# EFFICIENT SEMIPARAMETRIC SCORING ESTIMATION OF SAMPLE SELECTION MODELS

by

Lung-Fei Lee

Discussion Paper No. 255
February, 1990

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

**Efficient Semiparametric Scoring Estimation of Sample Selection Models**

by Lung-Fei Lee

Department of Economics
University of Minnesota
Minneapolis, Mn. 55455

Date: February, 1990

ABSTRACT

A semiparametric profile likelihood method is proposed for estimation of sample selection models. The method is a two step scoring semiparametric estimation procedure based on index formulation and kernel density estimation. Under some regularity conditions, the estimator is asymptotically normal. This method can be applied to estimation of general sample selection models with multiple regimes and sequential choice models with selectivity. For the binary choice sample selection model, the estimator is asymptotically efficiency in the sense that its asymptotic variance matrix attains the asymptotic bound of G. Chamberlain.

JEL classification number: 211
Keywords:
Sample selectivity, discrete choice, sequential choice, semiparametric model, index model, semiparametric estimation, method of scoring, profile likelihood, kernel density, efficient estimator

Corresponding address:
1035 Business and Economics Building, Department of Economics, Univeristy of Minnesota, 271 19th Avenue S., Minneapolis, Minnesota 55455

**Efficient Semiparametric Scoring Estimation of Sample Selection Models**

by Lung-Fei Lee *

## 1. Introduction

The sample selection model was introduced in Gronau [1974] and Heckman [1974] for the study of female labor suppy problem. Subsequently, it has been extended to incorporate more general discrete choice settings (eg. Dubin and McFadden [1984]). Under the normal distribution assumption, parametric estimation methods such as the maximum likelihood procedure and simple two stage methods are available for estimation of such models. Recently, alternative approaches which relax the normality assumption have been proposed for estimation of such models. Based on Edgeworth approximation of unknown distribution, Lee [1982] includes additional terms in a two stage estimation method. Cosselett [1984] has shown that the models can be consistently estimated without parametric distribution assumptions. Based on series approximation to unknown density functions, Gallant and Nychka [1987] has proposed a consistent semi-nonparametric maximum likelihood method. The asymptotic distributions of both the estimators of Cosslett and Gallant and Nychka are unknown. Semiparametric estimates which are $\sqrt{n}$ consistent and asymptotically normal are available in Robinson [1988], Powell [1987], Ichimura and Lee [1988] and Newey [1988]. The approaches in Robinson and Powell are single equation two stage estimation methods. The approach in Ichimura and Lee is a semiparametric nonliear least squares method. The asymptotic efficiency issue for semiparametric estimation of a two equations sample selection model has been investigated in Chamberlain [1986]. Chamberlain has derived an asymptotic efficiency bound for the variances of $\sqrt{n}$-consistent semiparametric estimators for the model. Newey [1988] claims that based on series approximation of conditional expectations, if more moment equations are included in a two stage estimation as sample size increases, the estimate can attain the efficiency bound of Chamberlain.

In this article, we propose a semiparametric profile likelihood estimation method based on index formulation and kernel estimation. The estimation method is motivated by methods of adaptive estimation (Bickel [1982], Manski [1984] and Schick [1986] and semiparametric estimation of index models (Ichimura [1987], Ichimura and Lee [1988] and Klein and Spady [1987]). The estimator is asymptotically normal. For the binary choice sample selection model, under some regularity conditions, the estimate can be asymptotically efficient in the sense its asymptotic variance attains the asymptotic efficiency bound of Chamberlain [1986]. This approach involves the construction of effective score functions and is a two step maximum profile likelihood estimation procedure. This procedure can be applied to estimation of general sample selection models with multiple discrete choice regimes and sequential choice models with selectivity.

This article is organized as follows. Section 2 introduces the semiparametric scoring estimator. Section 3 lists the sufficient regularity conditions for the asymptotic properties of the estimator. In that section, we explain briefly the logic and steps for our asymptotic anlysis. The main results are described. The detailed proofs are, however, collected in subsequent sections. Section 4 points out generalization of the procedure to estimation of sequential choice models with selectivity. In Section 5, we report some Monte Carlo results for our estimator. Detailed asymptotic analysis is provided in Section 5 to Section 7 and two appendices.

* I appreciate having financial support for my research under NSF grant no. SES-8809939 and computer time support from the Minnesota Supercomputer institute.

## 2. Sample Selection Models and Semiparametric Estimation

Sample selection models consist of some discrete choice equations and outcome equations. The outcome equations are defined on the whole population but their outcome values can be observed only selectively. Within an utility maximization framework (McFadden [1974]), suppose there are $L$ different alternatives, let

$$U_l = x\delta_l + u_l \tag{2.1}$$

be the associated utility for alternative $l, l = 1, \cdots, L$. Let $I_l$ be a dichotomous indicator that the alternative $l$ will be chosen:

$$I_l = 1 \Longleftrightarrow x\delta_l - x\delta_j \geq u_j - u_l, \quad j \neq l, \; j = 1, \cdots, L \tag{2.2}$$

and $I_l = 0$, otherwise. Associated with the first choice alternative, there is a vector of continuous outcomes:

$$y = x\gamma + \epsilon \tag{2.3}$$

where $y$ and $\epsilon$ are $k$ dimensional row vectors and $\gamma$ is a matrix of coefficients. The values of $y$ can be observed only when $I_1 = 1$. The general model may allow continuous outcome equations associated with each of the alternatives. However, for notational simplicity, we will concentrate on the model with outcome equations available only for $I_1 = 1$. The generalization to the switching regression models is straightforward.

Let $\tilde{f}(\epsilon, u)$ be the joint density of $\epsilon$ and $u$ where $u = (u_1, \cdots, u_L)$ and $h(u)$ be the marginal density of $u$. Under the assumption that the disturbances $\epsilon$ and $u$ are independent with $x$, the choice probability for the alternative $l$ is

$$P(I_l = 1|x) = \int_{-\infty}^{\infty} \left( \prod_{s \neq l}^{L} \int_{-\infty}^{x\delta_l - x\delta_s + u_l} \right) h(u) \left( \prod_{s \neq l}^{L} du_s \right) du_l. \tag{2.4}$$

The density function of $y$ conditional on $I_1 = 1$ and $x$ is

$$f(y|I_1 = 1, x) = \int_{-\infty}^{\infty} \left( \prod_{s=2}^{L} \int_{-\infty}^{x\delta_1 - x\delta_s + u_1} \right) \tilde{f}(y - x\gamma, u) \left( \prod_{s=2}^{L} du_s \right) du_1 / P(I_1 = 1|x). \tag{2.5}$$

Let $x\alpha_l = x\delta_l - x\delta_L$, $l = 1, \cdots, L-1$. The choice probabilities for the discrete alternatives are all functions of the indices $x\alpha_1, \cdots, x\alpha_{L-1}$ and the conditional density function of $y$ is function of the indices $x\alpha_1, \cdots, x\alpha_{L-1}$ and $y - x\gamma$.

As a generalization, let us assume that the number of indices in the choice probabilities are $m$ where $m$ can be greater than, equal to or less than the number of choices $L$. This generalization will include models with ordered choices and systems with multivariate qualitative dependent variables. To capture constraints, let $\theta$ be the vector of deep parameters in $\Theta$, $\Theta \subseteq R^{\bar{k}}$. Let $P_\theta$ denote the probability measure under $\theta$ being the true parameter vector. The choice probabilities in (2.4) are

$$\begin{aligned} &P(I_l = 1|x) \\ &= P_\theta(I_l = 1|x\alpha(\theta)) \\ &= E_\theta(I_l|x\alpha(\theta)) \end{aligned} \tag{2.6}$$

where $E_\theta(\cdot)$ denotes the expectation operator under $P_\theta$, and $x\alpha(\theta) = (x\alpha_1(\theta), \cdots, x\alpha_m(\theta))$. Let $p(x\alpha(\theta))$ be the density of $x\alpha(\theta)$, $g(y - x\gamma(\theta), x\alpha(\theta))$ be the density of $(y - x\gamma(\theta), x\alpha(\theta))$ , $g(y - x\gamma(\theta), x\alpha(\theta)|I_1 = 1)$ be the joint density of $(y - x\gamma(\theta), x\alpha(\theta))$ conditional on $I_1 = 1$ , $p(x\alpha(\theta)|I_1 = 1)$ be the marginal density of $x\alpha(\theta)$ conditional on $I_1 = 1$ and $f(y - x\gamma(\theta)|I_1 = 1, x\alpha(\theta))$ be the conditional density of $y - x\gamma(\theta)$ conditional on $I_1 = 1$ and $x\alpha(\theta)$ under $P_\theta$. From (2.5), it is apparent that the density function $y$ conditional on $I_1 = 1$ and $x$ under $P_\theta$ is exactly the density function of $y - x\gamma(\theta)$ conditional on $I_1 = 1$ and $x\alpha(\theta)$ under $P_\theta$, i.e.,

$$\begin{aligned} f(y|I_1 = 1, x) &= f(y - x\gamma(\theta)|I_1 = 1, x\alpha(\theta)) \\ &= g(y - x\gamma(\theta), x\alpha(\theta)|I_1 = 1)/p(x\alpha(\theta)|I_1 = 1) \end{aligned} \tag{2.7}$$

The sample selection model implies the above specific index structures. Index structures play useful role for semiparametric estimation of econometric models. Taking into account the index structures, Powell [1987] proposed a two stage method (see also Newey [1988] for a generalization) and Ichimura and Lee [1988] proposed a semiparametric nonlinear least square method for estimation of sample selection models. Several studies including Ruud [1986], Ichimura [1987] and Klein and Spady [1987] have explored index structures for semiparametric estimation in different contexts. The estimation method proposed in this articles will also take advantages of the index structure. It is a kind of adaptive estimation procedure motivated by the works on adaptive estimation (Stone [1975], Bickel [1982] , Manski [1984] and Schick [1986]).

Our estimation procedure starts with a given $\sqrt{n}$-consistent estimate $\tilde{\theta}_n$ of $\theta$. As in Bickel [1982], the consistent estimate $\tilde{\theta}_n$ can be discretized with a device by Le Cam. Suppose $\theta$ is of dimension $\bar{k}$. Let $\| \theta \| = \max_{l=1,\cdots,\bar{k}} |\theta_l|$, where $\theta = (\theta_1, \cdots, \theta_{\bar{k}})' \in R^{\bar{k}}$, be the norm in $R^{\bar{k}}$. Let $R_n^{\bar{k}} = \{\frac{1}{\sqrt{n}}(i_1, \cdots, i_{\bar{k}})|i_1, \ldots, i_{\bar{k}}$ are integers $\}$, and let $\bar{\theta}_n$ be a point in $R_n^{\bar{k}}$ closest to $\tilde{\theta}_n$ under the norm $\| \cdot \|$. The $\bar{\theta}_n$ is a discretized estimate of $\theta$. Given a random sample of size $n$, the probability function $E(I_l|x_i\alpha(\theta))$ evaluated at $x_i$ and $\theta$ can be estimated by nonparametric kernel regression function (Ichimura [1987] and Klein and Spady [1987]):

$$P_{n,l}(x_i, \theta) = \frac{A_{n,l}(x_i, \theta)}{B_n(x_i, \theta)} \tag{2.8}$$

where

$$A_{n,l}(x_i, \theta) = \frac{1}{n-1} \sum_{j \neq i}^{n} I_{lj} \frac{1}{a_n^m} K(\frac{x_i\alpha(\theta) - x_j\alpha(\theta)}{a_n}), \tag{2.9}$$

$$B_n(x_i, \theta) = \frac{1}{n-1} \sum_{j \neq i}^{n} \frac{1}{a_n^m} K(\frac{x_i\alpha(\theta) - x_j\alpha(\theta)}{a_n}) \tag{2.10}$$

and $K(\cdot)$ is a kernel function on $R^m$ and $a_n > 0$ is a bandwidth sequence. Similarly, the conditional density function $f(y - x\gamma(\theta)|I_1 = 1, x\alpha(\theta))$ of $y - x\gamma(\theta)$ conditional on $I_1 = 1$ and $x\alpha(\theta)$ can be estimated by

$$f_n(y_i - x_i\gamma(\theta)|I_{1i} = 1, x_i\alpha(\theta)) = \frac{C_n(z_i, \theta)}{A_{n,1}(x_i, \theta)} \tag{2.11}$$

where

$$C_n(z_i, \theta) = \frac{1}{n-1} \sum_{j \neq i}^{n} I_{1j} \frac{1}{a_n^{m+k}} \bar{K}(\frac{(y_i - x_i\gamma(\theta)) - (y_j - x_j\gamma(\theta))}{a_n}, \frac{x_i\alpha(\theta) - x_j\alpha(\theta)}{a_n}) \tag{2.12}$$

and $\bar{K}(\cdot)$ is a kernel function on $R^{m+k}$ and $z_i$ denotes $(x_i, y_i)$ [1]. The two step estimate that will be studied is

$$\hat{\theta}_n = \bar{\theta}_n + I_n^{-1}(\bar{\theta}_n) \cdot S_n(\bar{\theta}_n) \tag{2.13}$$

where

$$\begin{aligned} &S_n(\bar{\theta}_n) \\ &= \frac{1}{n}\sum_{i=1}^{n}\{I_{\tilde{Z}_{\bar{\theta}_n}}(z_i)I_{1i}[\frac{1}{C_n(z_i,\bar{\theta}_n)}\cdot\frac{\partial C_n(z_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{A_{n,1}(x_i,\bar{\theta}_n)}\cdot\frac{\partial A_{n,1}(x_i,\bar{\theta}_n)}{\partial\theta}] \\ &+ I_{\tilde{X}_{\bar{\theta}_n}}(x_i)\sum_{l=1}^{L} I_{li}[\frac{1}{A_{n,l}(x_i,\bar{\theta}_n)}\cdot\frac{\partial A_{n,l}(x_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{B_n(x_i,\bar{\theta}_n)}\cdot\frac{\partial B_n(x_i,\bar{\theta}_n)}{\partial\theta}]\}, \end{aligned} \tag{2.14}$$

$$\begin{aligned} &I_n(\bar{\theta}_n) \\ &= \frac{1}{n}\sum_{i=1}^{n}\{I_{\tilde{Z}_{\bar{\theta}_n}}(z_i)I_{1i}[\frac{1}{C_n(z_i,\bar{\theta}_n)}\cdot\frac{\partial C_n(z_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{A_{n,1}(x_i,\bar{\theta}_n)}\cdot\frac{\partial A_{n,1}(x_i,\bar{\theta}_n)}{\partial\theta}] \\ &\cdot[\frac{1}{C_n(z_i,\bar{\theta}_n)}\cdot\frac{\partial C_n(z_i,\bar{\theta}_n)}{\partial\theta'} - \frac{1}{A_{n,1}(x_i,\bar{\theta}_n)}\cdot\frac{\partial A_{n,1}(x_i,\bar{\theta}_n)}{\partial\theta'}] \\ &+ I_{\tilde{X}_{\bar{\theta}_n}}(x_i)\sum_{l=1}^{L} I_{li}[\frac{1}{A_{n,l}(x_i,\bar{\theta}_n)}\cdot\frac{\partial A_{n,l}(x_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{B_n(x_i,\bar{\theta}_n)}\cdot\frac{\partial B_n(x_i,\bar{\theta}_n)}{\partial\theta}] \\ &\cdot[\frac{1}{A_{n,l}(x_i,\bar{\theta}_n)}\cdot\frac{\partial A_{n,l}(x_i,\bar{\theta}_n)}{\partial\theta'} - \frac{1}{B_n(x_i,\bar{\theta}_n)}\cdot\frac{\partial B_n(x_i,\bar{\theta}_n)}{\partial\theta'}]\} \end{aligned} \tag{2.15}$$

and $I_{\tilde{Z}_{\bar{\theta}_n}}$ and $I_{\tilde{X}_{\bar{\theta}_n}}$ are the indicator functions of the sets $\tilde{Z}_{\bar{\theta}_n}$ and $\tilde{X}_{\bar{\theta}_n}$ respectively[2], where

$$\tilde{Z}_{\bar{\theta}_n} = \{(y,x)|\tilde{T}_1(\bar{\theta}_n) + \delta_n \le I_1(y - x\gamma(\bar{\theta}_n), x\alpha(\bar{\theta}_n)) \le \tilde{T}_2(\bar{\theta}_n) - \delta_n\} \tag{2.16}$$

$$\tilde{X}_{\bar{\theta}_n} = \{x|\tilde{T}_{1,X}(\bar{\theta}_n) + \delta_n \le x\alpha(\bar{\theta}_n) \le \tilde{T}_{2,X}(\bar{\theta}_n) - \delta_n\} \tag{2.17}$$

with

$$\tilde{T}_2(\bar{\theta}_n) = (\max_{i=1,\cdots,n}\{I_{1i}(y_i - x_i\gamma(\bar{\theta}_n))\}, \max_{i=1,\cdots,n}\{I_{1i}(x_i\alpha(\bar{\theta}_n))\})$$

$$\tilde{T}_1(\bar{\theta}_n) = (\min_{i=1,\cdots,n}\{I_{1i}(y_i - x_i\gamma(\bar{\theta}_n))\}, \min_{i=1,\cdots,n}\{I_{1i}(x_i\alpha(\bar{\theta}_n))\})$$

$$\tilde{T}_{2,X}(\bar{\theta}_n) = \max_{i=1,\cdots,n}\{x_i\alpha(\bar{\theta}_n)\}$$

[1] As a generalization, it is possible to allow a bandwidth parameter $a_n$ in the kernel estimation of the conditional density function in (2.11) and a bandwidth parameter $b_n$ with different rate in the kernel estimation of the choice probability in (2.8).

[2] The maximum and minimum operations applied to vectors below refer to component-wide operations.

and

$$\tilde{T}_{1,X}(\bar{\theta}_n) = \min_{i=1,\cdots,n} \{x_i\alpha(\bar{\theta}_n)\}.$$

$[\tilde{T}_1(\bar{\theta}_n),\ \tilde{T}_2(\bar{\theta}_n)]$ estimates the support of $(y - x\gamma(\bar{\theta}_n), x\alpha(\bar{\theta}_n))$ for $I_1 = 1$; $[\tilde{T}_{1,X}(\bar{\theta}_n),\ \tilde{T}_{2,X}(\bar{\theta}_n)]$ estimates the support of $x\alpha(\bar{\theta}_n)$ and $\delta_n > 0$ is a sequence of positive vectors such that $\lim_{n\to\infty} \delta_n = 0$ but $\lim_{n\to\infty} \frac{\delta_n}{a_n} = \infty$. The sets in (2.16) and (2.17) are designed to trim the observations from the tails of the indices. This trimming will avoid the troublesome caused in estimating the densities of the indices at their boundaries.

### 3. Regularity Conditions and The Main Results

Our proposed estimator can be shown to be consistent and asymptotically normal. It is also asymptotically efficient for semiparametric estimation with some infinite dimensional parameter space of distributions. To justify these asymptotic properties, we assume the following regularity conditions for the model.

**Assumption 1:**

(1) The samples $(I_{1i},\cdots,I_{Li},x_i,I_{1i}y_i)$, $i=1,\cdots,n$ are i.i.d.
(2) $\Theta$ is a compact neighborhood of the true parameter vector $\theta_o$ in a finite dimensional Euclidean space. The mapping $\alpha(\theta)$ and $\gamma(\theta)$ are twice continuously differentiable on $\Theta$.
(3) The explanatory variables $x$ are independent with the disturbances $\epsilon$.
(4) The density function $f(\epsilon|I_1=1,t)$ and the probability functions $P(I_l=1|t)$,where $t=x\alpha(\theta)$ and $\epsilon=y-x\gamma(\theta)$, are twice continuously differentiable in $t$ and $\epsilon$ on their supports.
(5) The functions $\sup_{\theta\in\Theta}\parallel\frac{\partial}{\partial w}f(y-x\gamma(\theta)|I_1=1,x\alpha(\theta))\parallel$, $\sup_{\theta\in\Theta}\parallel\frac{\partial^2}{\partial w\partial w'}f(y-x\gamma(\theta)|I_1=1,x\alpha(\theta))\parallel$, $\sup_{\theta\in\Theta}\parallel\frac{\partial}{\partial t}P(I_l=1|x\alpha(\theta))\parallel$ and $\sup_{\theta\in\Theta}\parallel\frac{\partial^2}{\partial t\partial t'}P(I_l=1|x\alpha(\theta))\parallel$ for all $l=1,\cdots,L$ are Lebesque integrable.
(6) The functions $\sup_{\theta\in\Theta}\parallel\frac{\partial^2}{\partial\theta\partial\theta'}\ln f(y-x\gamma(\theta)|I_1=1,x\alpha(\theta))\parallel$ and $\sup_{\theta\in\Theta}\parallel\frac{\partial^2}{\partial\theta\partial\theta'}\ln P(I_l=1|x\alpha(\theta))\parallel$ for all $l=1,\cdots,L$ are integrable under $P_{\theta_o}$.
(7) The information matrix $I(\theta_o)$ is invertible.
(8) $\tilde{\theta}_n$ is a $\sqrt{n}$ consistent estimator of $\theta_o$.

**Assumption 2:**

(1) Both the supports $S_y$ of $y$ and $S_x$ of $x$ are compact sets. Furthermore, for each $\theta\in\Theta$, the support of $x\alpha(\theta)$ where $x\alpha(\theta)=(x\alpha_1(\theta),\cdots,x\alpha_m(\theta))$ is a compact rectangle $[T_{1,X}(\theta),\ T_{2,X}(\theta)]$ in $R^m$ and the support $[T_1(\theta),\ T_2(\theta)]$ of $I_1(y-x\gamma(\theta),x\alpha(\theta))$ is a compact rectangle in $R^{k+m}$.
(2) The density functions $p(t)$, $f(\epsilon|I_1=1,t)$ and the probability functions $E_\theta(I_l|t)$, $l=1,\cdots,L$, where $t=x\alpha(\theta)$ and $\epsilon=y-x\gamma(\theta)$, are bounded away from zero on $S_y\times S_x\times\Theta$.
(3) The density functions $p(t)$, $g(\epsilon,t)$; the probability functions $E_\theta(I_l|t)$, $l=1,\cdots,L$; $E_\theta(I_1|\epsilon,t)$ and their derivatives $\frac{\partial}{\partial t}p(t)$, $\frac{\partial}{\partial w}g(\epsilon,t)$ ,$\frac{\partial}{\partial t}E_\theta(I_l|t)$, $l=1,\cdots,L$; $\frac{\partial}{\partial w}E_\theta(I_1|\epsilon,t)$ where $w=(\epsilon,t)$, are uniformly continuous in $t$ and $\epsilon$ on the interiors of $[T_{1,X}(\theta),\ T_{2,X}(\theta)]$ and $[T_1(\theta),\ T_2(\theta)]$, uniformly in $\theta\in\Theta$, and are uniformly bounded.
(4) The conditional expectation functions $E_\theta(x|t)$, $E_\theta(I_lx|t)$, $l=1,\cdots,L$; $E_\theta(I_1x|\epsilon,t)$ and their derivatives $\frac{\partial}{\partial t}E_\theta(x|t)$ ,$\frac{\partial}{\partial t}E_\theta(I_lx|t)$, $l=1,\cdots,L$; $\frac{\partial}{\partial w}E_\theta(I_1x|\epsilon,t)$ are uniformly continuous in $t$ and $\epsilon$ on the interiors of $[T_{1,X}(\theta),\ T_{2,X}(\theta)]$ and $[T_1(\theta),\ T_2(\theta)]$, uniformly in $\theta\in\Theta$, and are uniformly bounded.

**Assumption 3:**

(1) The kernel function $K(t)$ on $R^m$ is a function with bounded support such that $\int_{R^m}K(t)dt=1$ and $\int_{R^m}|K(t)|dt<\infty$. Similarly, $\bar{K}(w)$ is a kernel function with bounded support on $R^{k+m}$.
(2) The kernel functions $K(t)$ and $\bar{K}(w)$ are twice differentiable and their first and second order derivatives are bounded.
(3) The bandwidth sequence $\{a_n\}$ for the kernels has a rate such that $\lim_{n\to\infty}\frac{n}{\ln n}a_n^{k+m+2}=\infty$ and $\lim_{n\to\infty}a_n=0$.

**Assumption 4:**

(1) The kernel functions $K(t)$ and $\bar{K}(w)$ are high order kernel functions with zero moments up to the order $s^*$ where $s^* > m + k + 1$.

(2) The bandwidth sequence $\{a_n\}$ is chosen with a rate such that $\lim_{n\to\infty} n a_n^{2(m+k+1)} = \infty$ and $\lim_{n\to\infty} n a_n^{2s^*} = 0$ [3].

**Assumption 5:**

(1) The density functions $p(t)$ and $g(\epsilon, t)$ are differentiable in $t$ and $\epsilon$ to the order $s^*$ on the interiors of $[T_{1,X}(\theta),\ T_{2,X}(\theta)]$ and $[T_1(\theta),\ T_2(\theta)]$, and these derivatives are all bounded, uniformly in $\theta \in \Theta$.

(2) The conditional expectation functions $E_\theta(x|t)$,$E_\theta(I_l|t)$, $E_\theta(I_l x|t)$, $l = 1, \cdots, L$; $E_\theta(I_1|\epsilon, t)$, $E_\theta(I_1 x|\epsilon, t)$ and their derivatives $\frac{\partial}{\partial t}E_\theta(x|t)$, $\frac{\partial}{\partial t}E_\theta(I_l|t)$ ,$\frac{\partial}{\partial t}E_\theta(I_l x|t)$, $l = 1, \cdots, L$, $\frac{\partial}{\partial w}E_\theta(I_1|\epsilon, t)$ and $\frac{\partial}{\partial w}E_\theta(I_1 x|\epsilon, t)$ are differentiable in $(\epsilon, t)$ to the order $s^* + 1$ and these derivatives are bounded, uniformly in $\theta \in \Theta$.

The conditions in Assumption 1 are some basic regularity conditions for the model which justify some of the fundamental relations on the likelihood functions. The initial $\sqrt{n}$ consistent estimate can be constructed from several ways. For example, for the model with binary choice , we can estimate the choice equation by Ichimura's semiparametric nonlinear least square method (Ichimura[1987]) or the maximum likelihood method of Klein and Spady [1987], and the outcome equation by two stage methods similar to Powell [1987]. Alternatively, we can estimate all the parameters simultaneously by a semiparametric nonlinear least square method in Ichimura and Lee [1988]. For models with polytomous choice , the choice equations can be estimated by a profile likelihood method in Lee [1989].

The nonsingular information matrix assumption is an identification condition. Identitification of the sample selection model has been considered in Chamberlain [1986], Powell [1987] and Ichimura and Lee [1988]. Essentially, it requires that the regressors in the choice equations have variables which are not contained in the outcome equations. In addition, the indices in the polytomous choice case need be distinguishable from each other.

The asymptotic properties of the estimator depend on some convergence properties and the orders of asymptotic biases of the kernel function estimates. Proper bandwidth rate is chosen to quarantee convergence of the kernel estimates. High order kernels are needed to obtain the desired small order of asymptotic biases. The uniform continuity and boundedness conditions quarantee that the convergences of the kernel functions are uniform. These regularity conditions can be implied by some basic conditions on the joint density function of the dependent and explanatory variables in the model. These conditions, however, are more direct. Some of the conditions in the assumptions are redundant and can be eliminated. However, we keep them there so as to clarify the essential conditions needed in various parts of the analysis.

