0.965	0.290	0.291	174.743		0.986	0.213	0.212	162.485	
	$\beta_2$	0.991	0.241	0.240		1.019	0.245	0.245			0.994	0.193	0.192		
Ave	$\beta_1$	0.995	0.239	0.238		0.965	0.272	0.273			0.969	0.214	0.215		
	$\beta_2$	0.978	0.232	0.232		1.010	0.249	0.248			0.982	0.199	0.199		

In table 3, we report also some parametric two-stage estimates based on a normal distribution formulation. The first-stage estimate of  $\alpha$  in (6.1) is derived from the following nonlinear least squares procedure:

$$\min_{\alpha, \sigma_1} \sum_{i=1}^n \left( y_{1i} - s_i \alpha - \sigma_1 \frac{\phi(s_i \alpha / \sigma_1)}{\Phi(s_i \alpha / \sigma_1)} \right)^2, \quad (6.5)$$

where  $\phi$  and  $\Phi$  are, respectively, the standard normal density and distribution functions. With the first-stage estimates  $\hat{\alpha}$  and  $\hat{\sigma}_1$  from (6.5), the second-stage estimate of  $\beta$  is derived from

$$\min_{\beta, \sigma_{12}} \sum_{i=1}^n \left( y_{2i} - s_i \beta - \sigma_{12} \frac{\phi(s_i \hat{\alpha} / \hat{\sigma}_1)}{\Phi(s_i \hat{\alpha} / \hat{\sigma}_1)} \right)^2. \quad (6.6)$$

This parametric two-stage procedure is similar to Heckman's procedure [Heckman (1976)]. The parametric estimates of  $\beta$  are reported in table 3 with the legend 'PN-OLS'.<sup>12</sup> The estimates reported with the legend 'PN-ROLS' differ from the estimates 'PN-OLS' in that the intercept term of (6.2) is known and restricted to be zero. For the normal distribution model, this parametric two-stage procedure provides consistent estimates. However, for the mixed gamma-normal and mixed negative gamma-normal distributions, this procedure is, in general, inconsistent. The PN-OLS estimates do not perform well for our simulated models.<sup>13</sup> The variances and RMSEs for these estimates are larger than those of the semiparametric estimates. The restricted PN-ROLS estimates perform much better. The PN-OLS estimates apparently suffer from the problem of multicollinearity.<sup>14</sup> By knowing the intercept term to be zero, this exclusion restriction reduces the severity of multicollinearity for the parametric two-stage estimation. The PN-ROLS estimates have smaller variances than the semiparametric estimates in all cases. There is some evidence that for the misspecified distributions, the parametric two-stage estimates are biased.<sup>15</sup> The biases of the estimates of  $\beta_2$  of the mixed distributions models are larger

<sup>12</sup>Some of the first-stage parametric estimates of  $\alpha$  are reported in Lee (1992).

<sup>13</sup>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.

<sup>14</sup>The parametric two-stage method has not utilized the additional structure (2.12). It suffers from the multicollinearity issue mentioned in section 3.

<sup>15</sup>The estimates of the coefficient of the sample selection bias adjustment term are also biased. While the true coefficient implied by our data-generating process is 0.75, the estimates of the mixed gamma-normal model and the mixed negative gamma-normal model are about 0.83 and 0.675, respectively.

Table 3  
OLS, parametric normal two-stage, and semiparametric estimates.

N	OLS				PN-ROLS				Semiparametric			
	Mean	SD	RMSE		Mean	SD	RMSE		Mean	SD	RMSE	
30	$\beta_1$	0.791	0.317	0.379	1.050	1.238	1.237	1.001	0.328	0.327	0.746	0.761
	$\beta_2$	1.237	0.261	0.352	0.991	1.342	1.340	1.009	0.277	0.277	0.588	0.588
50	$\beta_1$	1.242	0.250	0.327	0.967	0.826	0.825	1.006	0.262	0.262	0.422	0.429
	$\beta_2$	0.788	0.163	0.267	1.033	0.488	0.487	1.014	0.183	0.183	0.269	0.268
100	$\beta_1$	1.242	0.137	0.278	0.984	0.456	0.454	1.001	0.149	0.148	0.248	0.247
	$\beta_2$	0.779	0.111	0.247	1.002	0.337	0.336	1.009	0.121	0.121	0.168	0.168
200	$\beta_1$	1.251	0.098	0.269	1.012	0.327	0.326	1.007	0.102	0.102	0.159	0.159
Gamma*Normal												
30	$\beta_1$	0.697	0.329	0.447	1.231	3.562	3.564	0.900	0.390	0.402	0.576	0.599
	$\beta_2$	1.316	0.272	0.417	1.362	2.814	2.833	1.108	0.306	0.324	0.539	0.539
50	$\beta_1$	0.706	0.229	0.371	0.963	1.758	1.755	0.934	0.256	0.262	0.516	0.524
	$\beta_2$	1.322	0.202	0.379	1.047	1.267	1.266	1.092	0.220	0.236	0.451	0.450
100	$\beta_1$	0.720	0.158	0.321	1.024	0.543	0.543	0.953	0.179	0.184	0.294	0.297
	$\beta_2$	1.312	0.141	0.342	1.009	0.582	0.581	1.079	0.156	0.174	0.275	0.274
200	$\beta_1$	0.724	0.110	0.297	1.033	0.347	0.348	0.958	0.124	0.131	0.224	0.223
	$\beta_2$	1.302	0.098	0.317	0.980	0.365	0.365	1.063	0.111	0.127	0.203	0.203
- Gamma*Normal												
30	$\beta_1$	0.796	0.301	0.363	0.988	1.004	1.002	1.017	0.310	0.310	0.544	0.546
	$\beta_2$	1.210	0.286	0.354	1.008	1.040	1.038	0.964	0.279	0.281	0.548	0.547
50	$\beta_1$	0.810	0.223	0.291	0.933	0.827	0.828	1.026	0.227	0.228	0.381	0.387
	$\beta_2$	1.189	0.206	0.278	1.043	0.746	0.746	0.947	0.196	0.203	0.358	0.360
100	$\beta_1$	0.800	0.158	0.254	0.965	0.421	0.420	1.020	0.154	0.155	0.223	0.224
	$\beta_2$	1.199	0.138	0.242	1.011	0.433	0.431	0.954	0.136	0.143	0.222	0.222
200	$\beta_1$	0.807	0.118	0.226	0.969	0.257	0.258	1.027	0.113	0.116	0.156	0.156
	$\beta_2$	1.203	0.102	0.227	1.021	0.282	0.282	0.959	0.097	0.105	0.139	0.139



than the biases of the semiparametric estimates. For the sample size 200, the biases of the estimates of  $\beta_1$  are also larger than the biases of the semiparametric estimates. These biases are, however, not very large; hence, in terms of RMSE, the parametric PN-ROLS estimates still perform better for all the cases considered.<sup>16</sup>

As we have pointed out before, sample selection models are index models. Since semiparametric methods based on index restrictions have broad applicability, it is worthwhile to compare our semiparametric estimates with semiparametric estimators derived with index restrictions. Identification in index models requires exclusion restrictions of explanatory variables on the outcome eq. (6.2). For comparison, simulated data are generated from a model with (6.2) replaced by

$$y_2^* = \beta_1 s_1 + \sigma_2 \bar{v}. \tag{6.2'}$$

The  $\beta_1$  can be identified with the index formulation. A semiparametric two-stage estimator of  $\beta_1$  based on the index restriction [see Powell (1987) and Ichimura and Lee (1991)] is

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

the weight function  $W_n^*(s_i, s_j, \alpha)$  is

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

<sup>16</sup>We should point out that the PN-ROLS is not a practical approach in practice. In this Monte Carlo design, the true intercept term is zero, so there is no omitted variable problem. In a practical application, there are usually no good reasons to impose zero restriction on the intercept term. For the semiparametric estimation, the intercept term in (2.2) is absorbed in the disturbance  $v$  and has implicitly been captured in the nonparametric estimates of the sample selection bias term. The introduction of the PN-ROLS is intended to demonstrate that the extremely poor performance of the PN-OLS is indeed due to multicollinearity but not due to programming errors.

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

$$K_4^*(t) = 2K^*(t) - \frac{1}{\sqrt{2}} K^*\left(\frac{t}{\sqrt{2}}\right), \quad (6.9)$$

with

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

and

$$b_n = c/n^{1/5}. \quad (6.11)$$

The function  $K^*(t)$  is a proper density function, which is twice continuously differentiable with a 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 a high-order kernel function from a density function is suggested in Bierens (1987). This kernel function and its bandwidth rate (6.11) satisfy the regularity conditions in Ichimura and Lee (1991). The simulations are based on the sample size 200. The estimates from (6.7) and our estimates (2.8) for the model (6.2') are presented in table 4. Various values of  $c$  in (6.11) have been tried. The semiparametric estimates based on the index formulation are relatively sensitive to the bandwidth factors. The magnitudes of the estimates as well as their variances decrease as the values of  $c$  in (6.11) increase from 1.0 to 20.0. The biases are the smallest around the factor value of 5.0 or 7.5. The corresponding RMSEs decrease but, eventually, increase as the bandwidth factor increases. Comparing these estimates with our semiparametric estimates from (2.8), the latter estimates have smaller biases and smaller RMSEs. Taking into account the independence property, our proposed semiparametric estimates seem likely to be more efficient than the estimates that have utilized only the index restriction.<sup>17</sup>

## Appendix 1

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

<sup>17</sup>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 4  
Index model and semiparametric estimation: sample size 200

Factor	Normal					Gamma*Normal					- Gamma*Normal					
	Mean	SD	RMSE	RSS	Mean	SD	RMSE	RSS	Mean	SD	RMSE	RSS	Mean	SD	RMSE	RSS
1.0 $\beta_1$	1.028	0.451	0.451	386 789	1.123	1.499	1.500	578 010	1.033	0.437	0.437	372 099	1.033	0.437	0.437	372 099
2.5 $\beta_1$	1.035	0.272	0.274	355 892	1.031	0.348	0.349	364 874	1.004	0.235	0.234	339 289	1.004	0.235	0.234	339 289
5.0 $\beta_1$	1.002	0.330	0.329	345 618	1.012	0.348	0.347	358 918	0.949	0.486	0.487	351 707	0.949	0.486	0.487	351 707
7.5 $\beta_1$	0.983	0.249	0.249	340 055	0.985	0.312	0.312	356 015	0.985	0.258	0.258	333 816	0.985	0.258	0.258	333 816
10.0 $\beta_1$	0.962	0.199	0.202	338 433	0.937	0.272	0.279	356 600	0.960	0.194	0.198	327 917	0.960	0.194	0.198	327 917
15.0 $\beta_1$	0.899	0.168	0.196	340 395	0.874	0.196	0.233	358 676	0.912	0.157	0.180	328 812	0.912	0.157	0.180	328 812
20.0 $\beta_1$	0.859	0.152	0.207	341 973	0.827	0.156	0.233	360 451	0.881	0.150	0.191	329 832	0.881	0.150	0.191	329 832
Min $\beta_1$	1.005	0.236	0.235	335 174	1.013	0.335	0.334	351 045	0.984	0.192	0.192	325 182	0.984	0.192	0.192	325 182
Tobit selection model																
1.0 $\beta_1$	1.001	0.168	0.168	340 513	0.986	0.214	0.214	357 208	0.993	0.153	0.153	329 468	0.993	0.153	0.153	329 468

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

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

where  $a_n = O(1/n^p)$ ,  $p > 0$ ,  $d \geq 0$ ,  $\alpha \in B$ , and  $s(y_{i_1}, \dots, y_{i_l}, \alpha)$  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  $t(y_1, \dots, y_l)$  is bounded by a finite-order (say, order  $\delta$ ) polynomial of  $y_1, \dots, y_l$ .
- (3) The first  $\delta \cdot r$  moments of  $y$  exist, where  $r \geq 2$ .
- (4) The function  $h(\cdot)$  is a bounded function.
- (5)  $E \left\{ t^2(y_1, \dots, y_l) h^2 \left[ y_1, \dots, y_l, \frac{s(y_1, \dots, y_l, \alpha)}{a_n} \right] \right\} = O(a_n^{\bar{d}})$  uniformly in  $\alpha \in B$ , where  $\bar{d} \leq d$ .
- (6) The functions  $h(y_1, \dots, y_l, a_n, s)$  and  $s(y_1, \dots, y_l, \alpha)$  satisfy the bounded Lipschitzian condition of order 1 with respect to  $\alpha$  and  $s$ , uniformly in  $y_1, \dots, y_l$ .