Discretization of $\sqrt{n}$-consistent estimate accompanied with the concept of contiguity of probability measures provides a simple technical device for asymptotic analysis in the literature of adaptive estimation (Bickel [1982], Manski [1986] and Schick [1986]). For any $\sqrt{n}$-consistent estimate $\tilde{\theta}_n$ with its discretized estimate $\bar{\theta}_n$, $\| \tilde{\theta}_n - \bar{\theta}_n \| \le \frac{1}{\sqrt{n}}$ and hence

$$\begin{aligned}\sqrt{n}(\bar{\theta}_n - \theta_o) &= \sqrt{n}(\bar{\theta}_n - \tilde{\theta}_n) + \sqrt{n}(\tilde{\theta}_n - \theta_o) \\ &= O_p(1),\end{aligned}$$

[3] If the bandwidth parameter $a_n$ is used only for the kernel estimation of the conditional density function in (2.11) and a different bandwidth parameter $b_n$ is used for estimating the choice probability function, the bandwidth sequence $b_n$ will be chosen with a rate such that $\lim_{n\to\infty} \frac{n}{\ln n} b_n^{m+2} = \infty$, $\lim_{n\to\infty} n b_n^{2(m+1)} = \infty$ and $\lim_{n\to\infty} n b_n^{2s^*} = 0$ where $s^* > m + 1$.

i.e., $\bar{\theta}_n$ is also $\sqrt{n}$-consistent. This device is useful in the following way. For any statistic $W_n(\theta)$ constructed from a sample of size $n$, if we want to show that $W_n(\bar{\theta}_n)\xrightarrow{P_{\theta_\circ}}0$ under the probability measure $P_{\theta_\circ}$, it is sufficient to show that $W_n(\theta_n)\xrightarrow{P_{\theta_n}}0$ for any nonstochastic sequence $\{\theta_n\}$ such that $\theta_n = \theta_\circ + \frac{1}{\sqrt{n}}h_n$ with $\{h_n\}$ being a bounded sequence. Effectively, we can replace $\bar{\theta}_n$ by the nonstochastic sequence $\theta_n$ in the proof of convergence in probability. This is so from the following arguments. For any $\epsilon > 0$ , the $\sqrt{n}$-consistency of $\bar{\theta}_n$ implies the existence of constant $M$ such that $P_{\theta_\circ}(\| \bar{\theta}_n - \theta_\circ \| < \frac{1}{\sqrt{n}}M) > 1 - \epsilon/2$ when $n$ is large enough. Denote

$$\Theta_{M,n} = \{\theta | \, \| \theta - \theta_\circ \| < \frac{1}{\sqrt{n}}M \text{ and } \theta \in R_n^k\}.$$

These sets $\Theta_{M,n}$ have two interesting properties. The cardinality of $\Theta_{M,n}$ is finite and bounded , say by $\bar{M}$ independent of $n$. Furthermore, $\theta_n \in \Theta_{M,n}$ can be written as $\theta_n = \theta_\circ + \frac{1}{\sqrt{n}}h_n$ where $h_n$ is a bounded sequence. This is so since $\theta_n = \frac{1}{\sqrt{n}}(i_1, \cdots, i_k)$ and $|i_l - \sqrt{n}\theta_{\circ,l}| < M$, $l = 1, \cdots, k$ with $i_1, \cdots, i_k$ being integers imply that the cardinalities of the sets $\{(i_1, \cdots, i_k) | i_1, \cdots, i_k$ are integers and $|i_l - \sqrt{n}\theta_{\circ,l}| < M, \; l = 1, \cdots, k\}$ are finite and bounded for all $n$. For any $\delta > 0$,

$$\begin{aligned}
&P_{\theta_\circ}(|W_n(\bar{\theta}_n)| > \delta) \\
&\leq P_{\theta_\circ}(\bar{\theta}_n \notin \Theta_{M,n}) + P_{\theta_\circ}(|W_n(\bar{\theta}_n)| > \delta, \; \bar{\theta}_n \in \Theta_{M,n}) \\
&\leq \epsilon/2 + P_{\theta_\circ}( \sup_{\theta_n \in \Theta_{M,n}} |W_n(\theta_n)| > \delta) \\
&\leq \epsilon/2 + \sum_{j=1}^{\bar{M}} P_{\theta_\circ}(|W_n(\theta_n^{(j)})| > \delta)
\end{aligned} \tag{3.1}$$

where $\Theta_{M,n} = \{\theta_n^{(1)}, \cdots, \theta_n^{(\bar{M})}\}$ for large $n$. Hence if $W_n(\theta_n)\xrightarrow{P_{\theta_\circ}}0$ for any sequence $\{\theta_n\}$ where $\theta_n = \theta_\circ + \frac{1}{\sqrt{n}}h_n$ with $\{h_n\}$ being bounded,

$$\sum_{j=1}^{\bar{M}} P_{\theta_\circ}(|W_n(\theta_n^{(j)})| > \delta) \leq \epsilon/2 \tag{3.2}$$

when $n$ is large enough. Combining (3.1) and (3.2), $P_{\theta_\circ}(|W_n(\bar{\theta}_n)| > \delta) \leq \epsilon$ for large $n$, i.e, $W_n(\bar{\theta}_n)\xrightarrow{P_{\theta_\circ}}0$.

The concept of contiguity was introduced by LeCam [1960]. Roussas [1972] provided detailed discussion on the subject. Contiguity is a concept of 'nearness' of sequences of probability measures. The probability measures $\{P_n\}$ and $\{P_n'\}$ defined on a sequence of probability spaces are said to be contiguous if, for any sequence of random variables $T_n$, $T_n \to 0$ in $P_n$-probability if and only if $T_n \to 0$ in $P_n'$-probability. A useful characterization of contiguity can be derived with the induced log likelihood ratio. Let $f_n(x)$ be the density function of $P_n$ and $g_n(x)$ be the one corresponding to $P_n'$. Let $D_n(x)$ be the distribution of the log likelihood ratio $\ln\left(\frac{g_n(x)}{f_n(x)}\right)$. The probability measures $P_n$ and $P_n'$ are contiguous if and only if $\{D_n(\cdot)\}$ is relatively compact (in the topology of weakly convergence) and for any subsequence $\{m\} \subseteq \{n\}$ for which $D_n(x)$ converges in distribution to $D(x)$ a probability distribution, we have $\int_{-\infty}^{\infty} \exp(x)dD(x) = 1$. Several alternative characterizations of contiguity are also provided in Roussas([1972],p.11). The above characterization of contiguity can be used to prove that the probability measures $P_{\theta_\circ}$ and $P_{\theta_n}$, where $\theta_n = \theta_\circ + \frac{1}{\sqrt{n}}h_n$ with $\{h_n\}$ being a bounded deterministic sequence, for the sample selection model, are contiguous. Thus to prove that $W_n(\theta_n)\xrightarrow{P_{\theta_\circ}}0$, it suffices to prove $W_n(\theta_n)\xrightarrow{P_{\theta_n}}0$. In several arguments in our subsequent analysis, the proofs of convergence of statistics evaluated at $\theta_n$ are much simpler when $\theta_n$ were treated as the true parameter vector.

Under the assumption that the sample observations $(y_i, I_{1i}, \cdots, I_{Li}, x_i)$, $i = 1, \cdots, n$ are independent and drawn from a common distribution for each $i$, the log likelihood function for our model under $P_{\theta_o}$ will be

$$\ln L_n(\theta_o) = \sum_{i=1}^{n} \{ I_{1i} \ln f(y_i | I_{1i} = 1, x_i, \theta_o) + \sum_{l=1}^{L} I_{li} \ln P(I_{li} = 1 | x_i, \theta_o) \} \tag{3.3}$$

where $f(y|I = 1, x, \theta_o)$ is the density (2.5) evaluated at $\theta = \theta_o$ and $P(I_l = 1|x, \theta_o)$ is the choice probability (2.4) evaluated at $\theta = \theta_o$. The log likelihood function under $P_{\theta_n}$ is

$$\ln L_n(\theta_n) = \sum_{i=1}^{n} \{ I_{1i} \ln f(y_i | I_{1i} = 1, x_i, \theta_n) + \sum_{l=1}^{L} I_{li} \ln P(I_{li} = 1 | x_i, \theta_n) \} \tag{3.4}$$

By the mean value theorem, since $\theta_n = \theta_o + \frac{1}{\sqrt{n}} h_n$,

$$\begin{aligned} &\ln L_n(\theta_n) - \ln L_n(\theta_o) \\ &= \frac{\partial}{\partial \theta} \ln L_n(\theta_o) \cdot \frac{1}{\sqrt{n}} h_n + \frac{1}{2} \left( \frac{1}{n} h_n{}' \frac{\partial^2}{\partial \theta \partial \theta'} \ln L_n(\hat{\hat{\theta}}_n) \cdot h_n \right) \end{aligned} \tag{3.5}$$

where $\hat{\hat{\theta}}_n$ lies between $\theta_n$ and $\theta_o$. Under the regularity conditions in Assumption 1, $\frac{1}{n} \frac{\partial^2}{\partial \theta \partial \theta'} \ln L_n(\hat{\hat{\theta}}_n) \xrightarrow{P_{\theta_o}} -I(\theta_o)$ and $\frac{1}{\sqrt{n}} \cdot \frac{\partial}{\partial \theta} \ln L_n(\theta_o) \xrightarrow{D} N(0, I(\theta_o))$ under $P_{\theta_o}$, where $I(\theta_o)$ is an information matrix. For any convergent sequence $h_n$ with limit $h$,

$$\ln L_n(\theta_n) - \ln L_n(\theta_o) \xrightarrow{D} N(-\frac{1}{2} h' I(\theta_o) h, \ \ h' I(\theta_o) h) \tag{3.6}$$

under $P_{\theta_o}$. The sequence of distributions of $\ln \left( \frac{L_n(\theta_n)}{L_n(\theta_o)} \right)$ under $P_{\theta_o}$ with $\theta_n = \theta_o + \frac{1}{\sqrt{n}} h_n$, where $\{h_n\}$ is bounded, is relatively compact because for each bounded sequence $\{h_n\}$ there exists a convergent subsequence $\{h_m\}$. The moment generating function $\phi(t)$ of the limiting normal distribution in (3.6) is $\exp\{-\frac{1}{2} h' I(\theta_o) h t + \frac{1}{2} h' I(\theta_o) h t^2\}$. When $t = 1$, $\phi(1) = 1$. By the characterization of contiguity, $\{P_{\theta_o}\}$ and $\{P_{\theta_o + \frac{1}{\sqrt{n}} h_n}\}$ will be contiguous.

The asymptotic distribution for our two step estimator $\hat{\theta}_n$ in (2.13) can be derived with several steps. The detailed analysis will be in subsequence sections. Here, we outline some basic steps of the analysis. Define the random vector $\theta_n^*$:

$$\theta_n^* = \bar{\theta}_n + I_n^{-1}(\bar{\theta}_n) S_n^*(\bar{\theta}_n) \tag{3.7}$$

where

$$\begin{aligned} &S_n^*(\bar{\theta}_n) \\ &= \frac{1}{n} \sum_{i=1}^{n} \{ I_{1i} \left[ \frac{1}{C(z_i, \bar{\theta}_n)} \frac{\partial C(z_i, \bar{\theta}_n)}{\partial \theta} - \frac{1}{A_1(x_i, \bar{\theta}_n)} \frac{\partial A_1(x_i, \bar{\theta}_n)}{\partial \theta} \right] \\ &+ \sum_{l=1}^{L} I_{li} \left[ \frac{1}{A_l(x_i, \bar{\theta}_n)} \frac{\partial A_l(x_i, \bar{\theta}_n)}{\partial \theta} - \frac{1}{B(x_i, \bar{\theta}_n)} \frac{\partial B(x_i, \bar{\theta}_n)}{\partial \theta} \right] \}, \end{aligned} \tag{3.8}$$

$$
\begin{aligned}
&\frac{1}{C(z_i,\theta)}\frac{\partial C(z_i,\theta)}{\partial\theta}-\frac{1}{A_1(x_i,\theta)}\frac{\partial A_1(x_i,\theta)}{\partial\theta}\\
&=\left\{E_\theta\left(\left[\frac{\partial x\gamma}{\partial\theta},-\frac{\partial x\alpha}{\partial\theta}\right]|x_i\alpha\right)-[\frac{\partial x_i\gamma}{\partial\theta},-\frac{\partial x_i\alpha}{\partial\theta}]\right\}\bar{\nabla}\ln f(y_i-x_i\gamma|I_{1i}=1,x_i\alpha)
\end{aligned}
\tag{3.9}
$$

with $\bar{\nabla}'\ln f(\epsilon|I_1=1,t)=\frac{\partial}{\partial(\epsilon,t)}\ln f(\epsilon|I_1=1,t)$, and

$$
\frac{1}{A_l(x_i,\theta)}\frac{\partial A_l(x_i,\theta)}{\partial\theta}-\frac{1}{B(x_i,\theta)}\frac{\partial B(x_i,\theta)}{\partial\theta}=\left[\frac{\partial x_i\alpha}{\partial\theta}-E_\theta\left(\frac{\partial x\alpha}{\partial\theta}|x_i\alpha\right)\right]\nabla\ln P(I_l=1|x_i\alpha) \tag{3.10}
$$

with $\nabla'\ln P(I_l=1|t)=\frac{\partial}{\partial t}\ln P(I_l=1|t)$.

Since

$$
\begin{aligned}
&\sqrt{n}(\hat{\theta}_n-\theta_n^*)\\
&=I_n^{-1}(\bar{\theta}_n)\cdot\sqrt{n}(S_n(\bar{\theta}_n)-S_n^*(\bar{\theta}_n)),
\end{aligned}
\tag{3.11}
$$

$\sqrt{n}(\hat{\theta}_n-\theta_\circ)$ will have the same limiting distribution as $\sqrt{n}(\theta_n^*-\theta_\circ)$ under $P_{\theta_\circ}$ if $I_n^{-1}(\bar{\theta}_n)$ is stochastically bounded and $\sqrt{n}(S_n(\bar{\theta}_n)-S_n^*(\bar{\theta}_n))$ converges to zero in probability under $P_{\theta_\circ}$. The proofs of these two properties are in Section 7. Under the identification condition that $I(\theta_\circ)$ is nonsingular, it can be shown that

$$
I_n^{-1}(\theta_n)\xrightarrow{P_{\theta_\circ}}I^{-1}(\theta_\circ). \tag{3.12}
$$

where the matrix $I(\theta_\circ)$ is

$$
\begin{aligned}
I(\theta_\circ)=E_{\theta_\circ}\{&I_1\Big([\frac{\partial x\gamma_\circ}{\partial\theta},-\frac{\partial x\alpha_\circ}{\partial\theta}]-E_{\theta_\circ}([\frac{\partial x\gamma_\circ}{\partial\theta},\ -\frac{\partial x\alpha_\circ}{\partial\theta}]|x\alpha_\circ)\Big)\\
&\cdot\bar{\nabla}\ln f(y-x\gamma_\circ|I_1=1,x\alpha_\circ)\bar{\nabla}'\ln f(y-x\gamma_\circ|I_1=1,x\alpha_\circ)\\
&\cdot\Big([\frac{\partial x\gamma_\circ}{\partial\theta},\ -\frac{\partial x\alpha_\circ}{\partial\theta}]-E_{\theta_\circ}([\frac{\partial x\gamma_\circ}{\partial\theta},\ -\frac{\partial x\alpha_\circ}{\partial\theta}]|x\alpha_\circ)\Big)'\\
&+[\frac{\partial x\alpha_\circ}{\partial\theta}-E_{\theta_\circ}(\frac{\partial x\alpha_\circ}{\partial\theta}|x\alpha_\circ)]\sum_{l=1}^{L}I_l\nabla\ln P(I_l=1|x\alpha_\circ)\nabla'\ln P(I_l=1|x\alpha_\circ)\\
&\cdot[\frac{\partial x\alpha_\circ}{\partial\theta}-E_{\theta_\circ}(\frac{\partial x\alpha_\circ}{\partial\theta}|x\alpha_\circ)]'\}.
\end{aligned}
\tag{3.13}
$$

The limiting distribution of $\sqrt{n}(\theta_n^*-\theta_\circ)$ can be derived from Taylor's expansion. By Tayor's expansion of $S_n^*(\bar{\theta}_n)$ at $\theta_\circ$,

$$
S_n^*(\bar{\theta}_n)=S_n^*(\theta_\circ)+\frac{\partial S_n^*(\bar{\bar{\theta}}_n)}{\partial\theta'}(\bar{\theta}_n-\theta_\circ) \tag{3.14}
$$

where $\bar{\bar{\theta}}_n$ lies between $\bar{\theta}_n$ and $\theta_\circ$. Therefore, (3.7) and (3.14) imply

$$
\begin{aligned}
&\sqrt{n}(\theta_n^*-\theta_\circ)\\
&=\sqrt{n}(\bar{\theta}_n-\theta_\circ)+\sqrt{n}I_n^{-1}(\bar{\theta}_n)[S_n^*(\theta_\circ)+\frac{\partial S_n^*(\bar{\bar{\theta}}_n)}{\partial\theta'}(\bar{\theta}_n-\theta_\circ)]\\
&=(I+I_n^{-1}(\bar{\theta}_n)\frac{\partial S_n^*(\bar{\bar{\theta}}_n)}{\partial\theta'})\cdot\sqrt{n}(\bar{\theta}_n-\theta_\circ)+I_n^{-1}(\bar{\theta}_n)\cdot\sqrt{n}S_n^*(\theta_\circ).
\end{aligned}
\tag{3.15}
$$

where $I$ is an identity matrix. Under the assumed regularity conditions, it can be shown that

$$\frac{\partial S_n^*(\bar{\theta}_n)}{\partial \theta'} \xrightarrow{P_{\theta_o}} -I(\theta_o), \tag{3.16}$$

$$\sqrt{n} S_n^*(\theta_o) \xrightarrow{D} N(0, I(\theta_o)) \tag{3.17}$$

and

$$(I + I_n^{-1}(\theta_n) \frac{\partial S_n^*(\bar{\theta}_n)}{\partial \theta'}) \cdot \sqrt{n}(\bar{\theta}_n - \theta_o) \xrightarrow{P_{\theta_o}} 0. \tag{3.18}$$

By Slutsky's lemma, it follows from (3.15) that

$$\sqrt{n}(\theta_n^* - \theta_o) \xrightarrow{D} N(0, I^{-1}(\theta_o)). \tag{3.19}$$

This two step estimator possesses some asymptotic efficient properties. For the binary choice sample selection model, Chamberlain [1986] has derived the asymptotic efficiency bound for semiparametric estimators of that model. From the asymptotic distribution in (3.19) and the precision matrix in (3.13), one can see that this two step scoring estimator attains Chamberlain's efficiency bound. For general polytomous choice selection models, the estimators are also efficient under the parameter space that the choice probability functions are functions of the indices and the disturbances of the outcome equations are independent with the regressors. The latter conclusion is drawn from the results in Severini and Wong [1987]. Severini and Wong have provided characterization of semiparametric efficient estimators in profile likelihood estimation framework. Semiparametric estimators related to profile likelihood estimation are asymptotic efficient under some appropriate infinite dimensional parameter space if the semiparametric estimators are asymptotically equivalent to some finite dimensional parametric estimators. For our estimation, this characterization is satisfied because $S_n^*(\theta)$ in (3.8) is a parameterized score function and $\sqrt{n}(S_n(\bar{\theta}_n) - S_n^*(\bar{\theta}_n))$ converges to zero in probability.

## 4. Estimation of Sequential Choice Sample Selection Models

The method can be easily generalized to estimation of sequential choice models with selectivity. Consider the model with two stage decisions and a single continuous outcome equation. There are $L_1$ alternatives in the first stage decision. If the first alternative was chosen in the first decsion, the second stage decision would involve $L_2$ alternatives. The disturbances in these decisions can be correlated. Define the $L_2$ mutually exclusive and exhaustive dichotomous indicators $J_{li}$ for individual $i$ where $J_{li} = 1$ if the $l$ alternative in the second stage decision is chosen, and 0 otherwise. The continuous dependent variable $y$ can be observed only when $J_{1i} = 1$. The observations will be the vectors $(I_i, J_i, J_{1i}y_i), i = 1, \cdots, n$ where $I_i = (I_{1i}, \cdots, I_{L_1 i})$,and $J_i = (J_{1i}, \cdots, J_{L_2 i})$. The log likelihood function for the sample observations will be

$$\begin{aligned}&\ln L_n(\theta)\\ &= \frac{1}{n}\sum_{i=1}^{n}[\sum_{l=1}^{L_1} I_{li}\ln P(I_{li}=1|x_i\alpha) + I_{1i}\sum_{j=1}^{L_2} J_{ji}\ln P(J_{ji}=1|I_{1i}=1, x_i\alpha, x_i\delta)\\ &+ I_{1i}J_{1i}\ln f(y_i - x_i\gamma|I_{1i}=1, J_{1i}=1, x_i\alpha, x_i\delta)].\end{aligned} \tag{4.1}$$

Suppose that $x\alpha$ is a vector of dimension $m_1$, $x\delta$ is a vector of dimension $m_2$ and $y - x\gamma$ is a variable, let $K(\cdot)$, $\bar{K}(\cdot)$ and $\bar{\bar{K}}(\cdot)$ be kernel functions on $R^{m_1}, R^{m_1+m_2}$ and $R^{m_1+m_2+1}$ respectively. Define

$$A_{n,l}(x_i,\theta) = \frac{1}{n-1}\sum_{j\neq i}^{n} I_{lj}\frac{1}{a_n^{m_1}}K(\frac{x_i\alpha - x_j\alpha}{a_n}), \quad l = 1, \cdots, L_1$$

$$B_n(x_i,\theta) = \frac{1}{n-1}\sum_{j\neq i}^{n}\frac{1}{a_n^{m_1}}K(\frac{x_i\alpha - x_j\alpha}{a_n}),$$

$$\bar{A}_{n,l}(x_i,\theta) = \frac{1}{n-1}\sum_{j\neq i}^{n} I_{1j}J_{lj}\frac{1}{a_n^{m_1+m_2}}\bar{K}(\frac{x_i\alpha - x_j\alpha}{a_n}, \frac{x_i\delta - x_j\delta}{a_n}), \quad l = 1, \cdots, L_2$$

$$\bar{B}_n(x_i,\theta) = \frac{1}{n-1}\sum_{j\neq i}^{n} I_{1j}\frac{1}{a_n^{m_1+m_2}}\bar{K}(\frac{x_i\alpha - x_j\alpha}{a_n}, \frac{x_i\delta - x_j\delta}{a_n})$$

and

$$C_n(z_i,\theta) = \frac{1}{n-1}\sum_{j\neq i}^{n} I_{1j}J_{1j}\frac{1}{a_n^{m_1+m_2+1}}\bar{\bar{K}}(\frac{(y_i - x_i\gamma) - (y_j - x_j\gamma)}{a_n}, \frac{x_i\alpha - x_j\alpha}{a_n}, \frac{x_i\delta - x_j\delta}{a_n}).$$

The probabilities $P(I_{li} = 1|x_i\alpha)$ and $P(J_{ji} = 1|I_{1i} = 1, x_i\alpha, x_i\delta)$ can be estimated by

$$P_{n,l}(x_i,\theta) = \frac{A_{n,l}(x_i,\theta)}{B_n(x_i,\theta)}$$

and

$$\bar{P}_{n,l}(x_i,\theta) = \frac{\bar{A}_{n,l}(x_i,\theta)}{\bar{B}_n(x_i,\theta)}$$

respectively. The conditional density function $f(y_i - x_i\gamma | I_{1i} = 1, J_{1i} = 1, x_i\alpha, x_i\delta)$ can be estimated by

$$f_n(y_i - x_i\gamma | I_{1i} = 1, J_{1i} = 1, x_i\alpha, x_i\delta) = \frac{C_n(z_i,\theta)}{\bar{A}_{n,1}(x_i,\theta)}.$$

The two step estimator will be

$$\hat{\theta}_n = \bar{\theta}_n + I_n^{-1}(\bar{\theta}_n)S_n(\bar{\theta}_n) \tag{4.2}$$

where

$$\begin{aligned} S_n(\bar{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n}\{ & I_{\tilde{Z}_{\bar{\theta}_n}}(z_i)I_{1i}J_{1i}[\frac{1}{C_n(z_i,\bar{\theta}_n)}\frac{\partial C_n(z_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{\bar{A}_{n,1}(x_i,\bar{\theta}_n)}\frac{\partial \bar{A}_{n,1}(x_i,\bar{\theta}_n)}{\partial\theta}] \\ & + I_{\tilde{\tilde{X}}_{\bar{\theta}_n}}(x_i)I_{1i}\sum_{l=1}^{L_2}J_{li}[\frac{1}{\bar{A}_{n,l}(x_i,\bar{\theta}_n)}\frac{\partial \bar{A}_{n,l}(x_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{\bar{B}_n(x_i,\theta_n)}\frac{\partial \bar{B}_n(x_i,\theta_n)}{\partial\theta}] \\ & + I_{\tilde{X}_{\bar{\theta}_n}}(x_i)\sum_{l=1}^{L_1}I_{li}[\frac{1}{A_{n,l}(x_i,\bar{\theta}_n)}\frac{\partial A_{n,l}(x_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{B_n(x_i,\bar{\theta}_n)}\frac{\partial B_n(x_i,\bar{\theta}_n)}{\partial\theta}]\}, \end{aligned} \tag{4.3}$$

$$\begin{aligned} I_n(\bar{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n}\{ & I_{\tilde{Z}_{\bar{\theta}_n}}(z_i)I_{1i}J_{1i}[\frac{1}{C_n(z_i,\bar{\theta}_n)}\frac{\partial C_n(z_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{\bar{A}_{n,2}(x_i,\bar{\theta}_n)}\frac{\partial \bar{A}_{n,2}(x_i,\bar{\theta}_n)}{\partial\theta}] \\ & [\frac{1}{C_n(z_i,\bar{\theta}_n)}\frac{\partial C_n(z_i,\bar{\theta}_n)}{\partial\theta'} - \frac{1}{\bar{A}_{n,2}(x_i,\bar{\theta}_n)}\frac{\partial \bar{A}_{n,2}(x_i,\bar{\theta}_n)}{\partial\theta'}] \\ & + I_{\tilde{\tilde{X}}_{\bar{\theta}_n}}(x_i)I_{1i}\sum_{l=1}^{L_2}J_{li}[\frac{1}{\bar{A}_{n,l}(x_i,\bar{\theta}_n)}\frac{\partial \bar{A}_{n,l}(x_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{\bar{B}_n(x_i,\theta_n)}\frac{\partial \bar{B}_n(x_i,\theta_n)}{\partial\theta}] \\ & [\frac{1}{\bar{A}_{n,l}(x_i,\bar{\theta}_n)}\frac{\partial \bar{A}_{n,l}(x_i,\bar{\theta}_n)}{\partial\theta'} - \frac{1}{\bar{B}_n(x_i,\theta_n)}\frac{\partial \bar{B}_n(x_i,\theta_n)}{\partial\theta'}] \\ & + I_{\tilde{X}_{\bar{\theta}_n}}(x_i)\sum_{l=1}^{L_1}I_{li}[\frac{1}{A_{n,l}(x_i,\bar{\theta}_n)}\frac{\partial A_{n,l}(x_i,\bar{\theta}_n)}{\partial\theta} - \frac{1}{B_n(x_i,\bar{\theta}_n)}\frac{\partial B_n(x_i,\bar{\theta}_n)}{\partial\theta}] \\ & [\frac{1}{A_{n,l}(x_i,\bar{\theta}_n)}\frac{\partial A_{n,l}(x_i,\bar{\theta}_n)}{\partial\theta'} - \frac{1}{B_n(x_i,\bar{\theta}_n)}\frac{\partial B_n(x_i,\bar{\theta}_n)}{\partial\theta'}]\}, \end{aligned} \tag{4.4}$$

and the trimming sets $\tilde{Z}_{\bar{\theta}_n}, \tilde{\tilde{X}}_{\bar{\theta}_n}$ and $\tilde{X}_{\bar{\theta}_n}$ are

$$\tilde{Z}_{\bar{\theta}_n} = \{(y,x) | \tilde{T}_1(\bar{\theta}_n) + \delta_n \leq I_1J_1(y - x\gamma(\bar{\theta}_n), x\alpha(\bar{\theta}_n), x\delta(\bar{\theta}_n)) \leq \tilde{T}_2(\bar{\theta}_n) - \delta_n\},$$

$$\tilde{\tilde{X}}_{\bar{\theta}_n} = \{x | \tilde{T}_{1,\alpha,\delta}(\bar{\theta}_n) + \delta_n \leq I_1(x\alpha(\bar{\theta}_n), x\delta(\bar{\theta}_n)) \leq \tilde{T}_{2,\alpha,\delta}(\bar{\theta}_n) - \delta_n\}$$

and

$$\tilde{X}_{\bar{\theta}_n} = \{x | \tilde{T}_{1,\alpha}(\bar{\theta}_n) + \delta_n \leq x\alpha(\bar{\theta}_n) \leq \tilde{T}_{2,\alpha}(\bar{\theta}_n) - \delta_n\}$$

with

$$\tilde{T}_2(\bar{\theta}_n) = (\max_{i=1,\cdots,n}\{I_{1i}J_{1i}(y_i - x_i\gamma(\bar{\theta}_n))\}, \max_{i=1,\cdots,n}\{I_{1i}J_{1i}x_i\alpha(\bar{\theta}_n)\}, \max_{i=1,\cdots,n}\{I_{1i}J_{1i}x_i\delta(\bar{\theta}_n)\})$$

$$\tilde{T}_1(\bar{\theta}_n) = (\min_{i=1,\cdots,n}\{I_{1i}J_{1i}(y_i - x_i\gamma(\bar{\theta}_n))\}, \min_{i=1,\cdots,n}\{I_{1i}J_{1i}x_i\alpha(\bar{\theta}_n)\}, \min_{i=1,\cdots,n}\{I_{1i}J_{1i}x_i\delta(\bar{\theta}_n)\})$$

$$\tilde{T}_{2,\alpha,\delta}(\bar{\theta}_n) = (\max_{i=1,\cdots,n}\{I_{1i}x_i\alpha(\bar{\theta}_n)\}, \max_{i=1,\cdots,n}\{I_{1i}x_i\delta(\bar{\theta}_n)\})$$

$$\tilde{T}_{1,\alpha,\delta}(\bar{\theta}_n) = (\min_{i=1,\cdots,n}\{I_{1i}x_i\alpha(\bar{\theta}_n)\}, \min_{i=1,\cdots,n}\{I_{1i}x_i\delta(\bar{\theta}_n)\})$$

$$\tilde{T}_{2,\alpha}(\bar{\theta}_n) = \max_{i=1,\cdots,n}\{x_i\alpha(\bar{\theta}_n)\}$$

and

$$\tilde{T}_{1,\alpha}(\bar{\theta}_n) = \min_{i=1,\cdots,n}\{x_i\alpha(\bar{\theta}_n)\}.$$

Under similar regularity conditions,

$$\sqrt{n}(\hat{\theta}_n - \theta_\circ)\xrightarrow{D}N(0, I^{-1}(\theta_\circ))$$

where

$$\begin{aligned}
&I(\theta_\circ)\\
&= E_{\theta_\circ}\{I_1J_1\Big([\frac{\partial x\gamma_\circ}{\partial\theta}, -\frac{\partial x\alpha_\circ}{\partial\theta}, -\frac{\partial x\delta_\circ}{\partial\theta}] - E_{\theta_\circ}([\frac{\partial x\gamma_\circ}{\partial\theta}, -\frac{\partial x\alpha_\circ}{\partial\theta}, -\frac{\partial x\delta_\circ}{\partial\theta}]|x\alpha_\circ, x\delta_\circ)\Big)\\
&\quad\cdot\bar{\nabla}\ln f(y - x\gamma_\circ|I_1 = 1, J_1 = 1, x\alpha_\circ, x\delta_\circ)\bar{\nabla}'\ln f(y - x\gamma_\circ|I_1 = 1, J_1 = 1, x\alpha_\circ, x\delta_\circ)\\
&\quad\cdot\Big([\frac{\partial x\gamma_\circ}{\partial\theta}, -\frac{\partial x\alpha_\circ}{\partial\theta}, -\frac{\partial x\delta_\circ}{\partial\theta}] - E_{\theta_\circ}([\frac{\partial x\gamma_\circ}{\partial\theta}, -\frac{\partial x\alpha_\circ}{\partial\theta}, -\frac{\partial x\delta_\circ}{\partial\theta}]|x\alpha_\circ, x\delta_\circ)\Big)'\\
&\quad + I_1\sum_{l=1}^{L_2}J_l\Big([\frac{\partial x\alpha_\circ}{\partial\theta}, \frac{\partial x\delta_\circ}{\partial\theta}] - E_{\theta_\circ}([\frac{\partial x\alpha_\circ}{\partial\theta}, \frac{\partial x\delta_\circ}{\partial\theta}]|x\alpha_\circ, x\delta_\circ)\Big)\\
&\quad\cdot\nabla\ln P(J_l = 1|x\alpha_\circ, x\delta_\circ)\nabla'\ln P(J_l = 1|x\alpha_\circ, x\delta_\circ)\\
&\quad\cdot\Big([\frac{\partial x\alpha_\circ}{\partial\theta}, \frac{\partial x\delta_\circ}{\partial\theta}] - E_{\theta_\circ}([\frac{\partial x\alpha_\circ}{\partial\theta}, \frac{\partial x\delta_\circ}{\partial\theta}]|x\alpha_\circ, x\delta_\circ)\Big)'\\
&\quad + \sum_{l=1}^{L_1}I_l[\frac{\partial x\alpha_\circ}{\partial\theta} - E_{\theta_\circ}(\frac{\partial x\alpha_\circ}{\partial\theta}|x\alpha_\circ)]\\
&\quad\quad\cdot\nabla\ln P(I_l = 1|x\alpha_\circ)\nabla'\ln P(I_l = 1|x\alpha_\circ)\cdot[\frac{\partial x\alpha_\circ}{\partial\theta} - E_{\theta_\circ}(\frac{\partial x\alpha_\circ}{\partial\theta}|x\alpha_\circ)]'\}
\end{aligned}$$

which can be consistently estimated by $I_n(\bar{\theta}_n)$ from (4.4).