If  $\lim_{n \rightarrow \infty} (n/\ln n) a_n^{2(1+\delta/r)d-\bar{d}} = \infty$ , then  $(1/n^{(l)}) \sum_{n,l} \{g(y_{i_1}, \dots, y_{i_l}, a_n, \alpha) - E(g(y_1, \dots, y_l, a_n, \alpha))\} \xrightarrow{p} 0$ , 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)  $E(g(y_{i_1}, \dots, y_{i_l}, a_n, \alpha))$  converges to a limit function  $g^*(\alpha)$  uniformly in  $\alpha \in B$ , then

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

uniformly in  $\alpha \in B$ .

*Proof.* This is a uniform law of large number with a 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 (1991).

*Proposition 2.* Let  $K(w)$  be a function with zero mean and a bounded support. Let  $g(w|\alpha)$  denote the density function of  $(u(\alpha), x_1\alpha)$  with a support  $W = \{w|w_1 + w_2 > 0, \text{ where } w = (w_1, w_2)\}$ .

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

$$\sup_{\alpha} \sup_{w \in N_{\delta}(v)} \left| \frac{\partial^2}{\partial w \partial w'} [E(c(y, x)|(u(x), x_1\alpha) = w)g(w|\alpha)] \right| \leq h(v),$$

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

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

then

$$\begin{aligned} & \sup_{(x_{1i}, \alpha)} \left| \int_{x_{1i}\alpha + \Delta_n}^{\infty} \int_{-x_{1i}\alpha}^{\infty} E \left[ c(y, x) \frac{1}{a_n^2} K \left( \frac{u - u(\alpha)}{a_n}, \frac{z - x_{1i}\alpha}{a_n} \right) \right] du dz \right. \\ & \left. - \int_{x_{1i}\alpha + \Delta_n}^{\infty} \int_{-x_{1i}\alpha}^{\infty} E(c(y, x)|u(\alpha) = u, x_1\alpha = z)g(u, z|\alpha) du dz \right| = O(a_n^2), \end{aligned}$$

where  $\{\Delta_n\}$  is a sequence of positive numbers such that  $\lim_{n \rightarrow \infty} \Delta_n = 0$  and  $\lim_{n \rightarrow \infty} (\Delta_n/a_n) = \infty$ .

*Proof.* This is lemma 1 in Lee (1992).

*Proposition 3.* Let  $K(w)$  be a function with zero mean and a bounded support, and let  $g(w|\alpha)$  denote the density function of  $(u(\alpha), x_1\alpha)$  as in Proposition 2. Suppose that there exists a Lebesgue measurable function  $h(v)$  such that

$$\sup_{\alpha} \sup_{w \in N_{\delta}(v)} \left| \frac{\partial^2}{\partial w \partial w'} [E(c(y, x)|(u(\alpha), x_1\alpha) = w)g(w|\alpha)] \right| \leq h(v),$$

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

(i) If  $\int_{-\infty}^{\infty} \sup_2 h(u, z) du < \infty$ , then

$$\begin{aligned} & \sup_{(x_{1i}, \alpha)} \left| \int_{-x_{1i}\alpha}^{\infty} E \left[ c(y, x) \frac{1}{a_n^2} K \left( \frac{u - u(\alpha)}{a_n}, \frac{x_{1i}\alpha + \Delta_n - x_{1i}\alpha}{a_n} \right) \right] du \right. \\ & \left. - \int_{-x_{1i}\alpha}^{\infty} E \left[ c(y, x) | u(\alpha) = u, x_1\alpha = x_{1i}\alpha + \Delta_n \right] \right. \\ & \left. \times g(u, x_{1i}\alpha + \Delta_n|\alpha) du \right| = O(a_n^2). \end{aligned}$$

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

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

*Proof.* This is lemma 2 in Lee (1992).

*Proposition 4.* Let  $F(w)$  be a bounded function with a bounded support, and let  $g(w|\alpha)$  denote the density function of  $(u(\alpha), x_1\alpha)$  as in Proposition 2. Suppose that there exists a Lebesgue measurable function  $h(v)$  such that

$$\sup_{\alpha} \sup_{w \in N_{\delta}(v)} E[|c(y, x)| | (u(\alpha), x_1\alpha) = w] g(w|\alpha) \leq h(v),$$

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

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

$$\sup_{x_{1i}, \alpha} E \left( |c(y, x)| \left\{ \int_{(x_{1i}\alpha + \Delta_n - x_1\alpha)/a_n}^{\infty} F \left[ \frac{-x_{1i}\alpha - u(\alpha)}{a_n}, z \right] dz \right\}^2 \right) = O(a_n).$$

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

$$\sup_{x_{1i}, \alpha} E \left( |c(y, x)| \left\{ \int_{(-x_{1i}\alpha - u(\alpha))/a_n}^{\infty} F \left[ u, \frac{x_{1i}\alpha + \Delta_n - x_1\alpha}{a_n} \right] du \right\}^2 \right) = O(a_n).$$

*Proof.* This proposition is lemma 3 in Lee (1992).

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

- (1)  $E|h(z)| < \infty$ ,
- (2)  $\sup_{z_i} |E(f_n(z_1, \dots, z_n; z_i) | z_i) - \bar{f}_n(z_i)| = O(a_n^{\delta_1})$ ,
- (3)  $\sup_{z_i} |E(g_n(z_1, \dots, z_n; z_i) | z_i) - \bar{g}_n(z_i)| = O(a_n^{\delta_2})$ ,
- (4)  $\sup_{z_i} \text{var}(f_n(z_1, \dots, z_n; z_i) | z_i) = O(1/na_n^{\tau_1})$ , and
- (5)  $\sup_{z_i} \text{var}(g_n(z_1, \dots, z_n; z_i) | z_i) = O(1/na_n^{\tau_2})$ .

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

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

*Proof.* This is lemma 6 in Lee (1992) with some slight generalization. The result follows from the Markov and Cauchy inequalities.

*Proposition 6.* Let  $\{z_i\}$  be a sequence of i.i.d. random variables and  $\Phi_n(z_1, z_2, a_n)$  be a sequence of measurable functions with a 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  $\text{var}(\Phi_n(z_1, z_2, a_n)) = O(1/a_n^r)$ ,
- (2) there exist squared integrable functions  $h_j(z), j = 1, 2$ , such that

$$|E(\Phi_n(z_1, z_2, a_n)|z_1)| \leq h_1(z_1) \quad \text{and} \quad |E(\Phi_n(z_2, z_1, a_n)|z_1)| \leq h_2(z_1),$$

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

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

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

*Proof.* This is proposition 6 in Lee (1989).

*Proposition 7.* Let  $\{z_i\}$  be a sequence of i.i.d. random variables and  $\Phi_n(z_1, z_2, a_n)$  a sequence of measurable functions with a 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  $\text{var}(\Phi_n(z_1, z_2, a_n)) = O(1/a_n^r)$ ,
- (2)  $\lim_{n \rightarrow \infty} E(\Phi_n(z_1, z_2, a_n)|z_1) = f_1(z_1)$  and  $\lim_{n \rightarrow \infty} E(\Phi_n(z_1, z_2, a_n)|z_2) = f_2(z_2)$ , a.e., for some measurable functions  $f_1(z)$  and  $f_2(z)$ , and

$$\begin{aligned} (3) \quad \lim_{n \rightarrow \infty} E\{ [E(\Phi_n(z_1, z_2, a_n)|z_1) + E(\Phi_n(z_2, z_1, a_n)|z_1)] \\ \times [E(\Phi_n(z_1, z_2, a_n)|z_1) + E(\Phi_n(z_2, z_1, a_n)|z_1)]' \} \\ = E\{ [f_1(z_1) + f_2(z_1)] [f_1(z_1) + f_2(z_1)]' \} \\ = \Sigma, \end{aligned}$$

where  $\Sigma$  is a finite matrix.

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

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

*Proof.* This is lemma 8 in Lee (1992).

**Appendix 2: Proofs of asymptotic distribution**

From (4.1),

$$\begin{aligned} & \frac{\partial L_n(\alpha, \beta)}{\partial \alpha'} \\ &= -\frac{1}{n} \sum_{i=1}^n I_X(x_{1i}) \frac{1}{C_n(1|x_{1i}, \alpha)} [x_{2i} - E_n(x_2|x_{1i}, \alpha)]' \\ & \quad \times \left[ \frac{\partial C_n(y_2|x_{1i}, \alpha)}{\partial \alpha'} - \beta' \frac{\partial C_n(x'_2|x_{1i}, \alpha)}{\partial \alpha'} - E_n(v(\beta)|x_{1i}, \alpha) \frac{\partial C_n(1|x_{1i}, \alpha)}{\partial \alpha'} \right] \\ & \quad - \frac{1}{n} \sum_{i=1}^n I_X(x_{1i}) (v_i(\beta) - E_n(v(\beta)|x_{1i}, \alpha)) \frac{1}{C_n(1|x_{1i}, \alpha)} \\ & \quad \times \left[ \frac{\partial C_n(x'_2|x_{1i}, \alpha)}{\partial \alpha'} - E_n(x'_2|x_{1i}, \alpha) \frac{\partial C_n(1|x_{1i}, \alpha)}{\partial \alpha'} \right], \end{aligned} \tag{A.1}$$

where

$$\begin{aligned} & \frac{\partial C_n(r_2|x_{1i}, \alpha)}{\partial \alpha} \\ &= \frac{1}{n-1} \sum_{j \neq i}^n (x_{1i} - x_{1j})' r_{2j} \int_{(x_{1i}\alpha + \Delta_n - x_{1j}\alpha)/a_n}^{\infty} \frac{1}{a_n} K\left(\frac{-x_{1i}\alpha - u_j(\alpha)}{a_n}, z\right) dz \\ & \quad - \frac{1}{n-1} \sum_{j \neq i}^n (x_{1i} - x_{1j})' r_{2j} \int_{(-x_{1i}\alpha - u_j(\alpha))/a_n}^{\infty} \frac{1}{a_n} K\left(u, \frac{x_{1i}\alpha + \Delta_n - x_{1j}\alpha}{a_n}\right) du \\ &= \int_{x_{1i}\alpha + \Delta_n}^{\infty} \frac{1}{n-1} \sum_{j \neq i}^n (x_{1i} - x_{1j})' r_{2j} \frac{1}{a_n^2} K\left(\frac{-x_{1i}\alpha + u_j(\alpha)}{a_n}, \frac{z - x_{1j}\alpha}{a_n}\right) dz \\ & \quad - \int_{-x_{1i}\alpha}^{\infty} \frac{1}{n-1} \sum_{j \neq i}^n (x_{1i} - x_{1j})' r_{2j} \frac{1}{a_n^2} K\left(\frac{u - u_j(\alpha)}{a_n}, \frac{x_{1i}\alpha + \Delta_n - x_{1j}\alpha}{a_n}\right) du, \end{aligned} \tag{A.2}$$



with  $r_2 = 1, y_2,$  or  $x_2$ . Under properties (1.iii) and (2.ii) of Assumption 3, Proposition 4 implies that  $\text{var}((\partial C_n(r_2|x_{1i}, \alpha)/\partial \alpha)|x_{1i})$  has order  $O(1/na_n)$  uniformly in  $x_{1i} \in X$ . Under Assumption 3, Proposition 3 implies that

$$\begin{aligned} \sup_{(\alpha, x_{1i}) \in \Theta_1 \times X} & \left\| E\left(\frac{\partial C_n(r_2|x_{1i}, \alpha)}{\partial \alpha} \middle| x_{1i}\right) \right. \\ & - \left\{ \int_{x_{1i}\alpha + \Delta_n}^{\infty} E((x_{1i} - x_1)'r_2 | -x_{1i}\alpha, z, \alpha)g(-x_{1i}\alpha, z|\alpha) dz \right. \\ & - \left. \int_{-x_{1i}\alpha}^{\infty} E((x_{1i} - x_1)'r_2 | u, x_{1i}\alpha + \Delta_n, \alpha) \right. \\ & \left. \left. \times g(u, x_{1i}\alpha + \Delta_n|\alpha) du \right\} \right\| = O(a_n^2). \end{aligned} \tag{A.3}$$