## 5. Monte Carlo Simulation

To investigate finite sample properties of the scoring estimates, some small scale Monte Carlo simulations have been performed. Simulated data have been generated for two equation binary choice sample selection models.

The choice equation which determines the dichotomous indicator is

$$y_1 = \alpha_0 + s + z_1\alpha_1 + w\alpha_2 + \epsilon_1 \tag{5.1}$$

and the outcome equation associated with choice alternative 1 is

$$y_2 = \gamma_0 + z_2\gamma_1 + w\gamma_2 + \epsilon_2 \tag{5.2}$$

where $s$ is generated by a truncated normal $N(0,1)$ random variable with support on [-1.9, 1.9]; $z_1$ and $z_2$ are generated by independent uniform variables with support on [-1, 1], and $w = w^* - 2$ where $w^*$ is generated by a Poisson variable with mean 2 but truncated with support on [0, 10]. Such explanatory variables imply that the density function of the index from the choice equation are bounded alway from zero on its support. The true parameters throughout the experiment are $\alpha_0 = 0$, $\alpha_1 = 1$, $\alpha_2 = -1$, $\gamma_0 = 0$, $\gamma_1 = 1$ and $\gamma_2 = -1$. Three models with disturbances generated from different distributions will be tested. In the first model, the disturbance $\epsilon_1$ is generated by a N(0,1) distribution. To capture correlation across the disturbances in the two equations, the disturbance $\epsilon_2$ is generated as follows:

$$\epsilon_2 = \begin{cases} \epsilon_1 + \epsilon^*, & \text{if } |\epsilon_1 + \epsilon^*| \leq 3 \\ U[-3,\ 3], & \text{if } |\epsilon_1 + \epsilon^*| > 3 \end{cases} \tag{5.3}$$

where $\epsilon^*$ is another $N(0,1)$ random variable independent with $\epsilon_1$ and $U[-3,\ 3]$ is independent uniform variable with support on $[-3,\ 3]$. $\epsilon_2$ is a bounded random variable with density function strictly bounded away from zero on its support. Its distribution is a mixture of truncated normal distribution and uniform distribution but is very closed to be a normal variable with variance 2. For the second model, the disturbance $\epsilon_2$ is exactly a normal variable with variance 2:

$$\epsilon_2 = \epsilon_1 + \epsilon^* \tag{5.4}$$

These two models can be used to investigate whether the assumption of strictly positive density for the disturbance of the outcome equation has any practical impact on the finite sample performance of the proposed estimation procedure. For these two models, the probit specification is a correctly or nearly correctly specified parametric model. The third model differs from the first two models in that the probit specification would be a misspecified parametric model. In the third model, both the disturbances are generated from mixtures of gamma and normal distributions:

$$\epsilon_1 = \sqrt{0.8}Gamma(0,1) + \sqrt{0.2}\epsilon_1^* \tag{5.5}$$

and

$$\epsilon_2 = \sqrt{0.7}Gamma(0,1) + \sqrt{0.3}\epsilon_2^* \tag{5.6}$$

where $Gamma(0,1)$ is a gamma distribution with zero mean and variance one and $\epsilon_1^*$ and $\epsilon_2^*$ are two independent $N(0,1)$ random variables. The correlation coefficient of the disturbances of the choice equation and the outcome equation is about 0.7 for all these three models.

The initial $\sqrt{n}$-consistent estimates are derived from a two stage semiparametric estimation procedure. The choice equation can be estimated by a maximum profile likelihood procedure similar to the procedure proposed in Klein and Spady [1987] with slight modification [4]:

$$\max_{\alpha} \frac{1}{n} \sum_{i=1}^{n} I(|s_i| \leq 1.9 - \delta_n)\{I_{1i} \ln P_{n,1}(x_i, \alpha) + (1 - I_{1i}) \ln(1 - P_{n,1}(x_i, \alpha))\} \tag{5.7}$$

where $P_{n,1}(x_i, \alpha)$ is a kernel estimate of the choice probability as in (2.8) and $\alpha = (\alpha_1, \alpha_2)$. For the choice equation, the number of indices is $m = 1$. The kernel function chosen for estimating such equation is the following univariate kernel function $K_4(t)$:

$$K_4(t) = 2K(t) - \frac{1}{\sqrt{2}} K(\frac{t}{\sqrt{2}}) \tag{5.8}$$

where

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

The function $K(t)$ is a proper density function which is twice continuously differentiable on the real line with bounded third order derivative and is computationally simple. The kernel function $K_4(t)$ is a high order kernel with its first three moments being zero. The way of constructing such high order kernel functions from density functions is suggested in Bierens[1987]. The bandwith sequence $\{b_n\}$ is chosen as

$$b_n = \frac{c_1}{n^{1/5.5}} \tag{5.9}$$

where $c_1$ is a constant proportional factor independent with sample size $n$. The sequence $\{\delta_n\}$ for trimming the choice index is chosen as

$$\delta_n = c b_n^{9/10} \tag{5.10}$$

where $c$ is another constant factor. The constant factor $c$ in $\delta_n$ is set to 0.1 so that the regressors are not excessively trimmed. The chosen kernel and the bandwidth sequences satisfy the regularity conditions in our assumptions. To avoid the potential empty space phenomenon, i.e., empty neighborhoods at some points in the kernel density estimation, and difficulty with negative probability values for the log likelihood function, the estimated probabilities $P_{n,l}(x_i, \alpha)$ are modified by the following rules:

1. If $B_n(x_i, \alpha) = 0$ and $A_{n,l}(x_i, \alpha) = 0$, set $P_{n,l}(x_i, \alpha) = 1/L$ ;
2. if $B_n(x_i, \alpha) = 0$, but $A_{n,l}(x_i, \alpha) \neq 0$, set $B_n(x_i, \alpha) = 10^{-10}$ ;
3. if $P_{n,l}(x_i, \alpha) \leq 10^{-10}$, set $P_{n,l}(x_i, \alpha) = 10^{-10}$ ; and
4. if $P_{n,l}(x_i, \alpha) \geq 1 - 10^{-10}$, set $P_{n,l}(x_i, \alpha) = 1 - 10^{-10}$. These modifications which guarantee the computed probabilities are properly bounded between zero and one are intuitively appealing. As shown in Lee[1989], such modification will not change the asymptotic distribution of the semiparametric maximum likelhood estimator (SMLE) for the choice equation. The log profile likelihood functions are maximized with the subroutine of downhill simplex method in Press et al [1986] with the probit estimates as one of the starting values.

[4] An alternative procedure among others is the semiparametric nonlinear least squares procedure in Ichimura [1987].

With the SMLE $\hat{\alpha}$, the parameter $\gamma$ where $\gamma = (\gamma_1, \gamma_2)$ can be estimated by the following semiparametric two stage estimate (SPTWOE):

$$\hat{\gamma} = \{ \sum_{i=1}^{n} I_{1i} I(|s_i| \leq 1.9 - \delta_n)(x_{2i} - \sum_{j \neq i}^{n} x_{2j} W_{nij}(\hat{\alpha}))'(x_{2i} - \sum_{j \neq i}^{n} x_{2j} W_{nij}(\hat{\alpha}))\}^{-1}$$
$$\cdot \sum_{i=1}^{n} I_{1i} I(|s_i| \leq 1.9 - \delta_n)(x_{2i} - \sum_{j \neq i}^{n} x_{2j} W_{nij}(\hat{\alpha}))'(y_{2i} - \sum_{j \neq i}^{n} y_{2j} W_{nij}(\hat{\alpha})) \quad (5.11)$$

where $x_2 = (z_2, w)$ and

$$W_{nij}(\hat{\alpha}) = I_{1j} K(\frac{x_i \alpha(\hat{\theta}) - x_j \alpha(\hat{\theta})}{b_n}) / \sum_{k \neq i}^{n} I_{1k} K(\frac{x_i \alpha(\hat{\theta}) - x_k \alpha(\hat{\theta})}{b_n})$$

with $x\alpha(\hat{\theta}) = s + z_1 \hat{\alpha}_1 + w \hat{\alpha}_2$, is a wighted function. This estimator is derived from the approaches suggested in Robinson [1988], Powell [1987] and Ichimura and Lee [1988].

The semiparametric scoring estimates (SPSCORE) are derived with the above initial estimates. The kernel function $K(\cdot)$ for estimation of the choice probability is the univariate function $K_4(\cdot)$ in (5.8) and the bandwidth parameter $b_n$ is chosen with the rate in (5.9). The bivariate kernel function $\bar{K}(\cdot)$ for estimating the conditional density function in (2.11) is taken as the product of the univariate kernel function $K_4(\cdot)$ in (5.8) and the corresponding bandwidth parameter is chosen with the rate [5]:

$$a_n = \frac{c_2}{n^{1/7}}. \quad (5.12)$$

In Table 1, we report results for the two stage semiparametric estimates and the scoring estimates with sample sizes 100, 300 and 500. We will compare them with some parametric estimates. The distributions of the disturbance $\epsilon_2$ are generated from the bounded distribution as in (5.3) or from the normal distribution in (5.4). The disturbances $\epsilon_1$ in both models are generated from the same standard normal distribution. These models are also estimated by the biased corrected probit-least square two stage procedure (PROBIT-NMTWOE) and the parametric maximum likelihood procedure with bivariate normal distribution (NMMLE). For the normal distribution model, these parametric procedures are consistent and the NMMLE is asymptotic efficient. The values of both the bandwidth factors $c_1$ in (5.9) and $c_2$ in (5.12) are set to 5. Throughtout simulation, results reported in this table and other tables are based on 200 repetitions for each case. Summary statistics, namely, mean, standard deviation (SD) and root mean squared error (RMSE), are reported. For both models, the magnitudes of the biases are reasonably small. For sample size 100, the biases of the estimates for the choice equation are about 5.2% to 6.4% and the largerst bias for the outcome equation is 9.2%. As sample sizes increase, the biases decrease. At sample size 500, the largest bias of the estimates for the choice equation is 2.8% and the the largest bias for the outcome equation is 1.5%. The biases of the NMMLE for the outcome equation are slightly smaller in magnitude than the biases of the corresponding SPSCORE. However, the biases of the SPSCORE for the choice equation are smaller than the biases of the NMMLE. The biases of the initial consistent semiparametric estimates are also reasonably small. The single equation SMLE of

[5] To handle the empty space problem, when the kernel estimates of the functions in the denominators of (2.14) or (2.15) are zero, they are set to arbitrarily small values. Asymptotically, such events occur with zero probability.

the choice equation are the same for the two models because they have the same choice equation. For sample size 100, the biases of the SMLE for the choice equation are from 8% to 9.6% which are reduced to less than 4% for sample size 500. The biases of the SMLE are relatively larger than the biases of the corresponding SPSCORE estimates. The largest bias of the SPTWOE of the outcome equation for the normal distribution model is 5.7% for size 100 and is about 2% for size 500. The OLS estimates of the outcome equation which ignores the sample selection bias have the largest biases among all these estimates. The biases of the estimates of $\gamma_2$ are more than 20% for all the three sample sizes. The biases for the estimates of $\gamma_1$ are apparently small. This is so, since the variable $z_2$ in the outcome equation is independent with all the regressors in the choice equation. For the bounded distribution model, the parametric maximum likelihood procedure as well the bias corrected two stage estimation of the outcome equation are subject to small degree of misspecfication. The NMMLE of $\gamma_2$ for the bounded distribution model have provided some evidences of misspecification biases which do not seem to be reduced as sample size increases. However, the bias is small and only about 1.5%. The NMTWOE seem to be more robust. More comparison between the estimates of these two models will be reported later. The standard deviations of the semiparametric estimates are all larger than the corresponding parametric estimates. For the choice equation, the largest ratio of the SD of the SPSCORE over the SD of the NMMLE is about 2 to 1. For the outcome equation, the largest ratio of the SD is about 1.54 to 1. The SPSCORE of the choice equation have smaller variances than the single equation SMLE for all the three sample sizes. For estimation of the outcome equation, the SPSCORE have shown improvement upon the SPTWOE only for the larger samples. As expected, the variances of all the estimates decrease as sample sizes increase. As the biases for these estimates are relatively small and the variance component in the RMSE dominates the bias component, the semiparametric estimates have relatively larger RMSE than the parametric estimates. Overall, the SPSCORE have smaller RMSE over the initial consistent semiparametric estimates for larger sample sizes for both models.

In Table 2, we report estimates for the model with mixed gamma and normal disturbances. For this model, the parametric likelihood functions based on the normal distribution are misspecified. However, the biases of the PROBIT estimates and the NMMLE are small. Except for the estimation of the parameter $\alpha_1$, the biases are compatible with the biases of the other two models. The biases of the NMMLE and the PROBIT estimate of $\alpha_1$ are respectively 1.3% and 2.7% larger than the biases for the normal model with sample size 500. The biases of the SPSCORE for the choice equation on average tend to be smaller than the parametric estimates. However, the parametric estimates have smaller variances and their RMSE are still less than the semiparametric estimates with the three sample sizes.

To investigate the possible sensitivity of the semiparametric estimates with the bandwidth parameter, estimates with bandwidth factors 1, 2.5, 7.5, 10 and 15 in addition to the factor 5 are also computed for the bounded distribution model and the normal distribution model. The results for estimation of the bounded distribution model are reported in Table 3 and Table 4. Table 3 reports the results for the semiparametric two stage estimation and Table 4 reports the results for the semiparametric scoring estimation [6]. The estimates in both tables have shown sensitivity to the values of the bandwidth factors. Some patterns emerge from these estimates. The smallest biases for both semiparametric estimates tend to occur around factor values 2.5 and 5. The magnitudes of the biases become larger with larger bandwidth factors. The biases of

[6] For sample size 300, the set of SPTWOE with bandwidth factor 1 in Table 3 contains an outliner with estimated value -28.7 for $\gamma_1$. The corresponging SPSCORE in Table 4 inherit the same outlier. By dropping the outlier, the mean vector in Table 3 becomes (1.1167, -.9884, 1.0440, -.9702) with SD (.4769, .2924, .6404, .1715) and the revised mean vector in Table 4 is (1.1165, -.9884, 1.0444, -.9702) with SD (.4760, .2924, .6400, .1714).

the estimates with factors 1, 2.5 and 5 are much smaller than the biases with larger factors. As sample sizes increase, the biases are reduced. The bias reductions are quite substantial for the larger bandwith factors. The degree of sensitivity to the values of the bandwidth factor decreases as sample size increases as expected from our theoretical result. Comparing the estimates across the two tables, both the SPSCORE and the SMLE of the choice equation are sensitive to the bandwidth factors. The SPSCORE are slightly less sensitive to the bandwidth factors than the corresponding SMLE. System estimation provides more information for estimation of the choice equation. However, for the outcome equation, the SPSCORE of $\gamma_2$ are sensitive to the bandwidth factors but the SPTWOE are not. This might be due to the fact that bivariate kernel function is needed for the scoring procedure and only univariate kernel function is needed for nonparametric regression estimation. It is interesting to note that both the SPSCORE and the SPTWOE of $\gamma_1$ are not sensitive to the bandwidth factors which might be due to the property that the variable $z_2$ is independent with all the regressors in the choice equation. The SD of the SPSCORE for both the choice and outcome equations and the SD of the SMLE for the choice equation are also sensitive the bandwidth factors. This is so, especially for sample size 100. The SD of these estimates are smaller along the section of bandwidth factors with values 2.5 to 7.5 than the SD with the larger or the smaller bandwidth factors. This happens also for the SD of the SPSCORE of $\gamma_1$. The independency of $z_2$ with the other regressors does not seem helpful here. The SD of the SPTWOE are not sensitive to the bandwidth factors except that there is an outlier appeared in the estimates of $\gamma_1$ when the bandwidth factor is 1 [7]. Anyhow, as sample sizes increase, the degree of sensitivity is reduced. This is so, especially for the semiparametric estimates of the choice equation. At sample sizes of 300 and 500, the SD of both the SPSCORE and the SMLE have shown only some small differences relative to the biases. Combining the biases and the SD, the RMSE show the combined tendency of these components. The RMSE tend to be larger when the bandwidth factors are large. When the bandwidth factor is too small, the RMSE will also increase. The degree of sensitivity decreases as sample size increases.

The sensitivity of the semiparametric estimates with the bandwidth factors in these models raises question on how to choose the proper bandwidth factors in practice with moderate sample sizes. In the statistics literature, there are suggestions such as the use of cross validation procedure (Silverman [1986]). However, such procedure is designded for density estimation and whether it would be useful for the purpose of coefficient estimation in semiparametric models is questionable. For kernel density estimation, the statistics literature points out that small bandwidth introduces large variance but small bias, and large bandwidth will introduce smaller variance but larger bias. However, for estimation of our models, the evidences from Table 3 and Table 4 reveal different patterns. An intuitive procedure that seems to be useful for our model is based on the goodness-of-fit criterion. For each estimation, the log profile likelihood function at the semiparametric estimates can be computed. The columns with the label ALK report the mean values of the average maximized log profile likelihood funtions for each factor. In Table 3, the averaged profile likelihood function evaluated is the function in (5.7) for the binary choice model. In Table 5, the profile likelihood function is for the whole model. It is interesting to see that the models with the factors around 5 or 7.5 on average provide the better fitted models. For each sample, there are six estimates corresponding to six different bandwidth factors. In the rows marked 'Max', we report the summary statistics for the estimator which provides the largest profile likelihood function among the six ones. This 'Max' estimator has the same asymptotic properties as the corresponding semiparametric estimators with fixed bandwidth factors if the selected bandwidth factors are not data dependent. This selection procedure seems encouraging. The biases of the 'Max' SPSCORE are only slightly larger than the best ones with fixed bandwidth factors. The

[7] See previous footnote.

SD of these estimates for the outcome equation are slightly larger than the SD of the estimates with small fixed bandwidth factors for sample size 100. Otherwise, the resultant estimates are compatible with the good ones with fixed factors. In Table 4, we have also computed the average values for the six different estimates with various bandwidth factors. The results are reported in the row marked 'Ave'. These average estimates are not better than the 'max' estimates. An alternative procedure that seems also useful is to use the same factor corresponding to the best fitted SMLE to derive the SPSCORE. The results are reported on the row marked 'C. Max' in Table 4. These estimates have even small biases than the 'max' estimates for sample sizes 300 and 500.

Results for estimation of the normal distribution model with various bandwidth factors by the scoring procedure are reported in Table 5 [8]. The pattern of sensitivity of the SPSCORE to the bandwidth factors is similar to the SPSCORE of the bounded distribution model in Table 4 [9]. Comparing the estimates of the two models, there are only slight differences. For the choice equation, magnitudes of the biases of the estimates for the normal distribution model are slightly larger. For bandwidth factors 7.5 or larger, the RMSE of the estimates of the normal distribution are also larger. On the contrary, for estimation of the outcome equation, the SD and the RMSE of the estimates of the normal distribution model are smaller with such bandwidth factors. At small bandwidth factors, the differences are really small. For the normal distribution model, the density function is not strictly bounded away from zero that violates one of the regularity conditions for our anlysis. From this experiment, we may conclude that this violation does not have significant impact on practice. This would indeed be the case for estimation with finite samples because that the density is strictly bounded away from zero is not a testable hypothesis [10].

The SPSCORE from the above simulations are derived from the design that the bandwidth rate $b_n$ is of order $O(\frac{1}{n^{1/5.5}})$ for the nonparametric function estimation of the choice equation component and the rate $a_n$ is of order $O(\frac{1}{n^{1/7}})$ for estimation of the outcome equation component. To investigate more the role of the bandwidth rates on estimation of this model, we have experimented with two more designs for the bounded distribution model. The results in Table 6 are derived with both the bandwidth rates $a_n$ and $b_n$ equal to $\frac{c}{n^{1/7}}$. With the same factor value for $c$, the latter design has a larger bandwidth for estimating the choice equation component than the previous design but the bandwidth rate for estimating the outcome equation component is the same. The results in Table 6 should be compared with the SPSCORE in Table 4 [11]. For small bandwidth factors 1, 2.5 and 5, the estimates are in general similar and the differences are small. However, for larger bandwidth factors, the SPSCORE in Table 6 for the choice equation have larger biases than the SPSCORE in Table 4 and the SD also tend to be larger. The estimates of $\gamma_2$ for the

[8] To save space, the semiparametric two stage estimates of the outcome equation are not reported here.

[9] For sample size 300, the set of SPSCORE with bandwidth factor 1 contains a similar outlier as in Table 4. By dropping the outlier, the revised mean vector for the summary statistics is (1.1167, -.9883, 1.0461, -.9736) with SD (.4764, .2921 .6360, .1798).

[10] It might be possible that alternative analysis would not require such assumption. But for our anlysis, it is not obvious how this assumption could be relaxed. Bickel's analysis [1982] for the linear regression model does not require such assumption but it requires sample splitting. With sample splitting, the adaptive estimates behave much worse than the estimates without splitting in the Monte Carlo results in Manski[1984]. Our estimation procedure does not require sample splitting.

[11] By dropping the outlier inherited from the initial estimates in Table 3, the revised summary statistics for the SPSCORE with bandwidth factor 1 and sample size 300 have mean vector (1.1166, -.9885, 1.0444, -.9701) with SD (.4759, .2923, .6400, .1714)

outcome equation, on the contrary, have smaller biases and the SD also tend to be smaller. The RMSE comparison possesses the same pattern. This design with slower bandwidth rate for the choice component is a two edged sword. From the RMSE error comparsion, the damages to the estimates of the choice eqation are more severe than the benefits to the estimates of the outcome equation.

Table 7 reports the results for the design that the rates of $a_n$ and $b_n$ are the same as in the first design but the bandwidth factors $c_1$ and $c_2$ in $b_n$ and $a_n$ respectively can be different from each other. The SPSCORE are derived by using the 'Max' estimates of the choice equation and the corresponding SPTWOE for the outcome equation ,i.e., the estimates underlying the summary statistics in the last row of Table 3, as initial consistent estimates. For each sample, the factor $c_1$ corresponds to the bandwidth factor in the SMLE estimates but the value of $c_2$ is allowed to vary. As one might expect, the SPSCORE for the choice equation do not show much sensitivity to the values of $c_2$. The interest here is to see whether the estimates for the outcome equation can improve upon the corresponding estimates on Table 5. For small values of $c_2$, the estimates are again not much different. For large values of $c_2$, the estimates of $\gamma_2$ in Table 7 have larger biases. Even thought the corresponding SD of the estimates of $\gamma_2$ tend to be smaller, the SD of the estimates of $\gamma_1$ tend to be larger. The resultant RMSE of the estimates for the outcome equation in Table 7 are larger than the RMSE of the estimates in Table 5. The 'Max' estimates in Table 7 does not improve upon the 'Max' estimates in Table 5. Indeed, it is worse than the 'C. Max' estimates in the last row of Table 4. In conclusion, these two different designs do not provide better results than the results obtained from the first design.

## TABLE 1.

Parametric and Semiparametric estimates : Comparison

| Method | N | Size 100 Mean | SD | RMSE | Size 300 Mean | SD | RMSE | Size 500 Mean | SD | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Bounded Distribution | | | | | |
| OLS | $\gamma_1$ | 1.009 | .279 | .278 | .996 | .177 | .177 | 1.002 | .141 | .141 |
| | $\gamma_2$ | -.786 | .201 | .293 | -.794 | .101 | .229 | -.797 | .079 | .218 |
| PROBIT | $\alpha_1$ | 1.111 | .386 | .400 | 1.039 | .203 | .206 | 1.028 | .148 | .150 |
| | $\alpha_2$ | -1.132 | .248 | .280 | -1.044 | .131 | .138 | -1.028 | .095 | .099 |
| NMTWOE | $\gamma_1$ | 1.006 | .282 | .281 | .993 | .173 | .173 | .999 | .136 | .136 |
| | $\gamma_2$ | -.965 | .240 | .241 | -.992 | .126 | .126 | -.994 | .098 | .098 |
| NMMLE | $\alpha_1$ | 1.096 | .404 | .413 | 1.046 | .196 | .201 | 1.037 | .142 | .147 |
| | $\alpha_2$ | -1.127 | .274 | .301 | -1.044 | .134 | .141 | -1.028 | .095 | .099 |
| | $\gamma_1$ | 1.011 | .305 | .304 | .995 | .172 | .172 | 1.000 | .135 | .135 |
| | $\gamma_2$ | -.978 | .251 | .251 | -.984 | .124 | .125 | -.985 | .093 | .094 |
| SMLE | $\alpha_1$ | 1.096 | .619 | .624 | 1.046 | .288 | .291 | 1.039 | .264 | .266 |
| | $\alpha_2$ | -1.079 | .453 | .457 | -1.036 | .225 | .227 | -1.031 | .196 | .198 |
| SPTWOE | $\gamma_1$ | 1.010 | .342 | .341 | 1.000 | .225 | .225 | .979 | .222 | .223 |
| | $\gamma_2$ | -.958 | .289 | .291 | -.986 | .171 | .171 | -.993 | .169 | .169 |
| SPSCORE | $\alpha_1$ | 1.064 | .609 | .609 | 1.019 | .285 | .285 | 1.017 | .248 | .248 |
| | $\alpha_2$ | -1.056 | .430 | .431 | -1.012 | .211 | .211 | -1.010 | .185 | .185 |
| | $\gamma_1$ | .997 | .471 | .469 | .983 | .220 | .220 | .992 | .165 | .165 |
| | $\gamma_2$ | -.908 | .352 | .362 | -.983 | .169 | .170 | -.985 | .133 | .134 |
| | | | | | Normal Distribution | | | | | |
| OLS | $\gamma_1$ | 1.035 | .300 | .301 | 1.004 | .192 | .192 | 1.004 | .139 | .139 |
| | $\gamma_2$ | -.730 | .211 | .342 | -.744 | .110 | .279 | -.749 | .085 | .265 |
| PROBIT | $\alpha_1$ | 1.111 | .386 | .400 | 1.039 | .203 | .206 | 1.028 | .148 | .150 |
| | $\alpha_2$ | -1.132 | .248 | .280 | -1.044 | .131 | .138 | -1.028 | .095 | .099 |
| NMTWOE | $\gamma_1$ | 1.035 | .293 | .294 | 1.000 | .187 | .187 | 1.000 | .132 | .132 |
| | $\gamma_2$ | -.959 | .250 | .252 | -.998 | .134 | .134 | -.999 | .104 | .104 |
| NMMLE | $\alpha_1$ | 1.086 | .416 | .423 | 1.039 | .190 | .194 | 1.033 | .140 | .144 |
| | $\alpha_2$ | -1.125 | .279 | .304 | -1.040 | .129 | .135 | -1.028 | .092 | .096 |
| | $\gamma_1$ | 1.040 | .334 | .335 | .998 | .188 | .188 | 1.002 | .133 | .133 |
| | $\gamma_2$ | -.986 | .256 | .255 | -1.000 | .130 | .130 | -.999 | .099 | .099 |
| SMLE | $\alpha_1$ | 1.096 | .619 | .624 | 1.046 | .288 | .291 | 1.039 | .264 | .266 |
| | $\alpha_2$ | -1.079 | .453 | .457 | -1.036 | .225 | .227 | -1.031 | .196 | .198 |
| SPTWOE | $\gamma_1$ | 1.028 | .365 | .365 | .996 | .226 | .226 | .979 | .220 | .221 |
| | $\gamma_2$ | -.943 | .301 | .305 | -.986 | .189 | .189 | -.986 | .160 | .160 |
| SPSCORE | $\alpha_1$ | 1.053 | .609 | .609 | 1.023 | .280 | .280 | 1.028 | .251 | .252 |
| | $\alpha_2$ | -1.052 | .442 | .443 | -1.016 | .212 | .212 | -1.015 | .186 | .186 |
| | $\gamma_1$ | 1.043 | .440 | .440 | .995 | .217 | .217 | .990 | .163 | .163 |
| | $\gamma_2$ | -.915 | .352 | .360 | -.980 | .176 | .177 | -.987 | .137 | .137 |

## TABLE 2.

Parametric and Semiparametric estimates : Comparison
Mixed Gamma-Normal Distribution

| Method | N | Size 100 Mean | SD | RMSE | Size 300 Mean | SD | RMSE | Size 500 Mean | SD | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | $\gamma_1$ | .995 | .250 | .249 | 1.007 | .136 | .136 | 1.005 | .110 | .110 |
| | $\gamma_2$ | -.751 | .163 | .297 | -.761 | .088 | .255 | -.762 | .065 | .247 |
| PROBIT | $\alpha_1$ | 1.115 | .366 | .382 | 1.060 | .206 | .214 | 1.055 | .157 | .166 |
| | $\alpha_2$ | -1.094 | .236 | .253 | -1.036 | .128 | .133 | -1.028 | .103 | .107 |
| NMTWOE | $\gamma_1$ | .990 | .241 | .240 | 1.002 | .130 | .130 | 1.003 | .106 | .106 |
| | $\gamma_2$ | -.993 | .211 | .210 | -1.004 | .103 | .103 | -1.004 | .078 | .078 |
| NMMLE | $\alpha_1$ | 1.113 | .379 | .394 | 1.049 | .195 | .201 | 1.046 | .146 | .153 |
| | $\alpha_2$ | -1.099 | .253 | .270 | -1.033 | .128 | .132 | -1.021 | .101 | .103 |
| | $\gamma_1$ | 1.000 | .252 | .251 | 1.000 | .126 | .126 | 1.001 | .103 | .103 |
| | $\gamma_2$ | -.997 | .197 | .196 | -.996 | .099 | .099 | -.998 | .075 | .075 |
| SMLE | $\alpha_1$ | 1.106 | .472 | .481 | 1.067 | .348 | .354 | 1.057 | .261 | .267 |
| | $\alpha_2$ | -1.085 | .380 | .388 | -1.041 | .329 | .331 | -1.041 | .206 | .210 |
| SPTWOE | $\gamma_1$ | .989 | .267 | .266 | 1.010 | .187 | .187 | 1.008 | .139 | .139 |
| | $\gamma_2$ | -.986 | .247 | .246 | -.985 | .172 | .172 | -.983 | .126 | .127 |
| SPSCORE | $\alpha_1$ | 1.094 | .442 | .450 | 1.052 | .313 | .317 | 1.027 | .222 | .223 |
| | $\alpha_2$ | -1.069 | .357 | .362 | -1.027 | .320 | .321 | -1.023 | .166 | .167 |
| | $\gamma_1$ | .949 | .377 | .379 | 1.011 | .156 | .156 | 1.001 | .112 | .112 |
| | $\gamma_2$ | -.951 | .374 | .375 | -.982 | .152 | .153 | -.997 | .102 | .102 |

**TABLE 3.**

**Initial Consistent Estimates with Various Bandwidth Factors: Bounded Distribution SMLE for Choice Equation and SPTWOE for Outcome Equation**

| | N | Size 100 | | | | Size 300 | | | | Size 500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| c | | Mean | SD | RMSE | ALK | Mean | SD | RMSE | ALK | Mean | SD | RMSE | ALK |
| 1.0 | $\alpha_1$ | 1.016 | .638 | .635 | -.421 | 1.119 | .477 | .491 | -.389 | 1.087 | .441 | .449 | -.381 |
| | $\alpha_2$ | -.822 | .372 | .411 | | -.989 | .292 | .292 | | -1.019 | .267 | .268 | |
| | $\gamma_1$ | .994 | .429 | .427 | | .895 | 2.198 | 2.197 | | .984 | .209 | .209 | |
| | $\gamma_2$ | -.919 | .340 | .348 | | -.961 | .218 | .221 | | -.974 | .126 | .128 | |
| 2.5 | $\alpha_1$ | 1.084 | .607 | .609 | -.363 | 1.082 | .404 | .412 | -.364 | 1.022 | .322 | .323 | -.364 |
| | $\alpha_2$ | -.987 | .432 | .430 | | -1.068 | .291 | .298 | | -1.024 | .195 | .196 | |
| | $\gamma_1$ | 1.005 | .339 | .337 | | .994 | .275 | .274 | | 1.003 | .216 | .216 | |
| | $\gamma_2$ | -.950 | .278 | .281 | | -.991 | .207 | .207 | | -.999 | .119 | .119 | |
| 5.0 | $\alpha_1$ | 1.096 | .619 | .624 | -.356 | 1.046 | .288 | .291 | -.360 | 1.039 | .264 | .266 | -.362 |
| | $\alpha_2$ | -1.079 | .453 | .457 | | -1.036 | .225 | .227 | | -1.031 | .196 | .198 | |
| | $\gamma_1$ | 1.010 | .342 | .341 | | 1.000 | .225 | .225 | | .979 | .222 | .223 | |
| | $\gamma_2$ | -.958 | .289 | .291 | | -.986 | .171 | .171 | | -.993 | .169 | .169 | |
| 7.5 | $\alpha_1$ | 1.225 | .539 | .581 | -.364 | 1.106 | .281 | .300 | -.363 | 1.099 | .280 | .297 | -.364 |
| | $\alpha_2$ | -1.242 | .390 | .457 | | -1.106 | .188 | .215 | | -1.098 | .203 | .226 | |
| | $\gamma_1$ | 1.006 | .308 | .307 | | .988 | .209 | .209 | | .999 | .157 | .157 | |
| | $\gamma_2$ | -.966 | .268 | .268 | | -.986 | .190 | .190 | | -.974 | .159 | .161 | |
| 10.0 | $\alpha_1$ | 1.496 | .690 | .847 | -.379 | 1.259 | .298 | .394 | -.371 | 1.222 | .295 | .369 | -.370 |
| | $\alpha_2$ | -1.563 | .536 | .776 | | -1.279 | .194 | .340 | | -1.208 | .217 | .301 | |
| | $\gamma_1$ | 1.011 | .303 | .302 | | .987 | .184 | .184 | | .994 | .152 | .152 | |
| | $\gamma_2$ | -.980 | .298 | .297 | | -.992 | .149 | .149 | | -.951 | .157 | .165 | |
| 15.0 | $\alpha_1$ | 2.410 | 1.713 | 2.212 | -.3986 | 1.791 | .406 | .889 | -.390 | 1.670 | .318 | .741 | -.388 |
| | $\alpha_2$ | -2.489 | 1.448 | 2.071 | | -1.835 | .252 | .872 | | -1.691 | .227 | .728 | |
| | $\gamma_1$ | 1.009 | .302 | .301 | | .988 | .190 | .190 | | .996 | .150 | .149 | |
| | $\gamma_2$ | -.954 | .324 | .325 | | -.999 | .180 | .180 | | -.972 | .172 | .174 | |
| Max. | $\alpha_1$ | 1.199 | 1.358 | 1.365 | -.3422 | 1.037 | .278 | .280 | -.355 | 1.017 | .253 | .253 | -.358 |
| | $\alpha_2$ | -1.180 | .942 | .954 | | -1.046 | .208 | .213 | | -1.025 | .168 | .170 | |
| | $\gamma_1$ | 1.020 | .340 | .339 | | .991 | .245 | .244 | | .978 | .220 | .220 | |
| | $\gamma_2$ | -.967 | .312 | .312 | | -.993 | .187 | .187 | | -.979 | .115 | .117 | |

**TABLE 4.**

Scoring Estimates with Various Bandwidth Factors: Bounded Distribution

Design: Two Bandwidth Rates and One Factor

| | N | Size 100 | | | | Size 300 | | | | Size 500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| c | | Mean | SD | RMSE | ALK | Mean | SD | RMSE | ALK | Mean | SD | RMSE | ALK |
| 1.0 | $\alpha_1$ | 1.014 | .637 | .634 | -5.154 | 1.119 | .476 | .490 | -3.639 | 1.086 | .440 | .448 | -2.872 |
| | $\alpha_2$ | -.820 | .372 | .412 | | -.989 | .292 | .292 | | -1.018 | .267 | .267 | |
| | $\gamma_1$ | .993 | .430 | .428 | | .896 | 2.196 | 2.195 | | .984 | .209 | .209 | |
| | $\gamma_2$ | -.918 | .341 | .349 | | -.961 | .218 | .221 | | -.974 | .126 | .129 | |
| 2.5 | $\alpha_1$ | 1.057 | .592 | .592 | -1.893 | 1.062 | .399 | .403 | -1.509 | 1.005 | .321 | .321 | -1.387 |
| | $\alpha_2$ | -.968 | .430 | .429 | | -1.054 | .287 | .292 | | -1.010 | .193 | .193 | |
| | $\gamma_1$ | 1.012 | .342 | .340 | | .992 | .265 | .265 | | 1.004 | .204 | .204 | |
| | $\gamma_2$ | -.943 | .290 | .294 | | -.991 | .203 | .203 | | -.999 | .120 | .120 | |
| 5.0 | $\alpha_1$ | 1.064 | .609 | .609 | -1.311 | 1.019 | .285 | .285 | -1.230 | 1.017 | .248 | .248 | -1.225 |
| | $\alpha_2$ | -1.056 | .430 | .431 | | -1.012 | .211 | .211 | | -1.010 | .185 | .185 | |
| | $\gamma_1$ | .997 | .471 | .469 | | .983 | .220 | .220 | | .992 | .165 | .165 | |
| | $\gamma_2$ | -.908 | .352 | .362 | | -.983 | .169 | .170 | | -.985 | .133 | .134 | |
| 7.5 | $\alpha_1$ | 1.142 | .503 | .520 | -1.292 | 1.078 | .278 | .288 | -1.208 | 1.077 | .247 | .258 | -1.208 |
| | $\alpha_2$ | -1.147 | .340 | .369 | | -1.073 | .189 | .202 | | -1.064 | .191 | .201 | |
| | $\gamma_1$ | .990 | .717 | .713 | | .984 | .272 | .272 | | .994 | .192 | .192 | |
| | $\gamma_2$ | -.798 | .590 | .621 | | -.935 | .214 | .223 | | -.965 | .168 | .171 | |
| 10.0 | $\alpha_1$ | 1.374 | .733 | .820 | -1.360 | 1.231 | .282 | .364 | -1.234 | 1.186 | .270 | .328 | -1.220 |
| | $\alpha_2$ | -1.394 | .514 | .646 | | -1.194 | .168 | .256 | | -1.151 | .203 | .253 | |
| | $\gamma_1$ | .989 | 1.192 | 1.186 | | 1.002 | .381 | .380 | | .998 | .255 | .255 | |
| | $\gamma_2$ | -.560 | 1.011 | 1.098 | | -.782 | .299 | .370 | | -.878 | .256 | .283 | |
| 15.0 | $\alpha_1$ | 2.130 | 1.601 | 1.953 | -1.600 | 1.592 | .360 | .693 | -1.383 | 1.494 | .286 | .571 | -1.334 |
| | $\alpha_2$ | -2.139 | 1.417 | 1.812 | | -1.525 | .242 | .578 | | -1.473 | .237 | .529 | |
| | $\gamma_1$ | 1.115 | 2.391 | 2.382 | | 1.020 | .769 | .768 | | 1.009 | .481 | .481 | |
| | $\gamma_2$ | -.219 | 2.463 | 2.572 | | -.300 | .858 | 1.106 | | -.590 | .727 | .834 | |
| Ave. | $\alpha_1$ | 1.297 | .575 | .645 | | 1.183 | .251 | .310 | | 1.144 | .218 | .261 | |
| | $\alpha_2$ | -1.254 | .416 | .486 | | -1.141 | .157 | .211 | | -1.121 | .146 | .190 | |
| | $\gamma_1$ | 1.016 | .749 | .745 | | .979 | .464 | .464 | | .997 | .196 | .196 | |
| | $\gamma_2$ | -.724 | .659 | .711 | | -.825 | .224 | .284 | | -.898 | .178 | .205 | |
| Max. | $\alpha_1$ | 1.074 | .489 | .492 | -1.226 | 1.048 | .263 | .267 | -1.191 | 1.061 | .233 | .241 | -1.194 |
| | $\alpha_2$ | -1.109 | .406 | .418 | | -1.047 | .194 | .199 | | -1.051 | .187 | .194 | |
| | $\gamma_1$ | .990 | .592 | .589 | | .978 | .248 | .249 | | .999 | .198 | .198 | |
| | $\gamma_2$ | -.885 | .513 | .523 | | -.946 | .194 | .201 | | -.947 | .164 | .172 | |
| C. | $\alpha_1$ | 1.147 | 1.262 | 1.264 | -2.343 | 1.023 | .283 | .283 | -1.562 | 1.005 | .249 | .249 | -1.404 |
| Max. | $\alpha_2$ | -1.124 | .856 | .861 | | -1.032 | .203 | .205 | | -1.007 | .161 | .161 | |
| | $\gamma_1$ | 1.013 | .586 | .583 | | .988 | .252 | .252 | | .992 | .189 | .189 | |
| | $\gamma_2$ | -.904 | .438 | .446 | | -.974 | .179 | .181 | | -.984 | .131 | .132 | |

TABLE 5.

Scoring Estimates: Normal Disturbances

Design: Two Bandwidth Rates and One Factor

| | N | Size 100 | | | | Size 300 | | | | Size 500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| c | | Mean | SD | RMSE | ALK | Mean | SD | RMSE | ALK | Mean | SD | RMSE | ALK |
| 1.0 | $\alpha_1$ | 1.014 | .639 | .636 | -5.241 | 1.119 | .476 | .490 | -3.821 | 1.087 | .441 | .449 | -3.053 |
| | $\alpha_2$ | -.822 | .372 | .411 | | -.989 | .291 | .291 | | -1.019 | .267 | .267 | |
| | $\gamma_1$ | 1.019 | .439 | .437 | | 1.222 | 2.562 | 2.567 | | .981 | .203 | .204 | |
| | $\gamma_2$ | -.888 | .360 | .375 | | -.984 | .229 | .229 | | -.971 | .153 | .156 | |
| 2.5 | $\alpha_1$ | 1.066 | .602 | .603 | -2.034 | 1.067 | .397 | .402 | -1.635 | 1.011 | .319 | .319 | -1.492 |
| | $\alpha_2$ | -.970 | .430 | .429 | | -1.055 | .284 | .289 | | -1.013 | .196 | .196 | |
| | $\gamma_1$ | 1.031 | .362 | .362 | | 1.003 | .331 | .330 | | 1.005 | .215 | .215 | |
| | $\gamma_2$ | -.925 | .306 | .314 | | -.997 | .243 | .243 | | -.999 | .140 | .140 | |
| 5.0 | $\alpha_1$ | 1.053 | .609 | .608 | -1.386 | 1.023 | .280 | .280 | -1.301 | 1.028 | .251 | .252 | -1.289 |
| | $\alpha_2$ | -1.052 | .442 | .443 | | -1.016 | .212 | .212 | | -1.015 | .186 | .186 | |
| | $\gamma_1$ | 1.043 | .440 | .440 | | .995 | .217 | .217 | | .990 | .163 | .163 | |
| | $\gamma_2$ | -.915 | .352 | .360 | | -.980 | .176 | .177 | | -.987 | .137 | .137 | |
| 7.5 | $\alpha_1$ | 1.156 | .545 | .564 | -1.328 | 1.085 | .287 | .299 | -1.266 | 1.093 | .246 | .263 | -1.265 |
| | $\alpha_2$ | -1.154 | .361 | .391 | | -1.085 | .188 | .206 | | -1.079 | .189 | .205 | |
| | $\gamma_1$ | 1.037 | .595 | .593 | | .989 | .247 | .247 | | .990 | .178 | .178 | |
| | $\gamma_2$ | -.794 | .522 | .559 | | -.937 | .180 | .190 | | -.958 | .145 | .151 | |
| 10.0 | $\alpha_1$ | 1.397 | .677 | .782 | -1.381 | 1.256 | .286 | .383 | -1.273 | 1.201 | .272 | .338 | -1.262 |
| | $\alpha_2$ | -1.405 | .521 | .658 | | -1.225 | .179 | .287 | | -1.177 | .206 | .271 | |
| | $\gamma_1$ | 1.066 | 1.005 | 1.002 | | .983 | .313 | .313 | | .992 | .218 | .218 | |
| | $\gamma_2$ | -.569 | .887 | .982 | | -.789 | .249 | .326 | | -.886 | .195 | .226 | |
| 15.0 | $\alpha_1$ | 2.130 | 1.581 | 1.937 | -1.607 | 1.610 | .355 | .705 | -1.396 | 1.522 | .289 | .597 | -1.348 |
| | $\alpha_2$ | -2.154 | 1.407 | 1.814 | | -1.561 | .247 | .613 | | -1.504 | .231 | .554 | |
| | $\gamma_1$ | 1.186 | 2.169 | 2.166 | | .977 | .639 | .638 | | 1.010 | .379 | .379 | |
| | $\gamma_2$ | -.130 | 2.351 | 2.496 | | -.305 | .838 | 1.088 | | -.617 | .659 | .762 | |
| Ave. | $\alpha_1$ | 1.303 | .570 | .643 | | 1.193 | .250 | .316 | | 1.157 | .219 | .269 | |
| | $\alpha_2$ | -1.259 | .420 | .492 | | -1.155 | .157 | .220 | | -1.134 | .147 | .199 | |
| | $\gamma_1$ | 1.063 | .657 | .657 | | 1.028 | .489 | .489 | | .995 | .168 | .168 | |
| | $\gamma_2$ | -.703 | .604 | .670 | | -.832 | .204 | .264 | | -.903 | .155 | .183 | |
| Max. | $\alpha_1$ | 1.119 | .588 | .597 | -1.271 | 1.108 | .311 | .329 | -1.244 | 1.094 | .242 | .259 | -1.245 |
| | $\alpha_2$ | -1.124 | .393 | .410 | | -1.092 | .216 | .231 | | -1.081 | .179 | .196 | |
| | $\gamma_1$ | 1.075 | .570 | .572 | | .984 | .240 | .240 | | .993 | .176 | .176 | |
| | $\gamma_2$ | -.847 | .466 | .488 | | -.930 | .211 | .222 | | -.934 | .155 | .168 | |
| C. | $\alpha_1$ | 1.141 | 1.265 | 1.267 | -2.387 | 1.021 | .273 | .273 | -1.655 | 1.005 | .240 | .240 | -1.494 |
| Max. | $\alpha_2$ | -1.129 | .855 | .860 | | -1.035 | .203 | .206 | | -1.012 | .161 | .161 | |
| | $\gamma_1$ | 1.036 | .610 | .608 | | 1.000 | .233 | .233 | | .994 | .176 | .176 | |
| | $\gamma_2$ | -.888 | .459 | .470 | | -.974 | .181 | .183 | | -.983 | .126 | .127 | |

**TABLE 6.**

Scoring Estimates with Various Bandwidth Factors II: Bounded Distribution

Design: One Bandwidth Rate and One Factor

| c | N | Size 100 Mean | SD | RMSE | ALK | Size 300 Mean | SD | RMSE | ALK | Size 500 Mean | SD | RMSE | ALK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.0 | $\alpha_1$ | 1.014 | .638 | .635 | -4.983 | 1.119 | .476 | .489 | -3.545 | 1.086 | .440 | .448 | -2.807 |
| | $\alpha_2$ | -.821 | .372 | .411 | | -.989 | .292 | .291 | | -1.018 | .267 | .267 | |
| | $\gamma_1$ | .993 | .430 | .427 | | .896 | 2.198 | 2.197 | | .984 | .209 | .209 | |
| | $\gamma_2$ | -.918 | .341 | .349 | | -.961 | .218 | .221 | | -.974 | .126 | .129 | |
| 2.5 | $\alpha_1$ | 1.063 | .593 | .593 | -1.849 | 1.063 | .392 | .396 | -1.486 | 1.004 | .315 | .314 | -1.373 |
| | $\alpha_2$ | -.981 | .430 | .428 | | -1.057 | .284 | .289 | | -1.010 | .190 | .190 | |
| | $\gamma_1$ | 1.012 | .342 | .340 | | .992 | .265 | .265 | | 1.004 | .204 | .204 | |
| | $\gamma_2$ | -.945 | .290 | .294 | | -.991 | .202 | .202 | | -.997 | .120 | .120 | |
| 5.0 | $\alpha_1$ | 1.128 | .598 | .609 | -1.292 | 1.022 | .250 | .251 | -1.226 | 1.026 | .216 | .217 | -1.219 |
| | $\alpha_2$ | -1.144 | .420 | .442 | | -1.020 | .181 | .181 | | -1.012 | .163 | .163 | |
| | $\gamma_1$ | .997 | .461 | .459 | | .980 | .223 | .223 | | .995 | .166 | .166 | |
| | $\gamma_2$ | -.924 | .360 | .366 | | -.988 | .172 | .172 | | -.986 | .132 | .133 | |
| 7.5 | $\alpha_1$ | 1.464 | .598 | .754 | -1.281 | 1.260 | .265 | .371 | -1.210 | 1.205 | .228 | .306 | -1.205 |
| | $\alpha_2$ | -1.482 | .348 | .593 | | -1.249 | .154 | .293 | | -1.188 | .145 | .237 | |
| | $\gamma_1$ | .997 | .706 | .703 | | .986 | .278 | .278 | | .990 | .192 | .192 | |
| | $\gamma_2$ | -.842 | .582 | .601 | | -.972 | .212 | .214 | | -.985 | .161 | .162 | |
| 10.0 | $\alpha_1$ | 1.942 | .922 | 1.315 | -1.355 | 1.702 | .352 | .785 | -1.246 | 1.578 | .279 | .641 | -1.230 |
| | $\alpha_2$ | -1.960 | .548 | 1.104 | | -1.654 | .210 | .687 | | -1.540 | .190 | .573 | |
| | $\gamma_1$ | .998 | 1.149 | 1.143 | | .999 | .387 | .386 | | .994 | .252 | .252 | |
| | $\gamma_2$ | -.639 | .994 | 1.053 | | -.850 | .310 | .344 | | -.943 | .250 | .256 | |
| 15.0 | $\alpha_1$ | 3.025 | 1.790 | 2.697 | -1.571 | 2.465 | .550 | 1.565 | -1.377 | 2.337 | .377 | 1.389 | -1.336 |
| | $\alpha_2$ | -3.062 | 1.444 | 2.513 | | -2.402 | .321 | 1.438 | | -2.251 | .264 | 1.279 | |
| | $\gamma_1$ | 1.161 | 2.316 | 2.310 | | 1.032 | .759 | .759 | | 1.004 | .458 | .457 | |
| | $\gamma_2$ | -.363 | 2.376 | 2.449 | | -.400 | .843 | 1.033 | | -.654 | .661 | .746 | |
| Ave. | $\alpha_1$ | 1.606 | .648 | .885 | | 1.438 | .280 | .520 | | 1.373 | .217 | .431 | |
| | $\alpha_2$ | -1.575 | .417 | .709 | | -1.395 | .162 | .427 | | -1.337 | .140 | .365 | |
| | $\gamma_1$ | 1.026 | .724 | .721 | | .981 | .465 | .465 | | .995 | .192 | .192 | |
| | $\gamma_2$ | -.772 | .643 | .679 | | -.860 | .229 | .268 | | -.923 | .172 | .188 | |
| Max. | $\alpha_1$ | 1.304 | .750 | .806 | -1.222 | 1.156 | .319 | .355 | -1.195 | 1.154 | .261 | .303 | -1.196 |
| | $\alpha_2$ | -1.316 | .525 | .611 | | -1.147 | .240 | .281 | | -1.142 | .230 | .270 | |
| | $\gamma_1$ | 1.036 | .620 | .618 | | .984 | .245 | .245 | | 1.000 | .186 | .186 | |
| | $\gamma_2$ | -.884 | .432 | .445 | | -.983 | .196 | .196 | | -.962 | .157 | .161 | |
| C. | $\alpha_1$ | 1.256 | 1.450 | 1.465 | -2.260 | 1.076 | .299 | .308 | -1.544 | 1.058 | .278 | .284 | -1.394 |
| Max. | $\alpha_2$ | -1.238 | .998 | 1.021 | | -1.083 | .241 | .255 | | -1.057 | .211 | .218 | |
| | $\gamma_1$ | 1.005 | .606 | .603 | | .986 | .253 | .253 | | .990 | .186 | .186 | |
| | $\gamma_2$ | -.929 | .425 | .429 | | -.984 | .179 | .179 | | -.990 | .126 | .126 | |

**TABLE 7.**

Scoring Estimates with Various Bandwidth Factors III: Bounded Distribution

Design: Two Bandwidth Rates and Two Factors

| c | N | Size 100 Mean | SD | RMSE | ALK | Size 300 Mean | SD | RMSE | ALK | Size 500 Mean | SD | RMSE | ALK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.0 | $\alpha_1$ | 1.198 | 1.359 | 1.367 | -5.214 | 1.037 | .278 | .280 | -3.467 | 1.017 | .253 | .253 | -2.773 |
| | $\alpha_2$ | -1.180 | .940 | .952 | | -1.046 | .208 | .213 | | -1.024 | .168 | .170 | |
| | $\gamma_1$ | 1.021 | .341 | .340 | | .991 | .244 | .244 | | .978 | .220 | .221 | |
| | $\gamma_2$ | -.967 | .313 | .313 | | -.992 | .187 | .187 | | -.979 | .115 | .117 | |
| 2.5 | $\alpha_1$ | 1.185 | 1.357 | 1.363 | -1.963 | 1.020 | .276 | .276 | -1.474 | 1.001 | .251 | .251 | -1.381 |
| | $\alpha_2$ | -1.157 | .942 | .950 | | -1.030 | .206 | .208 | | -1.010 | .167 | .167 | |
| | $\gamma_1$ | 1.029 | .342 | .342 | | .990 | .239 | .239 | | .977 | .216 | .217 | |
| | $\gamma_2$ | -.959 | .320 | .321 | | -.995 | .181 | .181 | | -.979 | .116 | .118 | |
| 5.0 | $\alpha_1$ | 1.163 | 1.378 | 1.381 | -1.359 | 1.026 | .279 | .280 | -1.241 | 1.000 | .240 | .240 | -1.227 |
| | $\alpha_2$ | -1.140 | .945 | .951 | | -1.027 | .201 | .202 | | -1.006 | .160 | .160 | |
| | $\gamma_1$ | .989 | .466 | .464 | | .978 | .226 | .227 | | .993 | .160 | .160 | |
| | $\gamma_2$ | -.944 | .370 | .372 | | -.990 | .188 | .188 | | -.987 | .124 | .125 | |
| 7.5 | $\alpha_1$ | 1.170 | 1.351 | 1.355 | -1.317 | 1.035 | .278 | .280 | -1.216 | 1.022 | .249 | .250 | -1.207 |
| | $\alpha_2$ | -1.138 | .920 | .926 | | -1.041 | .204 | .208 | | -1.028 | .163 | .165 | |
| | $\gamma_1$ | .990 | .804 | .800 | | .992 | .303 | .303 | | 1.003 | .200 | .200 | |
| | $\gamma_2$ | -.754 | .586 | .633 | | -.905 | .204 | .224 | | -.934 | .163 | .176 | |
| 10.0 | $\alpha_1$ | 1.168 | 1.355 | 1.359 | -1.388 | 1.033 | .274 | .276 | -1.241 | 1.022 | .253 | .254 | -1.219 |
| | $\alpha_2$ | -1.149 | .903 | .911 | | -1.043 | .206 | .210 | | -1.030 | .164 | .167 | |
| | $\gamma_1$ | 1.018 | 1.342 | 1.335 | | .991 | .452 | .451 | | 1.013 | .291 | .291 | |
| | $\gamma_2$ | -.364 | .945 | 1.135 | | -.637 | .308 | .476 | | -.732 | .225 | .350 | |
| 15.0 | $\alpha_1$ | 1.149 | 1.267 | 1.269 | -1.694 | 1.024 | .272 | .273 | -1.412 | 1.006 | .245 | .245 | -1.350 |
| | $\alpha_2$ | -1.145 | .851 | .859 | | -1.047 | .201 | .206 | | -1.027 | .159 | .161 | |
| | $\gamma_1$ | 1.112 | 3.107 | 3.093 | | .965 | .939 | .938 | | 1.085 | .629 | .634 | |
| | $\gamma_2$ | .961 | 2.210 | 2.946 | | .341 | .721 | 1.522 | | .051 | .510 | 1.168 | |
| Ave. | $\alpha_1$ | 1.172 | 1.340 | 1.344 | | 1.029 | .270 | .271 | | 1.011 | .244 | .244 | |
| | $\alpha_2$ | -1.151 | .913 | .921 | | -1.039 | .201 | .204 | | -1.021 | .160 | .161 | |
| | $\gamma_1$ | 1.026 | .898 | .894 | | .984 | .300 | .300 | | 1.008 | .205 | .205 | |
| | $\gamma_2$ | -.505 | .593 | .770 | | -.696 | .193 | .360 | | -.760 | .144 | .280 | |
| Max. | $\alpha_1$ | 1.156 | 1.278 | 1.281 | -1.253 | 1.030 | .276 | .277 | -1.202 | 1.009 | .243 | .243 | -1.197 |
| | $\alpha_2$ | -1.128 | .859 | .864 | | -1.034 | .208 | .210 | | -1.018 | .165 | .166 | |
| | $\gamma_1$ | .962 | .667 | .665 | | .980 | .273 | .273 | | 1.004 | .209 | .209 | |
| | $\gamma_2$ | -.851 | .468 | .489 | | -.931 | .229 | .239 | | -.921 | .166 | .184 | |

## 6. Uniform Convergence of Nonparametric Functions

The asymptotic properties of the estimator $\hat{\theta}_n$ in (2.13) depend on some convergence properties of the nonparametric functions $A_{n,l}(x_i,\theta)$, $B_n(x_i,\theta)$, $C_n(z_i,\theta)$ and their derivatives with $\theta$. The following propositions are useful. The proofs of Propositions 2, 3 and 4 are in Appendix 1.