The limit functions in (A.3) are uniformly continuous in  $(x_{1i}, \alpha, \Delta_n)$  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  $\lim_{n \rightarrow \infty} (n/\ln n) a_n^3 = \infty,$

$$\begin{aligned} \frac{\partial C_n(r_2|x_{1i}, \bar{\alpha})}{\partial \alpha} & \xrightarrow{p} \int_{x_{1i}\alpha_0}^{\infty} E((x_{1i} - x_1)'r_2 | x_{1i}\alpha_0 = z)g(-x_{1i}\alpha_0, z|\alpha_0) dz \\ & - E((x_{1i} - x_1)'r_2 | x_{1i}\alpha_0 = x_{1i}\alpha_0) \\ & \times \int_{-x_{1i}\alpha_0}^{\infty} g(u, x_{1i}\alpha_0|\alpha_0) du, \end{aligned} \tag{A.4}$$

uniformly in  $x_{1i} \in X$ . With (3.5), (3.8), and (A.2.4),

$$\begin{aligned} & \left( \frac{\partial C_n(y_2|x_{1i}, \bar{\alpha})}{\partial \alpha'} - \beta'_0 \frac{\partial C_n(x'_2|x_{1i}, \bar{\alpha})}{\partial \alpha'} - E_n(v|x_{1i}, \bar{\alpha}) \frac{\partial C_n(1|x_{1i}, \bar{\alpha})}{\partial \alpha'} \right) \\ & \Big/ C_n(1|x_{1i}, \bar{\alpha}) \xrightarrow{p} G(x_{1i}, \alpha_0), \end{aligned} \tag{A.5}$$

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

$$\begin{aligned} G(x_{1i}, \alpha_0) & = \left( \int_{x_{1i}\alpha_0}^{\infty} \int_{-x_{1i}\alpha_0}^{\infty} g(u, z|\alpha_0) du dz \right)^{-1} \\ & \times \left\{ \int_{x_{1i}\alpha_0}^{\infty} E(v(x_{1i} - x_1) | -x_{1i}\alpha_0, z, \alpha_0)g(-x_{1i}\alpha_0, z|\alpha_0) dz \right. \\ & - \left. \int_{-x_{1i}\alpha_0}^{\infty} E(v(x_{1i} - x_1) | u, x_{1i}\alpha_0, \alpha_0)g(u, x_{1i}\alpha_0|\alpha_0) du \right. \end{aligned}$$

$$\begin{aligned}
& - E(v|u > -x_{1i}\alpha_0) \int_{x_{1i}\alpha_0}^{\infty} (x_{1i} - E(x_1|x_1\alpha_0 = z)) \\
& \times g(-x_{1i}\alpha_0, z|\alpha_0) dz \\
& + E(v|u > -x_{1i}\alpha_0)(x_{1i} - E(x_1|x_1\alpha_0 = x_{1i}\alpha_0)) \\
& \times \left. \int_{-x_{1i}\alpha_0}^{\infty} g(u, x_{1i}\alpha_0|\alpha_0) du \right\}. \tag{A.6}
\end{aligned}$$

The density  $g(u, z|\alpha_0)$  equals the product of the marginal density  $f_u(\cdot)$  of  $u$  and the marginal density  $h(z)$  of  $x_1\alpha_0$ , i.e.,

$$g(u, z|\alpha_0) = f_u(u)h(z). \tag{A.7}$$

Using (A.7) and that  $(u, v)$  is independent of  $x$  in the latent model,  $G(x_{1i}, \alpha_0)$  can be simplified to

$$\begin{aligned}
G(x_{1i}, \alpha_0) &= [E(v|u = -x_{1i}\alpha_0) - E(v|u > -x_{1i}\alpha_0)] \\
&\times \lambda(-x_{1i}\alpha_0)[x_{1i} - E(x_1|x_1\alpha_0 > x_{1i}\alpha_0)], \tag{A.8}
\end{aligned}$$

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  $E(v|u > z) = \int_z^{\infty} \int_{-\infty}^{\infty} v f(v, u) dv du / \int_z^{\infty} f_u(t) dt$ . It follows that

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

Therefore,

$$G(x_{1i}, \alpha_0) = -\tau(-x_{1i}\alpha_0)[x_{1i} - E(x_1|x_1\alpha_0 > x_{1i}\alpha_0)], \tag{A.9}$$

where  $\tau(z)$  denotes  $\partial E(v|u > z)/\partial z$ . With the uniform convergence of the non-parametric functions and their derivatives, the first term on the right-hand side of (A.1) will converge in probability to  $E(I_X(x_{1i})[x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0)]) \times \tau(-x_{1i}\alpha_0)[x_{1i} - E(x_1|x_1\alpha_0 > x_{1i}\alpha_0)]$  and the second term will converge in probability to a zero limit. This establishes (4.3).

The asymptotic distribution of  $\sqrt{n}L_n(\alpha_0, \beta_0)$  can be analyzed as follows. Define

$$L_{n,i}(\alpha_0, \beta_0) = I_X(x_{1i})[x_{2i} - E_n(x_2|x_{1i}, \alpha_0)]'(v_i - E_n(v|x_{1i}, \alpha_0)). \tag{A.10}$$

By a Taylor series expansion up to the second order,  $L_{n,i}(\alpha_0, \beta_0) = S_n(y_{2i}, x_i, \alpha_0, \beta_0) + R_n(y_{2i}, x_i, \alpha_0, \beta_0)$ , where