**Proposition 1. (A Uniform Law of Large Numbers)** *Let $y_1,\cdots,y_n$ be $n$ independent observations drawn from a common probability distribution $P_{\theta_*}$. Suppose that the measurable functions $g_n(y,\beta,a_n)$ can be represented by the form:*

$$g_n(y,\beta,a_n) = \frac{1}{a_n^d} h_n(y,\beta,\frac{s_n(y,\beta)}{a_n})$$

*where $a_n = O(\frac{1}{n^p})$ with $p>0, \beta \in B, s_n(y,\beta)$ is a $\bar{m}$ dimensional vector value function and $d \geq \bar{m}$. Suppose that the following conditions are satisfied:*

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

*(ii) The functions $h_n(y,\beta,s)$ are bounded and satisfy a Lipschitz condition with respect to $\beta$ and $s$, uniformly in $n$. The function $s_n(y,\beta)$ satisfies also a Lipschitz condition with $\beta$, uniformly in $n$.*

*(iii) $E_{\theta_*}(h_n^2(y,\beta,\frac{s_n(y,\beta)}{a_n})) = O(a_n^{\bar{m}})$ uniformly in $\beta \in B$ and $n$.*

*If $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{2d-\bar{m}} = \infty$, then $\frac{1}{n}\sum_{i=1}^n g_n(y_i,\beta,a_n) - E_{\theta_*}(g_n(y,\beta,a_n)) \xrightarrow{P_{\theta_*}} 0$, uniformly in $\beta \in B$.*

**Proof:** This is a slightly modified case of a uniform law of large numbers in Ichimura [1987]. This result follows from Bernstein's inequality under $P_{\theta_*}$.

**Proposition 2.** *Let $K(w)$ be a kernel function on $R^{\bar{m}}$ with a bounded support $D$. For each $n$, let $T_{2,n}$ and $T_{1,n}$ denote the upper and lower limits of the values of the continuous random vector $t_n(z)$ in $R^{\bar{m}}$. Suppose that the density function $g_n(t)$ of $t_n(z)$ and the conditional expectation $E_{\theta_*}(c_n(z,z_i)|t_n(z)=t,z_i)$ are bounded and uniformly continuous in $t$ on the interior of $[T_{1,n},\ T_{2,n}]$, uniformly in $z_i$ and $n$. Then*

$$\lim_{n\to\infty} \sup_{z_i\in Z_n} |E_{\theta_*}(c_n(z,z_i)\frac{1}{a_n^{\bar{m}}}K(\frac{t_n(z_i)-t_n(z)}{a_n})|z_i) - E_{\theta_*}(c_n(z,z_i)|t_n(z_i),z_i)g_n(t_n(z_i))| = 0,$$

*where $Z_n = \{z|T_{1,n}+\delta_n \leq t_n(z) \leq T_{2,n}-\delta_n\}$ with $\delta_n > 0$ and $\lim_{n\to\infty}\frac{\delta_n}{a_n} = \infty$. Furthermore, if $K(w)$ is a kernel function with zero moments up to the order $s*$, i .e., $\int_D w_1^{i_1}\cdots w_{\bar{m}}^{i_{\bar{m}}} K(w)dw = 0$, for all $i_j \geq 0$, $j=1,\ldots,\bar{m}$, $i_1+\cdots+i_{\bar{m}} < s^*$, and the functions $g_n(t)$ and $E_{\theta_*}(c_n(z,z_i)|t_n(z) = t,z_i)$ are differentiable in $t$ on the interior of $[T_{1,n},\ T_{2,n}]$ to the order $s^*$ and these derivatives are bounded, uniformly in $z_i$ and $n$, then*

$$\sup_{z_i\in Z_n} |E_{\theta_*}(c_n(z,z_i)\frac{1}{a_n^{\bar{m}}}K(\frac{t_n(z_i)-t_n(z)}{a_n})|z_i) - E_{\theta_*}(c_n(z,z_i)|t_n(z_i),z_i)g_n(t_n(z_i))| = O(a_n^{s^*}).$$

**Proposition 3.** *Let $K(v)$ be a differentiable kernel function on $R^{\bar{m}}$ with a bounded support $D$ such that its gradient $\frac{\partial}{\partial v}K(v)$ is bounded. Suppose that the density function $g_n(t)$ of $t_n(z)$, its gradient $\frac{\partial}{\partial t}g_n(t)$, and the conditional expectation $E_{\theta_*}(c_n(z,z_i)|t_n(z)=t,z_i)$ and its derivative*

*$\frac{\partial}{\partial t}E_{\theta_n}(c_n(z,z_i)|t_n(z)=t,z_i)$ are all bounded and uniformly continuous in $t$ on the interior of $[T_{1,n},\ T_{2,n}]$, uniformly in $z_i$ and $n$. Then*

$$\begin{aligned}&\lim_{n\to\infty}\sup_{z_i\in Z_n}|E_{\theta_n}(c_n(z,z_i)\frac{1}{a_n^{m+1}}\frac{\partial}{\partial v}K(\frac{t_n(z_i)-t_n(z)}{a_n})|z_i)\\&\quad-[g_n(t_n(z_i))\frac{\partial}{\partial t}E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)+E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)\frac{\partial}{\partial t}g_n(t_n(z_i))]|\\&=0\end{aligned}$$

*where $Z_n$ is defined in Proposition 2.*

*Furthermore, if the kernel function $K(v)$ has zero moments up to the order $s^*$, the density function $g_n(t)$ and the conditional expectation $E_{\theta_n}(c_n(z,z_i)|t_n(z)=t,z_i)$ are differentiable in $t$ on the interior of $[T_{1,n},\ T_{2,n}]$ up to the order $s^*+1$, and these derivatives are uniformly bounded, then*

$$\begin{aligned}&\sup_{(z_i,\theta)\in Z_n}|E_{\theta_n}(c_n(z,z_i)\frac{1}{a_n^{m+1}}\frac{\partial}{\partial v}K(\frac{t_n(z_i)-t_n(z)}{a_n})|z_i)\\&\quad-[g_n(t_n(z_i))\frac{\partial}{\partial t}E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)+E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)\frac{\partial}{\partial t}g_n(t_n(z_i))]|\\&=O(a_n^{s^*}).\end{aligned}$$

Under Assumption 3 that the kernel functions are bounded, the variances of $A_{n,l}(x_i,\theta)$, $B_n(x_i,\theta)$ and $C_n(z_i,\theta)$ have the familiar orders:

$$\sup_{S_x} var_{\theta_n}(A_{n,l}(x_i,\theta_n)|x_i)=O(\frac{1}{na_n^m})\tag{6.1}$$

$$\sup_{S_x} var_{\theta_n}(B_n(x_i,\theta_n)|x_i)=O(\frac{1}{na_n^m}).\tag{6.2}$$

and

$$\sup_{S_y\times S_x} var_{\theta_n}(C_n(z_i,\theta_n)|x_i,y_i)=O(\frac{1}{na_n^{m+k}}).\tag{6.3}$$

Under Assumption 2, Proposition 1 implies that if $\lim_{n\to\infty}\frac{n}{\ln n}a_n^m=\infty$,

$$\sup_{S_x}|A_{n,l}(x_i,\theta_n)-E_{\theta_n}(A_{n,l}(x_i,\theta_n)|x_i)|\xrightarrow{P_{\theta_n}}0\tag{6.4}$$

$$\sup_{S_x}|B_n(x_i,\theta_n)-E_{\theta_n}(B_n(x_i,\theta_n)|x_i)|\xrightarrow{P_{\theta_n}}0\tag{6.5}$$

and if $\lim_{n\to\infty}\frac{n}{\ln n}a_n^{m+k}=\infty$,

$$\sup_{S_y\times S_x}|C_n(z_i,\theta_n)-E_{\theta_n}(C_n(z_i,\theta_n)|z_i)|\xrightarrow{P_{\theta_n}}0.\tag{6.6}$$

Let

$$Z_{\theta_n}=\{(y,x)|T_1(\theta_n)+\delta_n\le I_1(y-x\gamma(\theta_n),x\alpha(\theta_n))\le T_2(\theta_n)-\delta_n\}\tag{6.7}$$

$$X_{\theta_n} = \{x | T_{1,X}(\theta_n) + \delta_n \le x\alpha(\theta_n) \le T_{2,X}(\theta_n) - \delta_n\} \tag{6.8}$$

Proposition 2 guarantees that

$$\sup_{X_{\theta_n}} |E_{\theta_n}(A_{n,l}(x_i, \theta_n)|x_i) - E_{\theta_n}(I_l | x_i\alpha_n) p(x_i\alpha_n)| \to 0 \tag{6.9}$$

$$\sup_{X_{\theta_n}} |E_{\theta_n}(B_n(x_i, \theta_n)|x_i) - p(x_i\alpha_n)| \to 0 \tag{6.10}$$

and

$$\sup_{Z_{\theta_n}} |E_{\theta_n}(C_n(z_i, \theta_n)|z_i) - E_{\theta_n}(I_{1i} | y_i - x_i\gamma_n, x_i\alpha_n) g(y_i - x_i\gamma_n, x_i\alpha_n)| \to 0. \tag{6.11}$$

Since $p(x_i\alpha)$ and $E_\theta(I_l|x_i\alpha)$ are bounded away from zero on $S_x \times \Theta$ under our Assumption 2, uniform convergences in (6.4), (6.5), (6.9) and (6.10) imply that $\inf_{X_n} B_n(x_i, \theta_n)$ and $\inf_{X_n} A_{n,l}(x_i, \theta_n)$ are bounded away from zero in $P_{\theta_n}$ probability. Similarly, since $f(y_i|I_{1i} = 1, x_i, \theta_n)$ is bounded away from zero on $S_y \times S_x \times \Theta$ under our assumption , $\inf_{Z_n} C_n(z_i, \theta_n)$ is also bounded away from zero in $P_{\theta_n}$ probability. In summary, if $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{m+k} = \infty$,

$$\sup_{X_{\theta_n}} |B_n(x_i, \theta_n) - p(x_i\alpha_n)| \xrightarrow{P_{\theta_n}} 0, \tag{6.12}$$

$$\sup_{X_{\theta_n}} |A_{n,l}(x_i, \theta_n) - E_{\theta_n}(I_l|x_i\alpha_n) p(x_i\alpha_n)| \xrightarrow{P_{\theta_n}} 0 \tag{6.13}$$

$$\sup_{Z_{\theta_n}} |C_n(z_i, \theta_n) - f(y_i|I_{1i} = 1, x_i, \theta_n) E_{\theta_n}(I_{1i}|x_i\alpha_n) p(x_i\alpha_n)| \xrightarrow{P_{\theta_n}} 0 \tag{6.14}$$

and there exists a positive constant $\delta > 0$ such that

$$\lim_{n\to\infty} P_{\theta_n}(\inf_{X_{\theta_n}} B_n(x_i, \theta_n) \ge \delta) = 1 \tag{6.15}$$

$$\lim_{n\to\infty} P_{\theta_n}(\inf_{X_{\theta_n}} A_{n,l}(x_i, \theta_n) \ge \delta) = 1 \tag{6.16}$$

and

$$\lim_{n\to\infty} P_{\theta_n}(\inf_{Z_{\theta_n}} C_n(z_i, \theta_n) \ge \delta) = 1. \tag{6.17}$$

The first order derivatives of $A_{n,l}(x_i, \theta), B_n(x_i, \theta)$ and $C_n(z_i, \theta)$ with respect to $\theta$ are

$$\frac{\partial A_{n,l}(x_i, \theta)}{\partial \theta} = \frac{1}{n-1} \sum_{j \ne i}^{n} I_{lj} \left( \frac{\partial x_i\alpha}{\partial \theta} - \frac{\partial x_j\alpha}{\partial \theta} \right) \cdot \frac{1}{a_n^{m+1}} \nabla K(\frac{x_i\alpha - x_j\alpha}{a_n}), \tag{6.18}$$

$$\frac{\partial B_n(x_i, \theta)}{\partial \theta} = \frac{1}{n-1} \sum_{j \ne i}^{n} \left( \frac{\partial x_i\alpha}{\partial \theta} - \frac{\partial x_j\alpha}{\partial \theta} \right) \cdot \frac{1}{a_n^{m+1}} \nabla K(\frac{x_i\alpha - x_j\alpha}{a_n}) \tag{6.19}$$

and

$$\frac{\partial C_n(z_i, \theta)}{\partial \theta} = \frac{1}{n-1} \sum_{j \ne i}^{n} I_{1j} \left( \frac{\partial x_j\gamma}{\partial \theta} - \frac{\partial x_i\gamma}{\partial \theta}, \frac{\partial x_i\alpha}{\partial \theta} - \frac{\partial x_j\alpha}{\partial \theta} \right) \cdot \frac{1}{a_n^{m+k+1}} \nabla \bar{K}(\frac{(y_i - x_i\gamma) - (y_j - x_j\gamma)}{a_n}, \frac{x_i\alpha - x_j\alpha}{a_n}) \tag{6.20}$$

where $\nabla' K(w) = \frac{\partial}{\partial w} K(w)$ and $\bar{\nabla}' \bar{K}(v) = \frac{\partial}{\partial v} \bar{K}(v)$ denote gradient row vectors of the kernel functions. Under the boundedness conditions in Assumptions 2 and 3, the variances of these derivatives have the familiar orders:

$$\sup_{S_x} Var_{\theta_n} \left( \frac{\partial A_{n,l}(x_i, \theta_n)}{\partial \theta} | x_i \right) = O\left( \frac{1}{n a_n^{m+2}} \right) \tag{6.21}$$

$$\sup_{S_x} Var_{\theta_n} \left( \frac{\partial B_n(x_i, \theta_n)}{\partial \theta} | x_i \right) = O\left( \frac{1}{n a_n^{m+2}} \right) \tag{6.22}$$

and

$$\sup_{S_y \times S_x} Var_{\theta_n} \left( \frac{\partial C_n(z_i, \theta_n)}{\partial \theta} | x_i \right) = O\left( \frac{1}{n a_n^{m+k+2}} \right). \tag{6.23}$$

Proposition 1 and Proposition 3 imply that if $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{m+2} = \infty$,

$$\sup_{X_{\theta_n}} | \frac{\partial B_n(x_i, \theta_n)}{\partial \theta} - \frac{\partial B(x_i, \theta_n)}{\partial \theta} | \xrightarrow{P_{\theta_n}} 0, \tag{6.24}$$

$$\sup_{X_{\theta_n}} | \frac{\partial A_{n,l}(x_i, \theta_n)}{\partial \theta} - \frac{\partial A_l(x_i, \theta_n)}{\partial \theta} | \xrightarrow{P_{\theta_n}} 0 \tag{6.25}$$

where, for any component $\theta_{(k)}$ of $\theta$,

$$\begin{aligned} &\frac{\partial B(x_i, \theta_n)}{\partial \theta_{(k)}} \\ &= -tr \nabla E_{\theta_n}(x | x_i \alpha_n) \frac{\partial \alpha(\theta_n)}{\partial \theta_{(k)}} \cdot p(x_i \alpha_n) + (x_i - E_{\theta_n}(x | x_i \alpha_n)) \frac{\partial \alpha(\theta_n)}{\partial \theta_{(k)}} \nabla p(x_i \alpha_n) \end{aligned} \tag{6.26}$$

and

$$\begin{aligned} &\frac{\partial A_l(x_i, \theta_n)}{\partial \theta_{(k)}} \\ &= tr \nabla E_{\theta_n}(I_l(x_i - x) | x_i \alpha_n) \frac{\partial \alpha(\theta_n)}{\partial \theta_{(k)}} p(x_i \alpha_n) + E_{\theta_n}(I_l(x_i - x) | x_i \alpha_n) \frac{\partial \alpha(\theta_n)}{\partial \theta_{(k)}} \nabla p(x_i \alpha_n) \end{aligned} \tag{6.27}$$

where $\nabla' E_{\theta_n}(\cdot | t) = \frac{\partial}{\partial t} E_{\theta_n}(\cdot | t)$ and $\nabla' p(t) = \frac{\partial}{\partial t} p(t)$. Also, if $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{m+k+2} = \infty$,

$$\sup_{Z_{\theta_n}} \left| \frac{\partial C_n(z_i, \theta_n)}{\partial \theta} - \frac{\partial C(z_i, \theta_n)}{\partial \theta} \right| \xrightarrow{P_{\theta_n}} 0 \tag{6.28}$$

where

$$\begin{aligned} &\frac{\partial C(z_i, \theta_n)}{\partial \theta_{(k)}} \\ &= tr \bar{\nabla} E_{\theta_n}(I_1[(x - x_i) \frac{\partial \gamma(\theta_n)}{\partial \theta_{(k)}}, (x_i - x) \frac{\partial \alpha(\theta_n)}{\partial \theta_{(k)}}] | y_i - x_i \gamma_n, x_i \alpha_n) g(y_i - x_i \gamma_n, x_i \alpha_n) \\ &+ E_{\theta_n}(I_1[(x - x_i) \frac{\partial \gamma(\theta_n)}{\partial \theta_{(k)}}, (x_i - x) \frac{\partial \alpha(\theta_n)}{\partial \theta_{(k)}}] | y_i - x_i \gamma_n, x_i \alpha_n) \bar{\nabla} g(y_i - x_i \gamma_n, x_i \alpha_n) \end{aligned} \tag{6.29}$$

where $\bar{\nabla}' g(w) = \frac{\partial}{\partial w} g(w)$.

Denote

$$B(x_i, \theta) = p(x_i\alpha(\theta)) \tag{6.30}$$

$$A_l(x_i, \theta) = E_\theta(I_l | x_i\alpha) p(x_i\alpha) \tag{6.31}$$

$$\begin{aligned} C(z_i, \theta) &= E_\theta(I_1 | y_i - x_i\gamma, x_i\alpha) g(y_i - x_i\gamma, x_i\alpha) \\ &= f(y_i | I_{1i} = 1, x_i, \theta) E_\theta(I_1 | x_i\alpha) p(x_i\alpha). \end{aligned} \tag{6.32}$$

Combining terms, as shown in Appendix 2, we have

$$\begin{aligned} &\frac{1}{C(z_i, \theta)} \frac{\partial C(z_i, \theta)}{\partial \theta} - \frac{1}{A_1(x_i, \theta)} \frac{\partial A_1(x_i, \theta)}{\partial \theta} \\ &= \left\{ E_\theta \left( \left[ \frac{\partial x\gamma}{\partial \theta}, -\frac{\partial x\alpha}{\partial \theta} \right] | x_i\alpha \right) - [\frac{\partial x_i\gamma}{\partial \theta}, -\frac{\partial x_i\alpha}{\partial \theta}] \right\} \bar{\nabla} \ln f(y_i - x_i\gamma | I_{1i} = 1, x_i\alpha) \end{aligned} \tag{6.33}$$

where $\bar{\nabla}' \ln f(\epsilon | I_1 = 1, t) = \frac{\partial}{\partial(\epsilon, t)} \ln f(\epsilon | I_1 = 1, t)$. Also

$$\frac{1}{A_l(x_i, \theta)} \frac{\partial A_l(x_i, \theta)}{\partial \theta} - \frac{1}{B(x_i, \theta)} \frac{\partial B(x_i, \theta)}{\partial \theta} = \left[ \frac{\partial x_i\alpha}{\partial \theta} - E_\theta \left( \frac{\partial x\alpha}{\partial \theta} | x_i\alpha \right) \right] \nabla \ln P(I_l = 1 | x_i\alpha) \tag{6.34}$$

where $\nabla' \ln P(I_l = 1 | t) = \frac{\partial}{\partial t} \ln P(I_l = 1 | t)$.

## 7. Effective Scores and Asymptotic Equivalence

In this section, we will investigate the limiting property of $\sqrt{n}(S_n(\bar{\theta}_n) - S_n^*(\bar{\theta}_n))$. The limiting property of $I_n^{-1}(\bar{\theta}_n)$ and the limiting distribution $\sqrt{n}(\theta_n^* - \theta_o)$ will be investigated in subsequent sections.

By contiguity and discretization, to show that $\sqrt{n}(S_n(\bar{\theta}_n) - S_n^*(\bar{\theta}_n))$ converges to zero in probability under $P_{\theta_o}$, it is sufficient to prove that, for any deterministic sequence $\theta_n = \theta_o + \frac{1}{\sqrt{n}} h_n$ with $h_n$ being a bounded sequence, $\sqrt{n}(S_n(\theta_n) - S_n^*(\theta_n))$ converges to zero in $P_{\theta_n}$ probability. By a Taylor expansion,

$$\begin{aligned}\sqrt{n}(S_n(\theta_n) - S_n^*(\theta_n)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_{\hat{Z}_{\theta_n}}(z_i) - 1)D_{1i}(\theta_n) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_{\hat{X}_{\theta_n}}(x_i) - 1)D_{2i}(\theta_n) \\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_{n,i}(\theta_n) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{n,i}(\theta_n)\end{aligned} \tag{7.1}$$

where

$$D_{1i}(\theta_n) = I_{1i}\left[\frac{1}{C(z_i,\theta_n)}\frac{\partial C(z_i,\theta_n)}{\partial\theta} - \frac{1}{A_1(x_i,\theta_n)}\frac{\partial A_1(x_i,\theta_n)}{\partial\theta}\right], \tag{7.2}$$

$$D_{2i}(\theta_n) = \sum_{l=1}^{L} I_{li}\left[\frac{1}{A_l(x_i,\theta_n)}\frac{\partial A_l(x_i,\theta_n)}{\partial\theta} - \frac{1}{B(x_i,\theta_n)}\frac{\partial B(x_i,\theta_n)}{\partial\theta}\right], \tag{7.3}$$

$$\begin{aligned}&L_{n,i}(\theta_n) \\ &= I_{\hat{Z}_{\theta_n}}(z_i) I_{1i}\Big[\frac{1}{C(z_i,\theta_n)}\frac{\partial C_n(z_i,\theta_n)}{\partial\theta} - \Big(\frac{1}{C(z_i,\theta_n)}\Big)^2\frac{\partial C(z_i,\theta_n)}{\partial\theta}C_n(z_i,\theta_n) \\ &- \frac{1}{A_1(x_i,\theta_n)}\frac{\partial A_{n,1}(x_i,\theta_n)}{\partial\theta} + \Big(\frac{1}{A_1(x_i,\theta_n)}\Big)^2\frac{\partial A_1(x_i,\theta_n)}{\partial\theta}A_{n,1}(x_i,\theta_n)\Big] \\ &+ I_{\hat{X}_{\theta_n}}(x_i)\sum_{l=1}^{L} I_{li}\Big[\frac{1}{A_l(x_i,\theta_n)}\frac{\partial A_{n,l}(x_i,\theta_n)}{\partial\theta} - \Big(\frac{1}{A_l(x_i,\theta_n)}\Big)^2\frac{\partial A_l(x_i,\theta_n)}{\partial\theta}A_{n,l}(x_i,\theta_n) \\ &- \frac{1}{B(x_i,\theta_n)}\frac{\partial B_n(x_i,\theta_n)}{\partial\theta} + \Big(\frac{1}{B(x_i,\theta_n)}\Big)^2\frac{\partial B(x_i,\theta_n)}{\partial\theta}B_n(x_i,\theta_n)\Big]\end{aligned} \tag{7.4}$$

and

$$
\begin{aligned}
&R_{n,i}(\theta_n)\\
&= I_{\tilde{Z}_{\theta_n}}(z_i)I_{1i}[-(\frac{1}{C_n(\tilde{z}_i,\theta_n)})^2(\frac{\partial C_n(z_i,\theta_n)}{\partial\theta}-\frac{\partial C(z_i,\theta_n)}{\partial\theta})(C_n(z_i,\theta_n)-C(z_i,\theta_n))\\
&+(\frac{1}{C_n(\tilde{z}_i,\theta_n)})^3\frac{\partial C_n(\tilde{z}_i,\theta_n)}{\partial\theta}(C_n(z_i,\theta_n)-C(z_i,\theta_n))^2\\
&+(\frac{1}{A_{n,1}(\tilde{x}_i,\theta_n)})^2(\frac{\partial A_{n,1}(x_i,\theta_n)}{\partial\theta}-\frac{\partial A_1(x_i,\theta_n)}{\partial\theta})(A_{n,1}(x_i,\theta_n)-A_1(x_i,\theta_n))\\
&-(\frac{1}{A_{n,1}(\tilde{x}_i,\theta_n)})^3\frac{\partial A_{n,1}(\tilde{x}_i,\theta_n)}{\partial\theta}(A_{n,1}(x_i,\theta_n)-A_1(x_i,\theta_n))^2]\\
&+I_{\tilde{X}_{\theta_n}}(x_i)\sum_{l=1}^{L}I_{li}[-(\frac{1}{A_{n,l}(\tilde{x}_i,\theta_n)})^2(\frac{\partial A_{n,l}(x_i,\theta_n)}{\partial\theta}-\frac{\partial A_l(x_i,\theta_n)}{\partial\theta})(A_{n,l}(x_i,\theta_n)-A_l(x_i,\theta_n))\\
&+(\frac{1}{A_{n,l}(\tilde{x}_i,\theta_n)})^3\frac{\partial A_{n,l}(\tilde{x}_i,\theta_n)}{\partial\theta}(A_{n,l}(x_i,\theta_n)-A_l(x_i,\theta_n))^2\\
&+(\frac{1}{B_n(\tilde{x}_i,\theta_n)})^2(\frac{\partial B_n(x_i,\theta_n)}{\partial\theta}-\frac{\partial B(x_i,\theta_n)}{\partial\theta})(B_n(x_i,\theta_n)-B(x_i,\theta_n))\\
&-(\frac{1}{B_n(\tilde{x}_i,\theta_n)})^3\frac{\partial B_n(\tilde{x}_i,\theta_n)}{\partial\theta}(B_n(x_i,\theta_n)-B(x_i,\theta_n))^2]
\end{aligned}
\tag{7.5}
$$

where $C_n(\tilde{z_i},\theta_n)$ lies between $C_n(z_i,\theta_n)$ and $C(z_i,\theta_n)$; $\frac{\partial C_n(\tilde{z}_i,\theta_n)}{\partial\theta}$ lies between $\frac{\partial C_n(z_i,\theta_n)}{\partial\theta}$ and $\frac{\partial C(z_i,\theta_n)}{\partial\theta}$, etc.