$$\begin{aligned} &S_n(y_{2i}, x_i, \alpha_0, \beta_0) \\ &= I_X(x_{1i}) \left\{ (x_{2i} - E_\infty(x_2|x_{1i}, \alpha_0, \Delta_n))'(v_i - E_\infty(v|x_{1i}, \alpha_0, \Delta_n)) \right. \\ &\quad + \left[ \frac{C(x_2'|x_{1i}, \alpha_0, \Delta_n)}{C^2(1|x_{1i}, \alpha_0, \Delta_n)} (v_i - E_\infty(v|x_{1i}, \alpha_0, \Delta_n)) \right. \\ &\quad \left. \left. + (x_{2i} - E_\infty(x_2|x_{1i}, \alpha_0, \Delta_n))' \frac{C(v|x_{1i}, \alpha_0, \Delta_n)}{C^2(1|x_{1i}, \alpha_0, \Delta_n)} \right] \right. \\ &\quad \times (C_n(1|x_{1i}, \bar{\alpha}) - C(1|x_{1i}, \alpha_0, \Delta_n)) \\ &\quad - (x_{2i} - E_\infty(x_2|x_{1i}, \alpha_0, \Delta_n))' \frac{1}{C(1|x_{1i}, \alpha_0, \Delta_n)} \\ &\quad \times (C_n(v|x_{1i}, \alpha_0) - C(v|x_{1i}, \alpha_0, \Delta_n)) \\ &\quad \left. - (v_i - E_\infty(v|x_{1i}, \alpha_0, \Delta_n)) \frac{1}{C(1|x_{1i}, \alpha_0, \Delta_n)} \right. \\ &\quad \left. \times (C_n(x_2|x_{1i}, \alpha_0) - C(x_2|x_{1i}, \alpha_0, \Delta_n))' \right\}, \tag{A.11} \end{aligned}$$

and

$$\begin{aligned}
& R_n(y_{2i}, x_i, \alpha_0, \beta_0) \\
&= I_X(x_{1i}) \left\{ \left[ \frac{\tilde{C}_n(v|x_{1i})}{\tilde{C}_n^4(1|x_{1i})} \tilde{C}_n(x'_2|x_{1i}) - \frac{\tilde{C}_n(x'_2|x_{1i})}{\tilde{C}_n^3(1|x_{1i})} \left( v_i - \frac{\tilde{C}_n(v|x_{1i})}{\tilde{C}_n(1|x_{1i})} \right) \right. \right. \\
&\quad \left. \left. - \left( x_{2i} - \frac{\tilde{C}_n(x_2|x_{1i})}{\tilde{C}_n(1|x_{1i})} \right)' \frac{\tilde{C}_n(v|x_{1i})}{\tilde{C}_n^4(1|x_{1i})} \right] (C_n(1|x_{1i}, \alpha_0) - C(1|x_{1i}, \alpha_0, \Delta_n))^2 \right. \\
&\quad \left. + \left[ \frac{1}{\tilde{C}_n^2(1|x_{1i})} \left( x_{2i} - \frac{\tilde{C}_n(x_2|x_{1i})}{\tilde{C}_n(1|x_{1i})} \right)' - \frac{\tilde{C}_n(x'_2|x_{1i})}{\tilde{C}_n^3(1|x_{1i})} \right] \right. \\
&\quad \times (C_n(1|x_{1i}, \alpha_0) - C(1|x_{1i}, \alpha_0, \Delta_n)) (C_n(v|x_{1i}, \alpha_0) - C(v|x_{1i}, \alpha_0, \Delta_n)) \\
&\quad \left. + \left[ \frac{1}{\tilde{C}_n^2(1|x_{1i})} \left( v_i - \frac{\tilde{C}_n(v|x_{1i})}{\tilde{C}_n(1|x_{1i})} \right) - \frac{\tilde{C}_n(v|x_{1i})}{\tilde{C}_n^3(1|x_{1i})} \right] \right. \\
&\quad \times (C_n(1|x_{1i}, \alpha_0) - C(1|x_{1i}, \alpha_0, \Delta_n)) \\
&\quad \times (C_n(x_2|x_{1i}, \alpha_0) - C(x_2|x_{1i}, \alpha_0, \Delta_n))' \\
&\quad \left. + \frac{1}{\tilde{C}_n^2(1|x_{1i})} (C_n(v|x_{1i}, \alpha_0) - C(v|x_{1i}, \alpha_0, \Delta_n)) \right. \\
&\quad \left. \times (C_n(x_2|x_{1i}, \alpha_0) - C(x_2|x_{1i}, \alpha_0, \Delta_n))' \right\},
\end{aligned}$$

with  $\tilde{C}_n(s|x_{1i})$  lying between  $C_n(s|x_{1i}, \alpha_0, \Delta_n)$  and  $C(s|x_{1i}, \alpha_0, \Delta_n)$  with  $s = 1, x_2, v$ . It follows that

$$\sqrt{n}L_n(\alpha_0, \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n S_n(y_{2i}, x_i, \alpha_0, \beta_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^n R_n(y_{2i}, x_i, \alpha_0, \beta_0). \tag{A.12}$$

The remainder  $(1/\sqrt{n}) \sum_{i=1}^n R_n(y_{2i}, x_i, \alpha_0, \beta_0)$  converges in probability to zero. Consider, for example, the last component in the remainder:

$$\begin{aligned}
& \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n I_X(x_{1i}) \frac{1}{\tilde{C}_n^2(1|x_{1i})} (C_n(v|x_{1i}, \alpha_0) - C(v|x_{1i}, \alpha_0, \Delta_n)) \right. \\
&\quad \left. \times (C_n(x_2|x_{1i}, \alpha_0) - C(x_2|x_{1i}, \alpha_0, \Delta_n))' \right\| \\
&\leq O_p(1) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n I_X(x_{1i}) |C_n(v|x_{1i}, \alpha_0) - C(v|x_{1i}, \alpha_0, \Delta_n)| \\
&\quad \times \|C_n(x_2|x_{1i}, \alpha_0) - C(x_2|x_{1i}, \alpha_0, \Delta_n)\|. \tag{A.13}
\end{aligned}$$

With the order of bias in (3.4) and the order of variance in (3.2), Proposition 5 implies that as  $\lim_{n \rightarrow \infty} na_n^8 = 0$  (A.13) converges in probability to zero. Similarly, the other components in the remainder converge to zero in probability. Some of the terms in  $(1/\sqrt{n}) \sum_{i=1}^n S_n(y_{2i}, x_i, \alpha_0, \beta_0)$  will also converge to zero in probability. Since  $v$  is independent of  $x$  in the latent model,  $E_\infty(v|x_{1i}, \alpha_0, \Delta_n) = E(v|u > -x_{1i}\alpha_0)$ . It follows that