Consider the first two terms in (7.1). With (6.33) and (6.34),

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_{\tilde{Z}_{\theta_n}}(z_i)-1)D_{1i}(\theta_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_{\tilde{Z}_{\theta_n}}(z_i)-1)I_{1i}\Big(E_{\theta_n}([\frac{\partial x\gamma_n}{\partial\theta},\ -\frac{\partial x\alpha_n}{\partial\theta}]|x_i\alpha_n)\\
-[\frac{\partial x_i\gamma_n}{\partial\theta},-\frac{\partial x_i\alpha_n}{\partial\theta}]\Big)\bar{\nabla}\ln f(y_i-x_i\gamma_n|I_{1i}=1,x_i\alpha_n)
\end{aligned}
\tag{7.6}
$$

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_{\tilde{X}_{\theta_n}}(x_i)-1)D_{2i}(\theta_n)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_{\tilde{X}_{\theta_n}}(x_i)-1)[\frac{\partial x_i\alpha_n}{\partial\theta}-E_{\theta_n}(\frac{\partial x\alpha_n}{\partial\theta}|x_i\alpha_n)]\cdot\sum_{l=1}^{L}I_{li}\nabla\ln P(I_l=1|x_i\alpha_n).
\end{aligned}
\tag{7.7}
$$

Both terms have zero mean under $P_{\theta_n}$. This is so, since $E_{\theta_n}(I_{\tilde{Z}_{\theta_n}}(z_i)|I_{1i},y_i,x_i)$ is a function of $x_i\alpha_n$ and $y_i-x_i\gamma_n$,

$$
\begin{aligned}
&E_{\theta_n}\big((I_{\tilde{Z}_{\theta_n}}(z_i)-1)D_{1i}(\theta_n)\big)\\
&= E_{\theta_n}\{\big(E_{\theta_n}(I_{\tilde{Z}_{\theta_n}}(z_i)|I_{i1}, y_i - x_i\gamma_n, x_i\alpha_n)-1\big)I_{1i}\Big[E_{\theta_n}([\frac{\partial x\gamma_n}{\partial\theta}, -\frac{\partial x\alpha_n}{\partial\theta}]|x_i\alpha_n)\\
&\quad - E_{\theta_n}([\frac{\partial x\gamma_n}{\partial\theta}, -\frac{\partial x\alpha_n}{\partial\theta}]|I_{1i}, y_i - x_i\gamma_n, x_i\alpha_n)\Big]\bar{\nabla}\ln f(y_i - x_i\gamma_n|I_{1i}=1, x_i\alpha_n)\}\\
&= 0
\end{aligned}
$$

owning to that

$$E_{\theta_n}\big([\frac{\partial x\gamma_n}{\partial\theta}, -\frac{\partial x\alpha_n}{\partial\theta}]|I_{1i}, x_i\alpha_n, y_i - x_i\gamma_n\big) = E_{\theta_n}\big([\frac{\partial x\gamma_n}{\partial\theta}, -\frac{\partial x\alpha_n}{\partial\theta}]|x_i\alpha_n\big).$$

Similarly, $E_{\theta_n}(I_{\tilde{X}_{\theta_n}}(x_i)|x_i)$ is a function of $x_i\alpha_n$,

$$
\begin{aligned}
&E_{\theta_n}\big((I_{\tilde{X}_{\theta_n}}(x_i)-1)D_{2i}(\theta_n)\big)\\
&= E_{\theta_n}\{(E_{\theta_n}(I_{\tilde{X}_{\theta_n}}(x_i)|x_i\alpha_n)-1)[\frac{\partial x_i\alpha_n}{\partial\theta} - E_{\theta_n}(\frac{\partial x\alpha_n}{\partial\theta}|x_i\alpha_n)]\cdot\sum_{l=1}^{L}E_{\theta_n}(I_l|x_i)\nabla\ln P(I_l=1|x_i\alpha_n)\}\\
&= 0
\end{aligned}
$$

owning to that $\nabla\sum_{l=1}^{L}P(I_l=1|x_i\alpha_n)=0$. On the other hand,

$$
\begin{aligned}
&Var_{\theta_n}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_{\tilde{Z}_{\theta_n}}(z_i)-1)D_{1i}(\theta_n))\\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E_{\theta_n}\{(1-I_{\tilde{Z}_{\theta_n}}(z_i))(1-I_{\tilde{Z}_{\theta_n}}(z_j))D_{1i}(\theta_n)D'_{1j}(\theta_n)\}\\
&= E_{\theta_n}\big((1-I_{\tilde{Z}_{\theta_n}}(z_i))D_{1i}(\theta_n)D'_{1i}(\theta_n)\big).
\end{aligned}
$$

This is so, since whenerver $i\neq j$,

$$
\begin{aligned}
&E_{\theta_n}\{(1-I_{\tilde{Z}_{\theta_n}}(z_i))(1-I_{\tilde{Z}_{\theta_n}}(z_j))D_{1i}(\theta_n)D'_{1j}(\theta_n)\}\\
&= E_{\theta_n}\{E_{\theta_n}\big((1-I_{\tilde{Z}_{\theta_n}}(z_i))(1-I_{\tilde{Z}_{\theta_n}}(z_j))D_{1i}(\theta_n)|I_{1j}, z_j\big)D'_{1j}(\theta_n)\}\\
&= E_{\theta_n}\{E_{\theta_n}\big((1-I_{\tilde{Z}_{\theta_n}}(z_i))(1-I_{\tilde{Z}_{\theta_n}}(z_j))D_{1i}(\theta_n)|I_{1j}, y_j - x_j\gamma_n, x_j\alpha_n\big)D'_{1j}(\theta_n)\}\\
&= E_{\theta_n}\{E_{\theta_n}\big((1-I_{\tilde{Z}_{\theta_n}}(z_i))(1-I_{\tilde{Z}_{\theta_n}}(z_j))D_{1i}(\theta_n)|I_{1j}, y_j - x_j\gamma_n, x_j\alpha_n\big)\\
&\quad\cdot E_{\theta_n}[D'_{1j}(\theta_n)|I_{1j}, y_j - x_j\gamma_n, x_j\alpha_n]\}\\
&= 0.
\end{aligned}
$$

Similarly,

$$Var_{\theta_n}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_{\tilde{X}_{\theta_n}}(x_i)-1)D_{2i}(\theta_n)) = E_{\theta_n}\{(1-I_{\tilde{X}_{\theta_n}}(x_i))D_{2i}(\theta_n)D'_{2i}(\theta_n)\}$$

owning to that

$$E_{\theta_n}(\frac{\partial x\alpha_n}{\partial\theta}|I_{1i},\cdots,I_{Li}, x_i\alpha_n) = E_{\theta_n}(\frac{\partial x\alpha_n}{\partial\theta}|x_i\alpha_n).$$

Since $I_{Z_{\theta_n}}(z_i) \xrightarrow{P_{\theta_n}} 1$ and $I_{X_{\theta_n}}(x_i) \xrightarrow{P_{\theta_n}} 1$ in $P_{\theta_n}$, these variances converge to zero by the Lebesque dominated convergence theorem.

The last term in (7.1) can be shown to converge to zero in probability under $P_{\theta_n}$ with the following proposition.

**Proposition 4.** *Let $C_{j,n}(z_1, \cdots, z_{i-1}, z_{i+1}, \cdots, z_n;\ z_i)$, $j = 1,2$ be two sequences of measurable functions of the random variables $z_1, \cdots, z_n$ which are i.i.d. under $P_{\theta_n}$. Suppose that, for each $j$,*

*(1)* $\sup_{Z_n} |E(C_{j,n}(z_1, \cdots, z_n;\ z_i)|z_i) - C_j(z_i)| = O(a_n^{s_j})$ *, for some measurable functions $C_j(z_i)$, and*

*(2)* $\sup_{Z_n} var(C_{j,n}(z_1, \cdots, z_n;\ z_i)|z_i) = O(\frac{1}{n a_n^{r_j}})$ *, $j = 1, 2$.*

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{Z_n}(z_i)\,|C_{1,n}(z_1, \cdots, z_n; z_i) - C_1(z_i)| \cdot |C_{2n}(z_1, \cdots, z_n; z_i) - C_2(z_i)| \xrightarrow{P_{\theta_n}} 0.$$

**Proof:** see Appendix 1.

Since $B_n(x_i, \theta_n), A_{n,l}(x_i, \theta_n)$ and $C_n(z_i, \theta_n)$ are bounded away from zero in $P_{\theta_n}$ probability on $Z_{\theta_n}$ in (6.12)-(6.14), $\tilde{B}_n(x_i, \theta_n), \tilde{A}_{n,l}(x_i, \theta_n)$ and $\tilde{C}_n(z_i, \theta_n)$ will also be bounded away from zero in $P_{\theta_n}$-probability on $Z_{\theta_n}$. Also, since $\frac{\partial B_n(x_i,\theta_n)}{\partial\theta} - \frac{\partial B(x_i,\theta_n)}{\partial\theta}$, $\frac{\partial A_{n,l}(x_i,\theta_n)}{\partial\theta} - \frac{\partial A_l(x_i,\theta_n)}{\partial\theta}$ and $\frac{\partial C_n(z_i,\theta_n)}{\partial\theta} - \frac{\partial C(z_i,\theta_n)}{\partial\theta}$ converge to zero in $P_{\theta_n}$ probability uniformly on $Z_{\theta_n}$ and $\frac{\partial B(x_i,\theta_n)}{\partial\theta}$, $\frac{\partial A_l(x_i,\theta_n)}{\partial\theta}$ and $\frac{\partial C(z_i,\theta_n)}{\partial\theta}$ are bounded uniformly in $n$, $\frac{\partial B_n(\tilde{x}_i,\theta_n)}{\partial\theta}$, $\frac{\partial A_{n,l}(\tilde{x}_i,\theta_n)}{\partial\theta}$ and $\frac{\partial C_n(\tilde{z}_i,\theta_n)}{\partial\theta}$ will be bounded in $P_{\theta_n}$ probability. Since $\tilde{Z}_{\theta_n} \subseteq Z_{\theta_n}$ and $\tilde{X}_{\theta_n} \subseteq X_{\theta_n}$, it follows that

$$
\begin{aligned}
&\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{n,i}(\theta_n)\| \\
&\leq \sup_{Z_{\theta_n}} \frac{1}{|C_n(\tilde{z_i},\theta_n)|^2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Z_{\theta_n}}(z_i)\|\frac{\partial C_n(z_i,\theta_n)}{\partial\theta}-\frac{\partial C(z_i,\theta_n)}{\partial\theta}\|\cdot|C_n(z_i,\theta_n)-C(z_i,\theta_n)| \\
&\quad+\sup_{Z_{\theta_n}} \frac{1}{|C_n(\tilde{z_i},\theta_n)|^3}\|\frac{\partial C_n(\tilde{z_i},\theta_n)}{\partial\theta}\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Z_{\theta_n}}(z_i)|C_n(z_i,\theta_n)-C(z_i,\theta_n)|^2 \\
&\quad+\sup_{Z_{\theta_n}} \frac{1}{|A_{n,1}(\tilde{x_i},\theta_n)|^2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Z_{\theta_n}}(z_i)\|\frac{\partial A_{n,1}(x_i,\theta_n)}{\partial\theta}-\frac{\partial A_1(x_i,\theta_n)}{\partial\theta}\|\cdot|A_{n,1}(x_i,\theta_n)-A_1(x_i,\theta_n)| \\
&\quad+\sup_{Z_{\theta_n}} \frac{1}{|A_{n,1}(\tilde{x_i},\theta_n)|^3}\|\frac{\partial A_{n,1}(\tilde{x_i},\theta_n)}{\partial\theta}\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Z_{\theta_n}}(z_i)|A_{n,1}(x_i,\theta_n)-A_1(x_i,\theta_n)|^2 \\
&\quad+\sum_{l=1}^{L}\{\sup_{X_{\theta_n}} \frac{1}{|A_{n,l}(\tilde{x_i},\theta_n)|^2} \\
&\qquad\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{X_{\theta_n}}(x_i)\|\frac{\partial A_{n,l}(x_i,\theta_n)}{\partial\theta}-\frac{\partial A_l(x_i,\theta_n)}{\partial\theta}\|\cdot|A_{n,l}(x_i,\theta_n)-A_l(x_i,\theta_n)| \\
&\quad+\sup_{X_{\theta_n}} \frac{1}{|A_{n,l}(\tilde{x_i},\theta_n)|^3}\|\frac{\partial A_{n,l}(\tilde{x_i},\theta_n)}{\partial\theta}\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{X_{\theta_n}}(x_i)|A_{n,l}(x_i,\theta_n)-A_l(x_i,\theta_n)|^2 \\
&\quad+\sup_{X_{\theta_n}} \frac{1}{|B_n(\tilde{x_i},\theta_n)|^2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{X_{\theta_n}}(x_i)\|\frac{\partial B_n(x_i,\theta_n)}{\partial\theta}-\frac{\partial B(x_i,\theta_n)}{\partial\theta}\|\cdot|B_n(x_i,\theta_n)-B(x_i,\theta_n)| \\
&\quad+\sup_{X_{\theta_n}} \frac{1}{|B_n(\tilde{x_i},\theta_n)|^3}\|\frac{\partial B(\tilde{x_i},\theta_n)}{\partial\theta}\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{X_{\theta_n}}(x_i)|B_n(x_i,\theta_n)-B(x_i,\theta_n)|^2\}
\end{aligned}
$$

which is less than

$$
\begin{aligned}
&O_{P_{\theta_n}}(1)\cdot\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Z_{\theta_n}}(z_i)\|\frac{\partial C_n(z_i,\theta_n)}{\partial\theta}-\frac{\partial C(z_i,\theta_n)}{\partial\theta}\|\cdot|C_n(z_i,\theta_n)-C(z_i,\theta_n)| \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Z_{\theta_n}}(z_i)\|\frac{\partial A_{n,1}(x_i,\theta_n)}{\partial\theta}-\frac{\partial A_1(x_i,\theta_n)}{\partial\theta}\|\cdot|A_{n,1}(x_i,\theta_n)-A_1(x_i,\theta_n)| \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Z_{\theta_n}}(z_i)|C_n(z_i,\theta_n)-C(z_i,\theta_n)|^2+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Z_{\theta_n}}(z_i)|A_{n,1}(x_i,\theta_n)-A_1(x_i,\theta_n)|^2 \\
&+\sum_{l=1}^{L}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{X_{\theta_n}}(x_i)\|\frac{\partial A_{n,l}(x_i,\theta_n)}{\partial\theta}-\frac{\partial A_l(x_i,\theta_n)}{\partial\theta}\|\cdot|A_{n,l}(x_i,\theta_n)-A_l(x_i,\theta_n)| \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{X_{\theta_n}}(x_i)\|\frac{\partial B_n(x_i,\theta_n)}{\partial\theta}-\frac{\partial B(x_i,\theta_n)}{\partial\theta}\|\cdot|B_n(x_i,\theta_n)-B(x_i,\theta_n)| \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{X_{\theta_n}}(x_i)|A_{n,l}(x_i,\theta_n)-A_l(x_i,\theta_n)|^2+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{X_{\theta_n}}(x_i)|B_n(x_i,\theta_n)-B(x_i,\theta_n)|^2]\}.
\end{aligned}
\tag{7.8}
$$

As the kernel functions $K(\cdot)$ and $\bar{K}(\cdot)$ have zero moments up to the order $s^*$, Proposition 2 and Proposition 3 imply that, under Assumptions 4 and 5,

$$\sup_{X_{\theta_n}} |E_{\theta_n}(B_n(x_i,\theta_n)|x_i) - B(x_i,\theta_n)| = O(a_n^{s^*}), \tag{7.9}$$

$$\sup_{X_{\theta_n}} |E_{\theta_n}(A_{n,l}(x_i,\theta_n) - A_l(x_i,\theta_n)| = O(a_n^{s^*}), \tag{7.10}$$

$$\sup_{Z_{\theta_n}} |E_{\theta_n}(C_n(z_i,\theta_n)|z_i) - C(z_i,\theta_n)| = O(a_n^{s^*}), \tag{7.11}$$

$$\sup_{X_{\theta_n}} \|E_{\theta_n}(\frac{\partial B_n(x_i,\theta_n)}{\partial\theta}|x_i) - \frac{\partial B(x_i,\theta_n)}{\partial\theta}\| = O(a_n^{s^*}), \tag{7.12}$$

$$\sup_{X_{\theta_n}} \|E_{\theta_n}(\frac{\partial A_{n,l}(x_i,\theta_n)}{\partial\theta} - \frac{\partial A_l(x_i,\theta_n)}{\partial\theta}\| = O(a_n^{s^*}), \tag{7.13}$$

and

$$\sup_{Z_{\theta_n}} \|E_{\theta_n}(\frac{\partial C_n(z_i,\theta_n)}{\partial\theta}|x_i,y_i) - \frac{\partial C(z_i,\theta_n)}{\partial\theta}\| = O(a_n^{s^*}). \tag{7.14}$$

As it has been pointed out before, the variances of $A_{n,l}(x_i,\theta_n)$ and $B_n(x_i,\theta_n)$ under $P_{\theta_n}$ have order $O\left(\frac{1}{na_n^m}\right)$, the variances of $\frac{\partial A_{n,l}(x_i,\theta_n)}{\partial\theta}$ and $\frac{\partial B_n(x_i,\theta_n)}{\partial\theta}$ have order $O\left(\frac{1}{na_n^{m+2}}\right)$; the variance of $C_n(z_i,\theta_n)$ has order $O\left(\frac{1}{na_n^{m+k}}\right)$ and the variance of $\frac{\partial C_n(z_i,\theta_n)}{\partial\theta}$ has order $O\left(\frac{1}{na_n^{m+k+2}}\right)$. Since $s^* \geq m+k+2$, $\lim_{n\to\infty} na_n^{2(m+k+1)} = \infty$ and $\lim_{n\to\infty} na_n^{4s^*} = 0$ by our assumptions, Proposition 4 implies that all the terms in (7.8) converge to zero in $P_{\theta_n}$ probability.

The third term in (7.1) will also converge to zero in $P_{\theta_n}$ probability. Analysis of the convergence property of the third term in (7.1) can proceed in two steps. Define $L^*_{n,i}(\theta_n)$ as the term $L_{n,i}(\theta_n)$ in (7.4) with the indicator functions $I_{\hat{Z}_{\theta_n}}$ and $I_{\hat{X}_{\theta_n}}$ replaced respectively by the indicator functions $I_{Z_{\theta_n}}$ and $I_{X_{\theta_n}}$. Let

$$\begin{aligned}
&\Phi_1(z_i,I_i,z_j,I_j,a_n,\theta_n) \\
&= I_{Z_{\theta_n}}(z_i)I_{1i}I_{1j}[\frac{1}{C(z_i,\theta_n)}(\frac{\partial x_j\gamma_n}{\partial\theta} - \frac{\partial x_i\gamma_n}{\partial\theta}, \frac{\partial x_i\alpha_n}{\partial\theta} - \frac{\partial x_j\alpha_n}{\partial\theta}) \\
&\quad \cdot \frac{1}{a_n^{m+k+1}}\bar{\nabla}\bar{K}(\frac{(y_i - x_i\gamma_n)-(y_j-x_j\gamma_n)}{a_n}, \frac{x_i\alpha_n - x_j\alpha_n}{a_n}) \\
&\quad - \frac{\partial C(z_i,\theta_n)}{\partial\theta}(\frac{1}{C(z_i,\theta_n)})^2\frac{1}{a_n^{m+k}}\bar{K}(\frac{(y_i - x_i\gamma_n)-(y_j-x_j\gamma_n)}{a_n}, \frac{x_i\alpha_n - x_j\alpha_n}{a_n}) \\
&\quad - \frac{1}{A_1(x_i,\theta_n)}(\frac{\partial x_i\alpha_n}{\partial\theta} - \frac{\partial x_j\alpha_n}{\partial\theta})\frac{1}{a_n^{m+1}}\nabla K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n}) \\
&\quad + \frac{\partial A_1(x_i,\theta_n)}{\partial\theta}(\frac{1}{A_1(x_i,\theta_n)})^2\frac{1}{a_n^m}K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n})]
\end{aligned} \tag{7.15}$$

and

$$
\begin{aligned}
&\Phi_2(x_i, I_i, x_j, I_j, a_n, \theta_n) \\
&= I_{X_{\theta_n}}(x_i) \sum_{l=1}^{L} I_{li}[I_{lj} \frac{1}{A_l(x_i,\theta_n)} (\frac{\partial x_i \alpha_n}{\partial \theta} - \frac{\partial x_j \alpha_n}{\partial \theta}) \frac{1}{a_n^{m+1}} \nabla K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n}) \\
&- I_{lj} \frac{\partial A_l(x_i,\theta_n)}{\partial \theta} (\frac{1}{A_l(x_i,\theta_n)})^2 \frac{1}{a_n^m} K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n}) \\
&- \frac{1}{B(x_i,\theta_n)} (\frac{\partial x_i \alpha_n}{\partial \theta} - \frac{\partial x_j \alpha_n}{\partial \theta}) \frac{1}{a_n^{m+1}} \nabla K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n}) \\
&+ \frac{\partial B(x_i,\theta_n)}{\partial \theta} (\frac{1}{B(x_i,\theta_n)})^2 \frac{1}{a_n^m} K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n})]
\end{aligned} \tag{7.16}
$$

It follows that

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} L_{n,i}^*(\theta_n) \\
&= \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \Phi_1(z_i, I_i, z_j, I_j, a_n, \theta_n) + \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \Phi_2(z_i, I_i, z_j, I_j, a_n, \theta_n).
\end{aligned} \tag{7.17}
$$

The first step is to show that these terms converge in probability to zero. The second step is to show that the difference of (7.17) and the third term on the right hand side of (7.1) will converge in probability to zero also. The following result on U statistic will be useful for the first purpose.

**Proposition 5.** *Let $z_1, \cdots, z_n$ be independent observations drawn from a common distribution $P_{\theta_n}$ with density (probability function for discrete variable) $f_{\theta_n}(z)$, and $\Phi_n(z_1, z_2, a_n, \theta_n)$ be a sequence of measurable functions. Suppose that*
*(1) $f_{\theta_n}(z)$ is bounded by a constant M, uniformly in n;*
*(2) there exist Lebesque square integrable functions $h_j(z), j = 1,2$ such that*

$$
|\, E_{\theta_n}(\Phi_n(z_1, z_2, a_n, \theta_n)|z_1) \,| \le h_1(z_1), \quad |\, E_{\theta_n}(\Phi_n(z_1, z_2, a_n, \theta_n)|z_2) \,| \le h_2(z_2)
$$

*for all n;*
*(3) $E_{\theta_n}(\Phi_n(z_1, z_2, a_n, \theta_n)) = O(a_n^s)$ and $Var_{\theta_n}(\Phi_n(z_1, z_2, a_n, \theta_n)) = O(\frac{1}{a_n^r})$ , and*
*(4) $\lim_{n\to\infty} E_{\theta_n}(\Phi_n(z_1, z_2, a_n, \theta_n)|z_j) = 0$, a.e., $j = 1, 2$.*
*If $\lim_{n\to\infty} \sqrt{n} a_n^s = 0$ and $\lim_{n\to\infty} n a_n^r = \infty$, then*

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

**Proof:** See Appendix 1.

Since the kernel functions $K(\cdot)$ and $\bar{K}(\cdot)$ are of order $s^*$, Proposition 2 and Proposition 3 imply that

$$
E_{\theta_n}(\Phi_1(z_i, I_i, z_j, I_j, a_n, \theta_n)|z_i, I_i) = O(a_n^{s^*})
$$

and

$$E_{\theta_n}(\Phi_2(x_i, I_i, x_j, I_j, a_n, \theta_n)|x_i, I_i) = O(a_n^{s^*})$$

uniformly in $z_i$ and $x_i$ in $Z_{\theta_n}$ and $X_{\theta_n}$, which in turn imply that

$$E_{\theta_n}(\Phi_1(z_i, I_i, z_j, I_j, a_n, \theta_n)) = O(a_n^{s^*})$$

and

$$E_{\theta_n}(\Phi_2(x_i, I_i, x_j, I_j, a_n, \theta_n)) = O(a_n^{s^*}).$$

The equations (6.1)-(6.3) and (6.21)-(6.23) imply that

$$Var_{\theta_n}(\Phi_1(z_i, I_i, z_j, I_j, a_n, \theta_n)) = O(\frac{1}{a_n^{m+k+2}})$$

and

$$Var_{\theta_n}(\Phi_2(x_i, I_i, x_j, I_j, a_n, \theta_n)) = O(\frac{1}{a_n^{m+2}}).$$

Since

$$\begin{aligned}
&E_{\theta_n}(\Phi_1(z_i, I_i, z_j, I_j, a_n, \theta_n)|z_i, I_i) \\
&= I_{Z_{\theta_n}}(z_i) I_{1i}\{\frac{1}{C(z_i,\theta_n)} E_{\theta_n}[I_{1j}(\frac{\partial x_j\gamma_n}{\partial\theta} - \frac{\partial x_i\gamma_n}{\partial\theta}, \frac{\partial x_i\alpha_n}{\partial\theta} - \frac{\partial x_j\alpha_n}{\partial\theta}) \\
&\cdot \frac{1}{a_n^{m+k+1}} \nabla\bar{K}(\frac{(y_i - x_i\gamma_n) - (y_j - x_j\gamma_n)}{a_n}, \frac{x_i\alpha_n - x_j\alpha_n}{a_n})|z_i] \\
&- \frac{\partial C(z_i,\theta_n)}{\partial\theta}(\frac{1}{C(z_i,\theta_n)})^2 E_{\theta_n}[I_{1j}\frac{1}{a_n^{m+k}}\bar{K}(\frac{(y_i - x_i\gamma_n) - (y_j - x_j\gamma_n)}{a_n}, \frac{x_i\alpha_n - x_j\alpha_n}{a_n})|z_i] \\
&- \frac{1}{A_1(x_i,\theta_n)} E_{\theta_n}[I_{1j}(\frac{\partial x_i\alpha_n}{\partial\theta} - \frac{\partial x_j\alpha_n}{\partial\theta})\frac{1}{a_n^{m+1}}\nabla K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n})|x_i] \\
&+ \frac{\partial A_1(x_i,\theta_n)}{\partial\theta}(\frac{1}{A_1(x_i,\theta_n)})^2 E_{\theta_n}[I_{1j}\frac{1}{a_n^m}K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n})|x_i]\},
\end{aligned}$$

it follows that

$$\begin{aligned}
&\lim_{n\to\infty} E_{\theta_n}(\Phi_1(z_i, I_i, z_j, I_j, a_n, \theta_n)|z_i, I_i) \\
&= I_{1i}[\frac{1}{C(z_i,\theta_o)}\frac{\partial C(z_i,\theta_o)}{\partial\theta} - (\frac{1}{C(z_i,\theta_o)})^2\frac{\partial C(z_i,\theta_o)}{\partial\theta}C(z_i,\theta_o) \\
&- \frac{1}{A_1(x_i,\theta_o)}\frac{\partial A_1(x_i,\theta_o)}{\partial\theta} + (\frac{1}{A_1(x_i,\theta_o)})^2\frac{\partial A_1(x_i,\theta_o)}{\partial\theta}A_1(x_i,\theta_o)] \\
&= 0.
\end{aligned}$$

Similarly,

$$
\begin{aligned}
&E_{\theta_n}(\Phi_2(x_i, I_i, x_j, I_j, a_n, \theta_n)|x_i, I_i) \\
&= I_{X_{\theta_n}}(x_i) \sum_{l=1}^{L} I_{li}\{\frac{1}{A_l(x_i,\theta_n)} E_{\theta_n}[I_{lj}(\frac{\partial x_i\alpha_n}{\partial\theta} - \frac{\partial x_j\alpha_n}{\partial\theta})\frac{1}{a_n^{m+1}}\nabla K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n})|x_i] \\
&- \frac{\partial A_l(x_i,\theta_n)}{\partial\theta}(\frac{1}{A_l(x_i,\theta_n)})^2 E_{\theta_n}[I_{lj}\frac{1}{a_n^m}K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n})|x_i] \\
&- \frac{1}{B(x_i,\theta_n)} E_{\theta_n}[(\frac{\partial x_i\alpha_n}{\partial\theta} - \frac{\partial x_j\alpha_n}{\partial\theta})\frac{1}{a_n^{m+1}}\nabla K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n})|x_i] \\
&+ \frac{\partial B(x_i,\theta_n)}{\partial\theta}(\frac{1}{B(x_i,\theta_n)})^2 E_{\theta_n}[\frac{1}{a_n^m}K(\frac{x_i\alpha_n - x_j\alpha_n}{a_n})|x_i]\}
\end{aligned}
$$

and hence

$$
\begin{aligned}
&\lim_{n\to\infty} E_{\theta_n}(\Phi_2(x_i, I_i, x_j, I_j, a_n, \theta_n)|x_i, I_i) \\
&= \sum_{l=1}^{L} I_{li}\{\frac{1}{A_l(x_i,\theta_\circ)}\frac{\partial A_l(x_i,\theta_\circ)}{\partial\theta} - \frac{\partial A_l(x_i,\theta_\circ)}{\partial\theta}(\frac{1}{A_l(x_i,\theta_\circ)})^2 A_l(x_i,\theta_\circ) \\
&- \frac{1}{B(x_i,\theta_\circ)}\frac{\partial B(x_i,\theta_\circ)}{\partial\theta} + \frac{\partial B(x_i,\theta_\circ)}{\partial\theta}(\frac{1}{B(x_i,\theta_\circ)})^2 B(x_i,\theta)\} \\
&= 0.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&E_{\theta_n}(\Phi_1(z_j, I_j, z_i, I_i, a_n, \theta_n)|z_i, I_i) \\
&= I_{1i}\{E_{\theta_n}[T_1(z_j, I_j, z_i, a_n, \theta_n)|z_i] - E_{\theta_n}[T_2(z_j, I_j, z_i, a_n, \theta_n)|z_i] \\
&\quad - E_{\theta_n}[T_3(x_j, I_j, x_i, a_n, \theta_n)|x_i] + E_{\theta_n}[T_4(x_j, I_j, x_i, a_n, \theta_n)|x_i]\}
\end{aligned} \tag{7.18}
$$

where

$$
\begin{aligned}
T_1(z_j, I_j, z_i, a_n, \theta_n) &= I_{Z_{\theta_n}}(z_j) E_{\theta_n}[I_{1j}(\frac{\partial x_i\gamma_n}{\partial\theta} - \frac{\partial x_j\gamma_n}{\partial\theta}, \frac{\partial x_j\alpha_n}{\partial\theta} - \frac{\partial x_i\alpha_n}{\partial\theta})|y_j - x_j\gamma_n, x_j\alpha_n] \\
&\cdot \frac{1}{a_n^{m+k+1}}\bar{\nabla}\bar{K}(\frac{(y_j - x_j\gamma_n) - (y_i - x_i\gamma_n)}{a_n}, \frac{x_j\alpha_n - x_i\alpha_n}{a_n})\frac{1}{C(z_j,\theta_n)},
\end{aligned} \tag{7.19}
$$

$$
\begin{aligned}
T_2(z_j, I_j, z_i, a_n, \theta_n) &= I_{Z_{\theta_n}}(z_j) E_{\theta_n}[I_{1j}\frac{\partial C(z_j,\theta_n)}{\partial\theta}|y_j - x_j\gamma_n, x_j\alpha_n] \\
&\cdot \frac{1}{a_n^{m+k}}\bar{K}(\frac{(y_j - x_j\gamma_n) - (y_i - x_i\gamma_n)}{a_n}, \frac{x_j\alpha_n - x_i\alpha_n}{a_n})(\frac{1}{C(z_j,\theta_n)})^2,
\end{aligned} \tag{7.20}
$$

$$
T_3(x_j, I_j, x_i, a_n, \theta_n) = I_{X_{\theta_n}}(x_j) E_{\theta_n}[I_{1j}(\frac{\partial x_j\alpha_n}{\partial\theta} - \frac{\partial x_i\alpha_n}{\partial\theta})|x_j\alpha_n]\frac{1}{a_n^{m+1}}\nabla K(\frac{x_j\alpha_n - x_i\alpha_n}{a_n})\frac{1}{A_1(x_j,\theta_n)} \tag{7.21}
$$

and

$$T_4(x_j, I_j, x_i, a_n, \theta_n) = I_{X_{\theta_n}}(x_j) E_{\theta_n}[I_{1j} \frac{\partial A_1(x_j, \theta_n)}{\partial \theta} | x_j \alpha_n] \frac{1}{a_n^m} K(\frac{x_j \alpha_n - x_i \alpha_n}{a_n}) (\frac{1}{A_1(x_j, \theta_n)})^2 \tag{7.22}$$

As shown in Appendix 2, for each component $k$,

$$\lim_{n \to \infty} E_{\theta_n}(T_{1,k}(z_j, I_j, z_i, a_n, \theta_n) | z_i) = -tr \nabla E_{\theta_o}(\frac{\partial x \alpha_o}{\partial \theta_k} | x_i \alpha_o) \quad a.e., \tag{7.23}$$

$$\lim_{n \to \infty} E_{\theta_n}(T_{2,k}(z_j, I_j, z_i, a_n, \theta_n) | z_i) = -tr \nabla E_{\theta_o}(\frac{\partial x \alpha_o}{\partial \theta_k} | x_i \alpha_o) \quad a.e., \tag{7.24}$$

$$\lim_{n \to \infty} E_{\theta_n}(T_{3,k}(x_j, I_j, x_i, a_n, \theta_n) | x_i) = -tr \nabla E_{\theta_o}(\frac{\partial x \alpha_o}{\partial \theta_k} | x_i \alpha_o) \quad a.e., \tag{7.25}$$

and

$$\lim_{n \to \infty} E_{\theta_n}(T_{4,k}(x_j, I_j, x_i, a_n, \theta_n) | x_i) = -tr \nabla E_{\theta_o}(\frac{\partial x \alpha_o}{\partial \theta_k} | x_i \alpha_o) \quad a.e.. \tag{7.26}$$

Hence it follows that

$$\lim_{n \to \infty} E_{\theta_n}(\Phi_1(z_j, I_j, z_i, I_i, a_n, \theta_n) | z_i, I_i) = 0 \quad a.e.. \tag{7.27}$$

It remains to analyze the following term:

$$\begin{aligned} & E_{\theta_n}(\Phi_2(x_j, I_j, x_i, I_i, a_n, \theta_n) | x_i, I_i) \\ &= \sum_{l=1}^{L} \{ I_{li} E_{\theta_n}[W_1(x_j, I_{lj}, x_i, a_n, \theta_n) | x_i] - I_{li} E_{\theta_n}[W_2(x_j, I_{lj}, x_i, a_n, \theta_n) | x_i] \\ &- E_{\theta_n}[W_3(x_j, I_{lj}, x_i, a_n, \theta_n) | x_i] + E_{\theta_n}[W_4(x_j, I_{lj}, x_i, a_n, \theta_n) | x_i] \} \end{aligned} \tag{7.28}$$

where

$$\begin{aligned} & W_1(x_j, I_{lj}, x_i, a_n, \theta_n) \\ &= I_{X_{\theta_n}}(x_j) E_{\theta_n}[I_{lj}(\frac{\partial x_j \alpha_n}{\partial \theta} - \frac{\partial x_i \alpha_n}{\partial \theta}) | x_j \alpha_n] \frac{1}{a_n^{m+1}} \nabla K(\frac{x_j \alpha_n - x_i \alpha_n}{a_n}) \frac{1}{A_l(x_j, \theta_n)}, \end{aligned} \tag{7.29}$$

$$W_2(x_j, I_{lj}, x_i, a_n, \theta_n) = I_{X_{\theta_n}}(x_j) E_{\theta_n}[I_{lj} \frac{\partial A_l(x_j, \theta_n)}{\partial \theta} | x_j \alpha_n] \frac{1}{a_n^m} K(\frac{x_j \alpha_n - x_i \alpha_n}{a_n}) (\frac{1}{A_l(x_j, \theta_n)})^2, \tag{7.30}$$

$$\begin{aligned} & W_3(x_j, I_{lj}, x_i, a_n, \theta_n) \\ &= I_{X_{\theta_n}}(x_j) E_{\theta_n}[I_{lj}(\frac{\partial x_j \alpha_n}{\partial \theta} - \frac{\partial x_i \alpha_n}{\partial \theta}) | x_j \alpha_n] \frac{1}{a_n^{m+1}} \nabla K(\frac{x_j \alpha_n - x_i \alpha_n}{a_n}) \frac{1}{B(x_j, \theta_n)} \end{aligned} \tag{7.31}$$

and

$$W_4(x_j, I_{lj}, x_i, a_n, \theta_n) = I_{X_{\theta_n}}(x_j) E_{\theta_n}[I_{lj} \frac{\partial B(x_j, \theta_n)}{\partial \theta} | x_j \alpha_n] \frac{1}{a_n^m} K(\frac{x_j \alpha_n - x_i \alpha_n}{a_n})(\frac{1}{B(x_j, \theta_n)})^2. \tag{7.32}$$

As shown in Appendix 2, we have

$$\lim_{n \to \infty} E_{\theta_n}(W_{1,k}(x_j, I_{lj}, x_i, a_n, \theta_n) | x_i) = -tr \nabla E_{\theta_\circ}(\frac{\partial x \alpha_\circ}{\partial \theta_k} | x_i \alpha_\circ), \quad a.e., \tag{7.33}$$

$$\lim_{n \to \infty} E_{\theta_n}(W_{2,k}(x_j, I_{lj}, x_i, a_n, \theta_n) | x_i) = -tr \nabla E_{\theta_\circ}(\frac{\partial x \alpha_\circ}{\partial \theta_k} | x_i \alpha_\circ), \quad a.e., \tag{7.34}$$

$$\begin{aligned} &\lim_{n \to \infty} E_{\theta_n}(W_{3,k}(x_j, I_{lj}, x_i, a_n, \theta_n) | x_i) \\ &= -tr \nabla E_{\theta_\circ}(\frac{\partial x \alpha_\circ}{\partial \theta_k} | x_i \alpha_\circ) \cdot E_{\theta_\circ}(I_l | x_i \alpha_\circ) + (\frac{\partial x_i \alpha_\circ}{\partial \theta_k} - E_{\theta_\circ}(\frac{\partial x \alpha_\circ}{\partial \theta_k} | x_i \alpha_\circ)) \nabla E_{\theta_\circ}(I_l | x_i \alpha_\circ), \quad a.e., \end{aligned} \tag{7.35}$$

and

$$\lim_{n \to \infty} E_{\theta_n}(W_{4,k}(x_j, I_{lj}, x_i, a_n, \theta_n) | x_i) = -tr \nabla E_{\theta_\circ}(\frac{\partial x \alpha_\circ}{\partial \theta_k} | x_i \alpha_\circ) \cdot E_{\theta_\circ}(I_l | x_i \alpha_\circ), \quad a.e.. \tag{7.36}$$

Therefore,

$$\begin{aligned} &\lim_{n \to \infty} E_{\theta_n}(\Phi_2(x_j, I_j, x_i, I_i, a_n, \theta_n) | x_i, I_i) \\ &= -(\frac{\partial x_i \alpha_\circ}{\partial \theta} - E_{\theta_\circ}(\frac{\partial x \alpha_\circ}{\partial \theta} | x_i \alpha_\circ)) \sum_{l=1}^{L} \nabla E_{\theta_\circ}(I_l | x_i \alpha_\circ) \\ &= 0 \quad a.e., \end{aligned}$$

because $\sum_{l=1}^{L} E_{\theta_\circ}(I_l | x_i \alpha_\circ) = 1$ implies $\sum_{l=1}^{L} \nabla E_{\theta_\circ}(I_l | x_i \alpha_\circ) = 0$. Finally, Proposition 5 implies that, as $\lim_{n \to \infty} n a_n^{2s^*} = 0$ and $\lim_{n \to \infty} n a_n^{m+k+2} = \infty$, both terms on the right hand side of (7.17) converge to zero in $P_{\theta_n}$ probability.

The difference $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( L_{n,i}(\theta_n) - L^*_{n,i}(\theta_n) \right)$ can be shown to converge in $P_{\theta_n}$-probability to zero. Since $I_{\hat{Z}_{\theta_n}}(z) I_{Z_{\theta_n}}(z) = I_{\hat{Z}_{\theta_n}}(z)$ and $I_{\hat{X}_{\theta_n}}(x) I_{X_{\theta_n}}(x) = I_{\hat{X}_{\theta_n}}(x)$,

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( L_{n,i}(\theta_n) - L^*_{n,i}(\theta_n) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I_{\hat{Z}_{\theta_n}}(z_i) - I_{Z_{\theta_n}}(z_i)) \frac{1}{n-1} \sum_{j \neq i}^{n} \Phi_1(z_i, I_i, z_j, I_j, a_n, \theta_n) \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I_{\hat{X}_{\theta_n}}(x_i) - I_{X_{\theta_n}}(x_i)) \frac{1}{n-1} \sum_{j \neq i}^{n} \Phi_2(x_i, I_i, x_j, I_j, a_n, \theta_n). \end{aligned} \tag{7.37}$$

The absolute moment of the first term on the right hand side of (7.37) is less than

$$\begin{aligned}
&\sqrt{n}E_{\theta_n}|(I_{\hat{Z}_{\theta_n}}(z_i) - I_{Z_{\theta_n}}(z_i))\frac{1}{n-1}\sum_{j\neq i}^{n}\Phi_1(z_i,I_i,z_j,I_j,a_n,\theta_n)| \\
&\leq \{nE_{\theta_n}(I_{\hat{Z}_{\theta_n}}(z_i) - I_{Z_{\theta_n}}(z_i))E_{\theta_n}(\frac{1}{n-1}\sum_{j\neq i}^{n}\Phi_1(z_i,I_i,z_j,I_j,a_n,\theta_n))^2\}^{\frac{1}{2}}.
\end{aligned} \tag{7.38}$$

Since

$$\begin{aligned}
&E_{\theta_n}(\Phi_1^2(z_i,I_i,z_j,I_j,a_n,\theta_n)) \\
&= var_{\theta_n}(\Phi_1^2(z_i,I_i,z_j,I_j,a_n,\theta_n)) + E_{\theta_n}^2(\Phi_1(z_i,I_i,z_j,I_j,a_n,\theta_n)) \\
&= O(\frac{1}{a_n^{m+k+2}}) + O(a_n^{2s^*})
\end{aligned}$$

and

$$\begin{aligned}
&|E_{\theta_n}\{\Phi_1(z_i,I_i,z_j,I_j,a_n,\theta_n)\Phi_1'(z_i,I_i,z_k,I_k,a_n,\theta_n)\}| \\
&= |E_{\theta_n}\{E_{\theta_n}(\Phi_1(z_i,I_i,z_j,I_j,a_n,\theta_n)|z_i,I_i)E_{\theta_n}(\Phi_1'(z_i,I_i,z_k,I_k,a_n,\theta_n)|z_i,I_i)\}| \\
&\leq O(a_n^{2s^*}),
\end{aligned}$$

it follows that

$$\begin{aligned}
&E_{\theta_n}(\frac{1}{n-1}\sum_{j\neq i}^{n}\Phi_1(z_i,I_i,z_j,I_j,a_n,\theta_n))^2 \\
&= \frac{1}{(n-1)^2}\sum_{j\neq i}^{n}E_{\theta_n}(\Phi_1^2(z_i,I_i,z_j,I_j,a_n,\theta_n)) \\
&\quad + \frac{1}{(n-1)^2}\sum_{j\neq i}^{n}\sum_{k\neq i,j}^{n}E_{\theta_n}(\Phi_1(z_i,I_i,z_j,I_j,a_n,\theta_n)\Phi_1'(z_i,I_i,z_k,I_k,a_n,\theta_n)) \\
&= O(\frac{1}{na_n^{m+k+2}}) + O(\frac{a_n^{2s^*}}{n}) + O(a_n^{2s^*})
\end{aligned}$$

which converges to zero. The term $nE_{\theta_n}(I_{\hat{Z}_{\theta_n}}(z) - I_{Z_{\theta_n}}(z))$ in (7.38) is bounded. This follows from repeated applications of the following Proposition to each component of the indices.

**Proposition 6.** *Suppose that $z_{n1},\cdots,z_{nn}$ are independent random variables with common density function $g_n(z)$ and support $[T_{n1},T_{n2}]$. If $g_n(z)$ is bounded from above and away from zero on its support $[T_{n1},T_{n2}]$ uniformly in $n$, then*

$$nP(T_{n1}+\delta_n \leq z_{ni} \leq \min_{j\in\{1,\cdots,n\}\setminus\{i\}} z_{nj}+\delta_n)$$

*and*

$$nP(\max_{j\in\{1,\cdots,n\}\setminus\{i\}} z_{nj}-\delta_n \leq z_{ni} \leq T_{n2}-\delta_n)$$

*where $\delta_n \geq 0$, are bounded sequences.*

**Proof:**

See Appendix 1.

So the first term on the rignt hand side in (7.37) converges in absolute moment and hence in $P_{\theta_n}$-probability to zero. Similarly, the second term on the right hand side of (7.37) converges in $P_{\theta_n}$-probability to zero.

Finally, we conclude from (7.1) that

$$\sqrt{n}(S_n(\theta_n) - S_n^*(\theta_n)) \xrightarrow{P_{\theta_n}} 0.$$

By contiguity,

$$\sqrt{n}(S_n(\bar{\theta}_n) - S_n^*(\bar{\theta}_n)) \xrightarrow{P_{\theta_o}} 0. \tag{7.39}$$

With uniform convergence of the nonparametric functions in (6.12), (6.13), (6.14), (6.24), (6.25) and (6.28) and uniform boundedness of such functions from above and away from zero, $I_n(\bar{\theta}_n)$ in (2.15) will converge in probability to a well defined limit. Define

$$\begin{aligned}
&I_n^*(\theta_n) \\
&= \frac{1}{n}\sum_{i=1}^{n}\{I_{Z_{\theta_n}}(z_i)I_{1i}[\frac{1}{C(z_i,\theta_n)}\frac{\partial C(z_i,\theta_n)}{\partial\theta} - \frac{1}{A_1(x_i,\theta_n)}\frac{\partial A_1(x_i,\theta_n)}{\partial\theta}] \\
&\cdot[\frac{1}{C(z_i,\theta_n)}\frac{\partial C(z_i,\theta_n)}{\partial\theta'} - \frac{1}{A_1(x_i,\theta_n)}\frac{\partial A_1(x_i,\theta_n)}{\partial\theta'}] \\
&+ I_{X_{\theta_n}}(x_i)\sum_{l=1}^{L} I_{li}[\frac{1}{A_l(x_i,\theta_n)}\frac{\partial A_l(x_i,\theta_n)}{\partial\theta} - \frac{1}{B(x_i,\theta_n)}\frac{\partial B(x_i,\theta_n)}{\partial\theta}] \\
&\cdot[\frac{1}{A_l(x_i,\theta_n)}\frac{\partial A_l(x_i,\theta_n)}{\partial\theta'} - \frac{1}{B(x_i,\theta_n)}\frac{\partial B(x_i,\theta_n)}{\partial\theta'}]\}
\end{aligned} \tag{7.40}$$

This matrix is constructed with the nonparametric functions in $I_n(\theta_n)$ replaced by their limit functions. With uniform convergence and uniform boundedness of the nonparametric functions, $I_n(\theta_n) - I_n^*(\theta_n) \xrightarrow{P_{\theta_n}} 0$. By contiguity, $I_n(\theta_n) - I_n^*(\theta_n) \xrightarrow{P_{\theta_o}} 0$. By Chebyshev's weak law of large number, $I_n^*(\theta_n)$ converges in $P_{\theta_o}$ probability to $I(\theta_o)$ where

$$\begin{aligned}
&I(\theta_o) \\
&= E_{\theta_o}\{I_1[\frac{1}{C(z,\theta_o)}\frac{\partial C(z,\theta_o)}{\partial\theta} - \frac{1}{A_1(x,\theta_o)}\frac{\partial A_1(x,\theta_o)}{\partial\theta}][\frac{1}{C(z,\theta_o)}\frac{\partial C(z,\theta_o)}{\partial\theta'} - \frac{1}{A_1(x,\theta_o)}\frac{\partial A_1(x,\theta_o)}{\partial\theta'}] \\
&+ \sum_{l=1}^{L} I_l[\frac{1}{A_l(x,\theta_o)}\frac{\partial A_l(x,\theta_o)}{\partial\theta} - \frac{1}{B(x,\theta_o)}\frac{\partial B(x,\theta_o)}{\partial\theta}][\frac{1}{A_l(x,\theta_o)}\frac{\partial A_l(x,\theta_o)}{\partial\theta'} - \frac{1}{B(x,\theta_o)}\frac{\partial B(x,\theta_o)}{\partial\theta'}]\}.
\end{aligned} \tag{7.41}$$

Under the identification condition that $I(\theta_{\theta_o})$ is nonsingular,

$$I_n^{-1}(\theta_n) \xrightarrow{P_{\theta_o}} I^{-1}(\theta_o). \tag{7.42}$$

Hence we conclude from (3.11),

$$\sqrt{n}(\hat{\theta}_n - \theta_n^*) \xrightarrow{P_{\theta_o}} 0. \tag{7.43}$$

With the expressions in (6.33) and (6.34), the matrix $I(\theta_o)$ in (7.41) can be rewritten as

$$
\begin{aligned}
I(\theta_\circ) = E_{\theta_\circ}\{I_1\Big([\frac{\partial x\gamma_\circ}{\partial\theta}, -\frac{\partial x\alpha_\circ}{\partial\theta}] - E_{\theta_\circ}([\frac{\partial x\gamma_\circ}{\partial\theta},\ -\frac{\partial x\alpha_\circ}{\partial\theta}]|x\alpha_\circ)\Big) \\
\cdot \bar{\nabla}\ln f(y - x\gamma_\circ|I_1 = 1, x\alpha_\circ)\bar{\nabla}'\ln f(y - x\gamma_\circ|I_1 = 1, x\alpha_\circ) \\
\cdot \Big([\frac{\partial x\gamma_\circ}{\partial\theta},\ -\frac{\partial x\alpha_\circ}{\partial\theta}] - E_{\theta_\circ}([\frac{\partial x\gamma_\circ}{\partial\theta},\ -\frac{\partial x\alpha_\circ}{\partial\theta}]|x\alpha_\circ)\Big)' \\
+ [\frac{\partial x\alpha_\circ}{\partial\theta} - E_{\theta_\circ}(\frac{\partial x\alpha_\circ}{\partial\theta}|x\alpha_\circ)]\sum_{l=1}^{L} I_l\nabla\ln P(I_l = 1|x\alpha_\circ)\nabla'\ln P(I_l = 1|x\alpha_\circ) \\
\cdot [\frac{\partial x\alpha_\circ}{\partial\theta} - E_{\theta_\circ}(\frac{\partial x\alpha_\circ}{\partial\theta}|x\alpha_\circ)]'\}.
\end{aligned}
\tag{7.44}
$$

## 8. Asymptotic Distribution

The previous section has shown that $\sqrt{n}(\hat{\theta}_n - \theta_o)$ and $\sqrt{n}(\theta_n^* - \theta_o)$ have the same limiting distribution. In this section, we will derive the limiting distribution for $\sqrt{n}(\theta_n^* - \theta_o)$. We will show that under $P_{\theta_o}$, $\frac{\partial S_n^*(\bar{\theta}_n)}{\partial \theta'} \xrightarrow{P_{\theta_o}} -I(\theta_o)$ and $\sqrt{n}S_n^*(\theta_o) \xrightarrow{D} N(0, I(\theta_o))$.

With (6.33) and (6.34), $S_n^*(\theta)$ in (3.8) can be rewritten as

$$\begin{aligned} &S_n^*(\theta) \\ &= \frac{1}{n}\sum_{i=1}^{n}\{I_{1i}(E_\theta([\frac{\partial x\gamma}{\partial \theta}, -\frac{\partial x\alpha}{\partial \theta}]|x_i\alpha) - [\frac{\partial x_i\gamma}{\partial \theta}, -\frac{\partial x_i\alpha}{\partial \theta}])\bar{\nabla}\ln f(y_i - x_i\gamma|I_{1i} = 1, x_i\alpha) \\ &+ \sum_{l=1}^{L} I_{li}[\frac{\partial x_i\alpha}{\partial \theta} - E_\theta(\frac{\partial x\alpha}{\partial \theta}|x_i\alpha)]\nabla \ln P(I_l = 1|x_i\alpha)\}. \end{aligned} \tag{8.1}$$

It follows that, for each component $\theta_k$ of $\theta$,

$$\begin{aligned} &\frac{\partial S_n^*(\theta)}{\partial \theta_k} \\ &= \frac{1}{n}\sum_{i=1}^{n} I_{1i}[H_1(x_i, y_i, \theta) + H_2(x_i, y_i, \theta)] + \frac{1}{n}\sum_{i=1}^{n}[H_3(x_i, I_i, \theta) + H_4(x_i, I_i, \theta)] \end{aligned} \tag{8.2}$$

where

$$H_1(x_i, y_i, \theta) = \frac{\partial}{\partial \theta_k}(E_\theta([\frac{\partial x\gamma}{\partial \theta}, -\frac{\partial x\alpha}{\partial \theta}]|x_i\alpha) - [\frac{\partial x_i\gamma}{\partial \theta}, -\frac{\partial x_i\alpha}{\partial \theta}])\bar{\nabla}\ln f(y_i - x_i\gamma|I_{1i} = 1, x_i\alpha), \tag{8.3}$$

$$H_2(x_i, y_i, \theta) = (E_\theta([\frac{\partial x\gamma}{\partial \theta}, -\frac{\partial x\alpha}{\partial \theta}]|x_i\alpha) - [\frac{\partial x_i\gamma}{\partial \theta}, -\frac{\partial x_i\alpha}{\partial \theta}])\frac{\partial}{\partial \theta_k}\bar{\nabla}\ln f(y_i - x_i\gamma|I_{1i} = 1, x_i\alpha), \tag{8.4}$$

$$H_3(x_i, I_i, \theta) = \sum_{l=1}^{L} I_{li}\frac{\partial}{\partial \theta_k}[\frac{\partial x_i\alpha}{\partial \theta} - E_\theta(\frac{\partial x\alpha}{\partial \theta}|x_i\alpha)]\nabla \ln P(I_l = 1|x_i\alpha) \tag{8.5}$$

and

$$H_4(x_i, I_i, \theta) = \sum_{l=1}^{L} I_{li}[\frac{\partial x_i\alpha}{\partial \theta} - E_\theta(\frac{\partial x\alpha}{\partial \theta}|x_i\alpha)] \cdot \frac{\partial}{\partial \theta_k}\nabla \ln P(I_l = 1|x_i\alpha). \tag{8.6}$$

By the uniform law of large number for i.i.d. variables (Amemiya [1985]), $\frac{1}{n}\sum_{i=1}^{n} I_{1i}H_1(x_i, y_i, \theta)$ converges in $P_{\theta_o}$-probability to $E_{\theta_o}(I_1 H_1(x, y, \theta))$ uniformly in $\theta \in \Theta$ where

$$\begin{aligned} E_{\theta_o}(I_1H_1(x, y, \theta)) = E_{\theta_o}\{P(I_1 = 1|x_i) \cdot \frac{\partial}{\partial \theta_k}\left(E_\theta([\frac{\partial x\gamma}{\partial \theta}, -\frac{\partial x\alpha}{\partial \theta}]|x_i\alpha) - [\frac{\partial x_i\gamma}{\partial \theta}, -\frac{\partial x_i\alpha}{\partial \theta}]\right) \\ \cdot E_{\theta_o}[\bar{\nabla}\ln f(y_i - x_i\gamma|I_{1i} = 1, x_i\alpha)|I_{1i} = 1, x_i]\}. \end{aligned} \tag{8.7}$$