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n I_X(x_{1i}) \frac{C(x'_2|x_{1i}, \alpha_0, \Delta_n)}{C^2(1|x_{1i}, \alpha_0, \Delta_n)} (v_i - E_\infty(v|x_{1i}, \alpha_0, \Delta_n)) \\ & \quad \times (C_n(1|x_{1i}, \alpha_0) - C(1|x_{1i}, \alpha_0, \Delta_n)) \\ & = \frac{1}{\sqrt{n(n-1)}} \sum_{i=1}^n \sum_{j \neq i}^n \Phi_{n,1}(r_i, r_j), \end{aligned} \tag{A.14}$$

and

$$\begin{aligned} & \Phi_{n,1}(r_i, r_j) \\ & = I_X(x_{1i}) \frac{C(x'_2|x_{1i}, \alpha_0, \Delta_n)}{C^2(1|x_{1i}, \alpha_0, \Delta_n)} \varepsilon_{2i} \\ & \quad \times \left( \int_{(x_{1i}\alpha_0 + \Delta_n - x_{1j}\alpha_0)/a_n}^{\infty} \int_{(-x_{1i}\alpha_0 - u_j)/a_n}^{\infty} K(u, z) du dz - C(1|x_{1i}, \alpha_0, \Delta_n) \right). \end{aligned} \tag{A.15}$$

The variance of  $\Phi_{n,1}(r_i, r_j)$  has order  $O(1)$ . Since  $E(\Phi_{n,1}(r_i, r_j)|x_i, r_j) = 0$ ,  $E(\Phi_{n,1}(r_i, r_j)) = 0$ . The conditions in Proposition 6 are satisfied for  $\Phi_{n,1}(\cdot)$ . It follows from Proposition 6 that (A.14) converges to zero in probability. Similarly, Proposition 6 implies that  $(1/\sqrt{n}) \sum_{i=1}^n I_X(x_{1i})(1/C(1|x_{1i}, \alpha_0, \Delta_n)) \times (C_n(x_2|x_{1i}, \alpha_0) - C(x_2|x_{1i}, \alpha_0, \Delta_n))' \varepsilon_{2i} \xrightarrow{p} 0$ . Hence, (A.12) is reduced to

$$\begin{aligned} & \sqrt{n}L_n(\alpha_0, \beta_0) \\ & = \frac{1}{\sqrt{n}} \sum_{i=1}^n I_X(x_{1i})(x_{2i} - E(x_2|x_{1i}\alpha_0 > x_{1i}\alpha_0 + \Delta_n))' \\ & \quad \times \left\{ \varepsilon_{2i} + \frac{1}{C(1|x_{1i}, \alpha_0, \Delta_n)} [E(v|u > x_{1i}\alpha_0) C_n(1|x_{1i}, \alpha_0) \right. \\ & \quad \left. - C_n(v|x_{1i}, \alpha_0)] \right\} + o_p(1) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{\sqrt{n(n-1)}} \sum_{i=1}^n \sum_{j \neq i}^n I_X(x_{1i})(x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0 + \Delta_n))' \\
 &\quad \times \left\{ \varepsilon_{2i} - \frac{(v_j - E(v|u > -x_{1i}\alpha_0))}{C(1|x_{1i}, \alpha_0, \Delta_n)} \right. \\
 &\quad \left. \times \int_{x_{1i}\alpha_0 + \Delta_n}^{\infty} \int_{-x_{1i}\alpha_0}^{\infty} \frac{1}{a_n^2} K\left(\frac{u - u_j}{a_n}, \frac{z - x_{1j}\alpha_0}{a_n}\right) du dz \right\} + o_p(1). \tag{A.16}
 \end{aligned}$$

The asymptotic distribution of (A.16) can be derived by the central limit theorem for  $U$ -statistics in Proposition 7. Denote

$$\begin{aligned}
 \Psi_{n,1}(r_i, r_j) &= I_X(x_{1i})(x_{2i} - E(x_2|x_1\alpha_0 > x_{1i}\alpha_0 + \Delta_n))' \\
 &\quad \times \left( \varepsilon_{2i} - \frac{(v_j - E(v|u > -x_{1i}\alpha_0))}{C(1|x_{1i}, \alpha_0, \Delta_n)} \right. \\
 &\quad \left. \times \int_{x_{1i}\alpha_0 + \Delta_n}^{\infty} \int_{-x_{1i}\alpha_0}^{\infty} \frac{1}{a_n^2} K\left(\frac{u - u_j}{a_n}, \frac{z - x_{1j}\alpha_0}{a_n}\right) du dz \right). \tag{A.17}
 \end{aligned}$$

Eq. (3.4) implies that

$$\begin{aligned}
 &\sup_{x_{1i}} \left| E\left( v_j \int_{x_{1i}\alpha_0 + \Delta_n}^{\infty} \int_{-x_{1i}\alpha_0}^{\infty} \frac{1}{a_n^2} K\left(\frac{u - u_j}{a_n}, \frac{z - x_{1j}\alpha_0}{a_n}\right) du dz \middle| x_{1i} \right) \right. \\
 &\quad \left. - E(v|u > -x_{1i}\alpha_0) C(1|x_{1i}, \alpha_0, \Delta_n) \right| \\
 &= O(a_n^2). \tag{A.18}
 \end{aligned}$$

It follows from (3.4) and (A.18) that  $E(\Psi_{n,1}(r_i, r_j)) = O(a_n^2)$ , and

$$\lim_{n \rightarrow \infty} E(\Psi_{n,1}(r_i, r_j)|r_i) = \Psi_1^{(1)}(r_i), \tag{A.19}$$

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

$$\begin{aligned}
 E(\Psi_{n,1}(r_j, r_i)|r_i) &= -E \left\{ I_X(x_{1j})(x_{2j} - E(x_2|x_1\alpha_0 > x_{1j}\alpha_0 + \Delta_n))' \right. \\
 &\quad \times \frac{v_i - E(v|u > -x_{1j}\alpha_0)}{C(1|x_{1j}, \alpha_0, \Delta_n)} \\
 &\quad \left. \times \int_{(x_{1j}\alpha_0 + \Delta_n - x_{1i}\alpha_0)/a_n}^{\infty} \int_{(-x_{1i}\alpha_0 - u_i)/a_n}^{\infty} K(u, z) du dz \middle| r_i \right\}. \tag{A.20}
 \end{aligned}$$

The function  $\int_{(x_{1j}\alpha_0 + \Delta_n - x_{1i}\alpha_0)/a_n}^{\infty} \int_{(-x_{1j}\alpha_0 - u_i)/a_n}^{\infty} K(u, z) du dz$  converges to 1 if  $x_{1j}\alpha_0 < x_{1i}\alpha_0$  and  $-x_{1j}\alpha_0 < u_i$ , to  $\int_{-\infty}^{\infty} \int_0^{\infty} K(u, z) du dz$  if  $x_{1j}\alpha_0 < x_{1i}\alpha_0$  and  $u_i = -x_{1j}\alpha_0$ , and to 0 otherwise. The event  $u_i = -x_{1j}\alpha_0$  occurs with zero probability. By the LDC theorem,

$$\lim_{n \rightarrow \infty} E(\Psi_{n,1}(r_j, r_i) | r_i) = \Psi_1^{(2)}(r_i), \tag{A.21}$$

where  $\Psi_1^{(2)}(r_i)$  is defined in (4.7). The limiting distribution of  $\sqrt{n}L_n(\alpha_0, \beta_0)$  follows from Proposition 7. This establishes (4.5).

With  $\hat{\alpha}$  in (4.8), it follows from (4.2) and (A.16) that

$$\sqrt{n}L_n(\hat{\alpha}, \beta_0) = \left[ I, \frac{\partial L_n(\bar{\alpha}, \beta_0)}{\partial \alpha'} C^{-1} \right] \frac{1}{\sqrt{n(n-1)}} \sum_{i=1}^n \sum_{j \neq i}^n \Psi_n(r_i, r_j) + o_p(1), \tag{A.22}$$

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

$$\begin{aligned} \Psi_{n,2}(r_i, r_j) &= I_X(x_{1i})(x_{1i} - E(x_1 | x_1\alpha_0 > x_{1i}\alpha_0 + \Delta_n))' \frac{\tau_1(x_{1i}\alpha_0)}{C(1 | x_{1i}, \alpha_0, \Delta_n)} \\ &\quad \times \int_{x_{1i}\alpha_0 + \Delta_n}^{\infty} \int_{-x_{1i}\alpha_0}^{\infty} (u_i - u) \frac{1}{a_n^2} K\left(\frac{u - u_j}{a_n}, \frac{z - x_{1j}\alpha_0}{a_n}\right) du dz. \end{aligned} \tag{A.23}$$

The CLT in Proposition 7 can be applied to derive the asymptotic distribution of (A.22). For  $\Psi_{n,2}(r_i, r_j)$ , it has been shown in Lee (1992) that  $E(\Psi_{n,2}(r_i, r_j)) = O(a_n^2)$ ,  $\text{var}(\Psi_{n,2}(r_i, r_j)) = O(1)$ ,

$$\lim_{n \rightarrow \infty} E(\Psi_{n,2}(r_i, r_j) | r_i) = I_X(x_{1i})(x_{1i} - E(x_1 | x_1\alpha_0 > x_{1i}\alpha_0))' \tau_1(x_{1i}\alpha_0) \varepsilon_{1i}, \tag{A.24}$$

and

$$\begin{aligned} &\lim_{n \rightarrow \infty} E(\Psi_{n,2}(r_j, r_i) | r_i) \\ &= - \int_{-u_i}^{x_{1i}\alpha_0} E[I_X(x_{1j})(x_{1j} - E(x_1 | x_1\alpha_0 > x_{1j}\alpha_0))' | x_{1j}\alpha_0 = z] \\ &\quad \times \frac{u_i - E(u | u > -z)}{\int_z^{\infty} h(t) dt} h(z) dz. \end{aligned} \tag{A.25}$$

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

$$\frac{1}{\sqrt{n(n-1)}} \sum_{i=1}^n \sum_{j \neq i}^n \Psi_n(r_i, r_j) \xrightarrow{D} N(0, \Sigma), \quad (\text{A.26})$$

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

## References

- Bierens, H.J., 1985, Kernel estimators of regression functions, in: T.W. Bewley ed., *Advances in the econometrics fifth world congress*, Vol. 1 (Cambridge University Press, New York, NY) 99–144.
- Chamberlain, G., 1986, Asymptotic efficiency in semi-parametric models with censoring, *Journal of Econometrics* 32, 189–218.
- Cosslett, S., 1991, Semiparametric estimation of a regression model with sample selectivity, Ch. 7 in: W.A. Barnett, J. Powell, and G. Tauchen, eds., *Nonparametric and semiparametric methods in econometrics and statistics* (Cambridge University Press, New York, NY).
- Heckman, J., 1974, Shadow prices, market wages, and labor supply, *Econometrica* 42, 679–694.
- 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.
- Ichimura, H., 1987, Estimation of single index models, Ph.D. thesis (Department of Economics, M.I.T., Cambridge, MA).
- Ichimura, H. and L.F. Lee, 1991, Semiparametric estimation of multiple index models: Single equation estimation, Ch. 1 in: W.A. Barnett, J. Powell, and G. Tauchen, eds., *Nonparametric and semiparametric methods in econometrics and statistics* (Cambridge University Press, New York, NY).
- Lee, L.F., 1992, Semiparametric nonlinear least square estimation of truncated regression models, *Econometric Theory* 8, 52–94.
- Lee, L.F., 1989, Semiparametric maximum likelihood estimation of polytomous and sequential choice models, Discussion paper no. 253 (Center for Economic Research, University of Minnesota, Minneapolis, MN).
- Lee, L.F., 1990, Efficient semiparametric scoring estimation of sample selection models, Discussion paper no. 255 (Center of Economics Research, University of Minnesota, Minneapolis, MN).
- Newey, W.K., 1988, Two step estimation of sample selection models, Manuscript (Department of Economics, Princeton University, Princeton, NJ).
- Olsen, R.J., 1980, A least squares correction for selectivity bias, *Econometrica* 48, 1815–1820.
- Powell, J.L., 1984, Least absolute deviations estimation for the censored regression model, *Journal of Econometrics* 25, 303–325.
- Powell, J.L., 1987, Semiparametric estimation of bivariate latent variable models, Discussion paper no. 8704 (Social Systems Research Institute, University of Wisconsin, Madison, WI).
- Rao, B.L.S.P., 1983, *Nonparametric functional estimation* (Academic Press, New York, NY).
- Robinson, P.M., 1988, Root- $n$ -consistent semiparametric regression, *Econometrica* 56, 931–954.
- Silverman, B.W., 1986, *Density estimation for statistics and data analysis* (Chapman and Hall, London).