For each $\theta$, $\int f(y-x\gamma|I_1=1,x\alpha)dy=1$ which implies that, under Assumption 1 which guarantee that the differentiation operator and the integral operator are interchangeable,

$$\int \bar{\nabla}\ln f(y-x\gamma|I_1=1,x\alpha)\cdot f(y-x\gamma|I_1=1,x\alpha)dy=0. \tag{8.8}$$

At $\theta=\theta_\circ$, $f(y-x\gamma_\circ|I_1=1,x\alpha_\circ)$ is the conditional density of $y$ given $I_1=1$ and $x$ and hence $E_{\theta_\circ}(\bar{\nabla}\ln f(y-x\gamma_\circ|I_1=1,x\alpha_\circ)|x)=0$ for all $x$. It follows from (8.7) that $E_{\theta_\circ}(I_1H_1(x,y,\theta_\circ))=0$. By continuity of $E_{\theta_\circ}(I_1H_1(x,y,\theta))$ in $\theta$ and $\bar{\bar{\theta}}_n$ being a consistent estimate of $\theta_\circ$,

$$\frac{1}{n}\sum_{i=1}^{n}I_{1i}H_1(x_i,y_i,\bar{\bar{\theta}}_n)\xrightarrow{P_{\theta_\circ}}0. \tag{8.9}$$

Similarly, $P(I_l=1|x)=P(I_l=1|x\alpha_\circ)$ and $\sum_{l=1}^{L}\nabla P(I_l=1|x\alpha_\circ)=0$ imply

$$\begin{aligned}
&E_{\theta_\circ}(H_3(x,I,\theta_\circ))\\
&=E_{\theta_\circ}\{\sum_{l=1}^{L}P(I_l=1|x)\frac{\partial}{\partial\theta_k}[\frac{\partial x\alpha_\circ}{\partial\theta}-E_{\theta_\circ}(\frac{\partial x\alpha_\circ}{\partial\theta}|x\alpha_\circ)]\nabla\ln P(I_l=1|x\alpha_\circ)\}\\
&=E_{\theta_\circ}\{\frac{\partial}{\partial\theta_k}[\frac{\partial x\alpha_\circ}{\partial\theta}-E_{\theta_\circ}(\frac{\partial x\alpha_\circ}{\partial\theta}|x\alpha_\circ)]\sum_{l=1}^{L}\nabla P(I_l=1|x\alpha_\circ)\}\\
&=0,
\end{aligned}$$

and hence

$$\frac{1}{n}\sum_{i=1}^{n}H_3(x_i,I_i,\bar{\bar{\theta}}_n)\xrightarrow{P_{\theta_\circ}}0. \tag{8.10}$$

On the other hand, $\frac{1}{n}\sum_{i=1}^{n}I_{1i}H_2(x_i,y_i,\theta)$ converges to $E_{\theta_\circ}(I_1H_2(x,y,\theta))$ in $P_{\theta_\circ}$ probability uniformly in $\theta$ by the uniform law of large number for i.i.d. variables, where

$$\begin{aligned}
&E_{\theta_\circ}(I_1H_2(x,y,\theta))\\
&=E_{\theta_\circ}\{P(I_1=1|x_i)\Big(E_\theta([\frac{\partial x\gamma}{\partial\theta},-\frac{\partial x\alpha}{\partial\theta}]|x_i\alpha)-[\frac{\partial x_i\gamma}{\partial\theta},-\frac{\partial x_i\alpha}{\partial\theta}]\Big)\\
&\cdot E_{\theta_\circ}(\frac{\partial}{\partial\theta_k}\bar{\nabla}\ln f(y-x\gamma|I_1=1,x\alpha)|I_{1i}=1,x_i)\}.
\end{aligned} \tag{8.11}$$

The identity in (8.8) implies that, by differentiation,

$$\begin{aligned}
&\int[\frac{\partial}{\partial\theta'}\bar{\nabla}\ln f(y-x\gamma|I_1=1,x\alpha)+\bar{\nabla}\ln f(y-x\gamma|I_1=1,x\alpha)\frac{\partial}{\partial\theta'}\ln f(y-x\gamma|I_1=1,x\alpha)]\\
&\quad\cdot f(y-x\gamma|I_1=1,x\alpha)dy\\
&=0.
\end{aligned} \tag{8.12}$$

Since $f(y-x\gamma|I_1=1,x\alpha)$ depends on $\theta$ only through $x\gamma$ and $x\alpha$,

$$\frac{\partial}{\partial\theta'}\ln f(y-x\gamma|I_1=1,x\alpha)=\bar{\nabla}'\ln f(y-x\gamma|I_1=1,x\alpha)(-\frac{\partial x\gamma}{\partial\theta},\frac{\partial x\alpha}{\partial\theta})'. \tag{8.13}$$

Substituting (8.13) into (8.12),

$$\int \frac{\partial}{\partial\theta'}\bar{\nabla}\ln f(y - x\gamma|I_1 = 1, x\alpha)\cdot f(y - x\gamma|I_1 = 1, x\alpha)dy$$
$$= -\int \bar{\nabla}\ln f(y - x\gamma|I_1 = 1, x\alpha)\bar{\nabla}'\ln f(y - x\gamma|I_1 = 1, x\alpha)\cdot f(y - x\gamma|I_1 = 1, x\alpha)dy$$
$$\cdot(-\frac{\partial x\gamma}{\partial\theta}, \frac{\partial x\alpha}{\partial\theta})'$$

and hence

$$E_{\theta_o}(\frac{\partial}{\partial\theta'}\bar{\nabla}\ln f(y - x\gamma_o|I_1 = 1, x\alpha_o)|I_1 = 1, x)$$
$$= -E_{\theta_o}[\bar{\nabla}\ln f(y - x\gamma_o|I_1 = 1, x\alpha_o)\bar{\nabla}'\ln f(y - x\gamma_o|I_1 = 1, x\alpha_o)|I_1 = 1, x](-\frac{\partial x\gamma_o}{\partial\theta}, \frac{\partial x\alpha_o}{\partial\theta})'. \tag{8.14}$$

Because $E_{\theta_o}[\bar{\nabla}\ln f(y - x\gamma_o|I_1 = 1, x\alpha_o)\bar{\nabla}'\ln f(y - x\gamma_o|I_1 = 1, x\alpha_o)|I_1 = 1, x]$ is a function of $x\alpha_o$,

$$\begin{aligned} &E_{\theta_o}\{P(I_1 = 1|x_i\alpha_o)\Big(E_{\theta_o}([\frac{\partial x\gamma_o}{\partial\theta}, -\frac{\partial x\alpha_o}{\partial\theta}]|x_i\alpha_o) - [\frac{\partial x_i\gamma_o}{\partial\theta}, -\frac{\partial x_i\alpha_o}{\partial\theta}]\Big) \\ &\cdot E_{\theta_o}[\bar{\nabla}\ln f(y - x\gamma_o|I_1 = 1, x\alpha_o)\bar{\nabla}'\ln f(y - x\gamma_o|I_1 = 1, x\alpha_o)|I_{1i} = 1, x_i] \\ &\cdot E_{\theta_o}([-\frac{\partial x\gamma_o}{\partial\theta}, \frac{\partial x\alpha_o}{\partial\theta}]'|x_i\alpha_o)\} \\ &= 0. \end{aligned} \tag{8.15}$$

It follows from (8.14) and (8.15) that, at $\theta = \theta_o$, (8.11) becomes

$$\begin{aligned} &E_{\theta_o}(I_1 H_2(x, y, \theta_o)) \\ &= -E_{\theta_o}\{P(I_1 = 1|x_i\alpha_o)\Big(E_{\theta_o}([\frac{\partial x\gamma_o}{\partial\theta}, -\frac{\partial x\alpha_o}{\partial\theta}]|x_i\alpha_o) - [\frac{\partial x_i\gamma_o}{\partial\theta}, -\frac{\partial x_i\alpha_o}{\partial\theta}]\Big) \\ &\cdot E_{\theta_o}[\bar{\nabla}\ln f(y - x\gamma_o|I_1 = 1, x\alpha_o)\bar{\nabla}'\ln f(y - x\gamma_o|I_1 = 1, x\alpha_o)|I_{1i} = 1, x_i\alpha_o] \\ &\Big(E_{\theta_o}([\frac{\partial x\gamma_o}{\partial\theta_k}, -\frac{\partial x_i\alpha_o}{\partial\theta_k}]|x_i\alpha_o) - [\frac{\partial x_i\gamma_o}{\partial\theta_k}, -\frac{\partial x_i\alpha_o}{\partial\theta_k}]\Big)'\}. \end{aligned} \tag{8.16}$$

Similarly, $\frac{1}{n}\sum_{i=1}^{n} H_4(x_i, I_i, \theta)$ converges uniformly in $P_{\theta_o}$ probability to $E_{\theta_o}(H_4(x_i, I_i, \theta))$ in $\theta$ where

$$\begin{aligned} &E_{\theta_o}(H_4(x_i, I_i, \theta)) \\ &= E_{\theta_o}\{[\frac{\partial x_i\alpha}{\partial\theta} - E_\theta(\frac{\partial x\alpha}{\partial\theta}|x_i\alpha)]\sum_{l=1}^{L} P(I_l = 1|x_i\alpha_o)\frac{\partial}{\partial\theta_k}\nabla\ln P(I_l = 1|x_i\alpha)\} \\ &= E_{\theta_o}\{[\frac{\partial x_i\alpha}{\partial\theta} - E_\theta(\frac{\partial x\alpha}{\partial\theta}|x_i\alpha)]\sum_{l=1}^{L} P(I_l = 1|x_i\alpha_o)\nabla^2\ln P(I_l = 1|x_i\alpha)\cdot(\frac{\partial x_i\alpha}{\partial\theta_k})'\} \end{aligned} \tag{8.17}$$

and $\nabla^2\ln P(\cdot|t) = \frac{\partial^2}{\partial t'\partial t}\ln P(\cdot|t)$, because $P(I_l = 1|x\alpha)$ depends on $\theta$ only through $x\alpha$. Since $\sum_{l=1}^{L} P(I_l = 1|x\alpha) = 1$,

$$\sum_{l=1}^{L} \nabla \ln P(I_l = 1|x\alpha) \cdot P(I_l = 1|x\alpha) = 0 \tag{8.18}$$

and

$$\sum_{l=1}^{L} [\nabla^2 \ln P(I_l = 1|x\alpha) + \nabla \ln P(I_l = 1|x\alpha)\nabla' \ln P(I_l = 1|x\alpha)]P(I_l = 1|x\alpha) \\ = 0. \tag{8.19}$$

At $\theta = \theta_o$, the equation (8.17) becomes

$$\begin{aligned}
&E_{\theta_o}(H_4(x_i, I_i, \theta_o)) \\
&= -E_{\theta_o}\{[\frac{\partial x_i\alpha_o}{\partial \theta} - E_{\theta_o}(\frac{\partial x\alpha_o}{\partial \theta}|x_i\alpha_o)] \\
&\quad \cdot \sum_{l=1}^{L} P(I_l = 1|x_i\alpha_o)\nabla \ln P(I_l = 1|x_i\alpha_o)\nabla' \ln P(I_l = 1|x_i\alpha_o)(\frac{\partial x_i\alpha_o}{\partial \theta_k})'\} \\
&= -E_{\theta_o}\{[\frac{\partial x_i\alpha_o}{\partial \theta} - E_{\theta_o}(\frac{\partial x\alpha_o}{\partial \theta}|x_i\alpha_o)] \\
&\quad \cdot \sum_{l=1}^{L} P(I_l = 1|x_i\alpha_o)\nabla \ln P(I_l = 1|x_i\alpha_o)\nabla' \ln P(I_l = 1|x_i\alpha_o)[\frac{\partial x_i\alpha_o}{\partial \theta_k} - E_{\theta_o}(\frac{\partial x\alpha_o}{\partial \theta_k}|x_i\alpha_o)]'\}
\end{aligned} \tag{8.20}$$

Combining these results,

$$\frac{\partial S_n^*(\bar{\theta}_n)}{\partial \theta'} \xrightarrow{P_{\theta_o}} -I(\theta_o). \tag{8.21}$$

At $\theta = \theta_o$,

$$\sqrt{n}S_n^*(\theta_o) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (S_{1i} + S_{2i})$$

where

$$S_{1i} = I_{1i}\Big(E_{\theta_o}([\frac{\partial x\gamma_o}{\partial \theta}, -\frac{\partial x\alpha_o}{\partial \theta}]|x_i\alpha_o) - [\frac{\partial x_i\gamma_o}{\partial \theta}, -\frac{\partial x_i\alpha_o}{\partial \theta}]\Big) \cdot \bar{\nabla} \ln f(y_i - x_i\gamma_o|I_{1i} = 1, x_i\alpha_o) \tag{8.22}$$

and

$$S_{2i} = \sum_{l=1}^{L} I_{li}[\frac{\partial x_i\alpha_o}{\partial \theta} - E_{\theta_o}(\frac{\partial x\alpha_o}{\partial \theta}|x_i\alpha_o)]\nabla P(I_l = 1|x_i\alpha_o). \tag{8.23}$$

The identity (8.8) implies that

$$
\begin{aligned}
E_{\theta_o}(S_{1i}) = E_{\theta_o}\Big(& P(I_1 = 1|x_i)\{E_{\theta_o}([\frac{\partial x\gamma_o}{\partial \theta}, -\frac{\partial x\alpha_o}{\partial \theta}]|x_i\alpha_o) - [\frac{\partial x_i\gamma_o}{\partial \theta}, -\frac{\partial x_i\alpha_o}{\partial \theta}]\} \\
& \cdot E_{\theta_o}(\bar{\nabla} \ln f(y_i - x_i\gamma_o|I_{1i} = 1, x_i\alpha_o)|I_{1i} = 1, x_i)\Big) \\
= 0&.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
E_{\theta_o}(S_{2i}) &= \sum_{l=1}^{L} E_{\theta_o}\{P(I_l = 1|x_i)[\frac{\partial x_i\alpha_o}{\partial \theta} - E_{\theta_o}(\frac{\partial x\alpha_o}{\partial \theta}|x_i\alpha_o)]\nabla \ln P(I_l = 1|x_i\alpha_o)\} \\
&= E_{\theta_o}\{[\frac{\partial x_i\alpha_o}{\partial \theta} - E_{\theta_o}(\frac{\partial x\alpha_o}{\partial \theta}|x_i\alpha_o)]\sum_{l=1}^{L} \nabla P(I_l = 1|x_i\alpha_o)\} \\
&= 0.
\end{aligned}
$$

Thus, $S_n^*(\theta_o)$ has zero mean. The identity (8.8) implies also that $S_{1i}$ and $S_{2i}$ are uncorrelated:

$$
\begin{aligned}
& E_{\theta_o}(S_{1i}S_{2i}') \\
&= E_{\theta_o}\{I_{1i}\Big(E_{\theta_o}([\frac{\partial x\gamma_o}{\partial \theta}, -\frac{\partial x\alpha_o}{\partial \theta}]|x_i\alpha_o) - [\frac{\partial x_i\gamma_o}{\partial \theta}, -\frac{\partial x_i\alpha_o}{\partial \theta}]\Big) \\
&\quad \cdot \bar{\nabla} \ln f(y_i - x_i\gamma_o|I_{1i} = 1, x_i\alpha_o)\nabla' \ln P(I_l = 1|x_i\alpha_o) \cdot [\frac{\partial x_i\alpha_o}{\partial \theta} - E_{\theta_o}(\frac{\partial x\alpha_o}{\partial \theta}|x_i\alpha_o)]\} \\
&= E_{\theta_o}\{(E_{\theta_o}([\frac{\partial x\gamma_o}{\partial \theta}, -\frac{\partial x\alpha_o}{\partial \theta}]|x_i\alpha_o) - [\frac{\partial x_i\gamma_o}{\partial \theta}, -\frac{\partial x_i\alpha_o}{\partial \theta}]) \\
&\quad \cdot E_{\theta_o}(\bar{\nabla} \ln f(y_i - x_i\gamma_o|I_{1i} = 1, x_i\alpha_o)|I_{1i} = 1, x_i) \cdot \nabla' P(I_l = 1|x_i\alpha_o)[\frac{\partial x_i\alpha_o}{\partial \theta} - E_{\theta_o}(\frac{\partial x\alpha_o}{\partial \theta}|x_i\alpha_o)]\} \\
&= 0.
\end{aligned}
$$

The variance matrix of $S_{1i} + S_{2i}$ is apparently $I(\theta_o)$. Finally, the central limit theorem for i.i.d. samples implies

$$
\sqrt{n} S_n^*(\theta_o) \xrightarrow{D} N(0, I(\theta_o)). \tag{8.24}
$$

Since $\bar{\theta}_n$ is $\sqrt{n}$-consistent, it follows from (7.42) and and (8.21) that

$$
(I + I_n^{-1}(\theta_n)\frac{\partial S_n^*(\bar{\bar{\theta}}_n)}{\partial \theta'}) \cdot \sqrt{n}(\bar{\theta}_n - \theta_o) \xrightarrow{P_{\theta_o}} 0.
$$

By Slutsky's lemma, we conclude from (3.15) that

$$
\sqrt{n}(\theta_n^* - \theta_o) \xrightarrow{D} N(0, I^{-1}(\theta_o)). \tag{8.25}
$$

## Appendix 1.

**Proof of Proposition 2:**

$$\begin{aligned}
&E_{\theta_n}(c_n(z,z_i)\frac{1}{a_n^{\bar{m}}}K(\frac{t_n(z_i)-t_n(z)}{a_n}|z_i)\\
&=\int_{T_{1,n}}^{T_{2,n}} E_{\theta_n}(c_n(z,z_i)|t,z_i)\frac{1}{a_n^{\bar{m}}}K(\frac{t_n(z_i)-t}{a_n})g_n(t)dt\\
&=\int_{\frac{1}{a_n}(t_n(z_i)-T_{2,n})}^{\frac{1}{a_n}(t_n(z_i)-T_{1,n})} E_{\theta_n}(c_n(z,z_i)|t_n(z_i)-a_nw,z_i)g_n(t_n(z_i)-a_nw)K(w)dw.
\end{aligned}$$

For any $z_i \in Z_n$, $\frac{1}{a_n}(t_n(z_i)-T_{1,n}) \geq \frac{\delta_n}{a_n}$ and $\frac{1}{a_n}(t_n(z_i)-T_{2,n}) \leq -\frac{\delta_n}{a_n}$. Since $\frac{\delta_n}{a_n}$ tends to infinity, when $n$ is large enough, $D$ will be contained in the rectangle $[-\frac{\delta_n}{a_n}, \ \frac{\delta_n}{a_n}]$. Hence, when $n$ is large enough, $t_n(z_i)-a_nw$ for all $w \in D$ are in the interior of $[T_{1,n}, \ T_{2,n}]$ and

$$\begin{aligned}
&\int_{\frac{1}{a_n}(t_n(z_i)-T_{2,n})}^{\frac{1}{a_n}(t_n(z_i)-T_{1,n})} E_{\theta_n}(c_n(z,z_i)|t_n(z_i)-a_nw,z_i)g_n(t_n(z_i)-a_nw)K(w)dw\\
&=\int_D E_{\theta_n}(c_n(z,z_i)|t_n(z_i)-a_nw,z_i)g_n(t_n(z_i)-a_nw)K(w)dw
\end{aligned}$$

for all $z_i \in Z_n$. Therefore,

$$\begin{aligned}
&\sup_{z_i\in Z_n} |E_{\theta_n}(c_n(z,z_i)\frac{1}{a_n^{\bar{m}}}K(\frac{t_n(z_i)-t_n(z)}{a_n})|z_i) - E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)g_n(t_n(z_i))|\\
&\leq \sup_{z_i\in Z_n}\sup_{w\in D} |E_{\theta_n}(c_n(z,z_i)|t_n(z_i)-a_nw,z_i)g_n(t_n(z_i)-a_nw) - E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)g_n(t_n(z_i))|\\
&\quad\cdot\int_D |K(v)|dv
\end{aligned}$$

which converges to zero by uniform continuity and uniform boundedness of the functions.

Let $R_n(z_i,a_nw) = E_{\theta_n}(c_n(z,z_i)|t_n(z_i)-a_nw,z_i)g_n(t_n(z_i)-a_nw)$. Let $R_{n,i_1\cdots i_k}(z_i,t)$ denote the kth partial derivatives of $R_n(z_i,t)$ with respect to the components $t_{i_1},\cdots,t_{i_k}$ where $t=(t_1,\cdots,t_{\bar{m}})$. By a Taylor expansion up to order $s^*$,

$$R_n(z,a_nw) = R_n(z,0) + \frac{1}{i!}\sum_{i=1}^{s^*-1}\frac{\partial^i R_n(z,0)}{\partial a_n^i}a_n^i + \frac{1}{s^*!}\frac{\partial^{s^*}R_n(z,\hat{a_n}w)}{\partial a_n^{s^*}}a_n^{s^*}.$$

We note that $\frac{\partial^i}{\partial a_n^i}R_n(z,a_nw) = \sum_{j_1=1}^{\bar{m}}\cdots\sum_{j_i=1}^{\bar{m}} R_{n,j_1\cdots j_i}(z,a_nw)w_{j_1}\cdots w_{j_i}$. As the first $s^*-1$ moments of $K(w)$ are zero, $\int_D \frac{\partial^i R_n(z,0)}{\partial a_n^i}dw = 0, \ 1 \leq i \leq s^*-1$. Hence

$$\int_D [R_n(z_i,a_nw) - R_n(z_i,0)]K(w)dw = a_n^{s^*}\frac{1}{s^*!}\int_D \frac{\partial^{s^*}R_n(z_i,\hat{a_n}w)}{\partial a_n^{s^*}}K(w)dw$$

and

$$\sup_{z_i \in Z_n} \left| \int_D [E_{\theta_n}(c_n(z,z_i,)|t_n(z_i) - a_n w, z_i) g_n(t_n(z_i) - a_n w) \right.$$
$$\left. - E_{\theta_n}(c_n(z,z_i)|t_n(z_i), z_i) g_n(t_n(z_i))] K(w) dw \right|$$
$$\leq a_n^{s^*} \frac{1}{s^*!} \sum_{i_1=1}^{\bar{m}} \cdots \sum_{i_{s^*}=1}^{\bar{m}} \sup_{z_i \in Z_n} \sup_t |R_{n,i_1 \cdots i_{s^*}}(z_i,t)| \cdot \int_D \| w \|^{s^*} |K(w)| dw$$
$$= O(a_n^{s^*})$$

by uniform boundedness of the derivatives.

Q.E.D.

**Proof of Proposition 3**

$$E_{\theta_n}(c_n(z,z_i) \frac{1}{a_n^{\bar{m}+1}} \frac{\partial}{\partial w} K(\frac{t_n(z_i) - t_n(z)}{a_n}) | z_i)$$
$$= \frac{1}{a_n} \int_{\frac{1}{a_n}(t_n(z_i) - T_{2,n})}^{\frac{1}{a_n}(t(z_i) - T_{1,n})} E_{\theta_n}(c_n(z,z_i)|t_n(z_i) - a_n w, z_i) \cdot g_n(t_n(z_i) - a_n w) \frac{\partial}{\partial w} K(w) dw$$
$$= \frac{1}{a_n} \int_D E_{\theta_n}(c_n(z,z_i)|t_n(z_i) - a_n w, z_i) \cdot g_n(t_n(z_i) - a_n w) \frac{\partial}{\partial w} K(w) dw$$

for all $z_i \in Z_n$ when $n$ is large enough. By Taylor expansion at $a_n = 0$,

$$E_{\theta_n}(c_n(z,z_i)|t_n(z_i) - a_n w, z_i) \cdot g_n(t_n(z_i) - a_n w)$$
$$= E_{\theta_n}(c_n(z,z_i)|t_n(z_i), z_i) \cdot g_n(t_n(z_i))$$
$$- a_n [g_n(t_n(z_i) - \hat{a}_n w) \frac{\partial}{\partial t} E_{\theta_n}(c_n(z,z_i)|t(z_i) - \hat{a}_n w, z_i)$$
$$+ E_{\theta_n}(c_n(z,z_i)|t_n(z_i) - \hat{a}_n w, z_i) \cdot \frac{\partial}{\partial t} g_n(t_n(z_i) - \hat{a}_n w)] w'.$$

Since $K(w)$ goes to zero at the boundary of $D$, $\int_D \frac{\partial}{\partial w} K(w) dw = 0$. It follows that

$$E_{\theta_n}(c_n(z,z_i) \frac{1}{a_n^{\bar{m}+1}} \frac{\partial}{\partial w} K(\frac{t_n(z_i) - t_n(z)}{a_n}) | z_i)$$
$$= - \int_D [g_n(t_n(z_i) - \hat{a}_n w) \frac{\partial}{\partial t} E_{\theta_n}(c_n(z,z_i)|t_n(z_i) - \hat{a}_n w, z_i)$$
$$+ E_{\theta_n}(c_n(z,z_i)|t_n(z_i) - \hat{a}_n w, z_i) \frac{\partial}{\partial t} g_n(t_n(z_i) - \hat{a}_n w)] w' \frac{\partial}{\partial w} K(w) dw$$

By integration by parts, $\int_D w' \frac{\partial}{\partial w} K(w) dw = -I$. Therefore,

$$\sup_{z_i\in Z_n} |E_{\theta_n}(c_n(z,z_i)\frac{1}{a_n^{\bar{m}+1}}\frac{\partial}{\partial w}K(\frac{t_n(z_i)-t_n(z)}{a_n}|z_i)$$
$$- g_n(t_n(z_i))\frac{\partial}{\partial t}E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i) - E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)\frac{\partial}{\partial t}g_n(t_n(z_i))|$$
$$= \sup_{z_i\in Z_n} |\int_D [g_n(t_n(z_i))\frac{\partial}{\partial t}E_{\theta_n}(c_n(z,z_i)|t(z_i),z_i) + E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)\frac{\partial}{\partial t}g_n(t_n(z_i))$$
$$- g_n(t_n(z_i)-\bar{a}_n w)\frac{\partial}{\partial t}E_{\theta_n}(c_n(z,z_i)|t_n(z_i)-\bar{a}_n w,z_i)$$
$$- E_{\theta_n}(c_n(z,z_i)|t_n(z_i)-\bar{a}_n w,z_i)\frac{\partial}{\partial t}g_n(t_n(z_i)-\bar{a}_n w)]\cdot w'\frac{\partial}{\partial w}K(w)dw|$$
$$\leq \{ \sup_{z_i\in Z_n}\sup_{w\in D} \| g_n(t_n(z_i))\frac{\partial}{\partial t}E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)$$
$$- g_n(t_n(z_i)-\bar{a}_n w)\frac{\partial}{\partial t}E_{\theta_n}(c_n(z,z_i)|t_n(z_i)-\bar{a}_n w,z_i) \|$$
$$+ \sup_{z_i\in Z_n}\sup_{w\in D} \| E_{\theta_n}(c_n(z,z_i)|t_n(z_i),z_i)\frac{\partial}{\partial t}g_n(t_n(z_i))$$
$$- E_{\theta_n}(c_n(z,z_i)|t_n(z_i)-\bar{a}_n w,z_i)\frac{\partial}{\partial t}g_n(t_n(z_i)-\bar{a}_n w) \| \}$$
$$\cdot \int_D \| w \| \cdot \| \frac{\partial}{\partial w}K(w) \| dw$$

which converges to zero under our assumptions.

Let

$$R_n(z_i,a_n w) = \frac{\partial}{\partial t}[g_n(t)E_{\theta_n}(c_n(z,z_i)|t,z_i)]|_{t=t_n(z_i)-a_n w}.$$

By Taylor expansion up to order $s^*$,

$$R_n(z,a_n w)$$
$$= R_n(z,0) + \frac{1}{i!}\sum_{i=1}^{s^*-1}\frac{\partial^i R_n(z,0)}{\partial a_n^i}a_n^i + \frac{1}{s^*!}\frac{\partial^{s^*}R_n(z,\bar{a}_n w)}{\partial a_n^{s^*}}a_n^{s^*}$$
$$= R_n(z,0) + \frac{1}{i!}\sum_{i=1}^{s^*-1}(\sum_{j_1=1}^{\bar{m}}\cdots\sum_{j_i=1}^{\bar{m}} R_{n,j_1\cdots j_i}(z,0)w_{j_1}\cdots w_{j_i})a_n^i$$
$$+ \frac{1}{s^*!}\sum_{j_1=1}^{\bar{m}}\cdots\sum_{j_{s^*}=1}^{\bar{m}} R_{n,j_1\cdots j_{s^*}}(z,\bar{a}_n w)w_{j_1}\cdots w_{j_{s^*}}a_n^{s^*}$$

where $R_{n,j_1\cdots j_i}(z,t) = \frac{\partial^i R_n(z,t)}{\partial t_{j_1}\cdots\partial t_{j_i}}$. For any $j \notin \{j_1\cdots,j_i\}$,

$$\int_D w_{j_1}w_{j_2}\cdots w_{j_i}\frac{\partial}{\partial w_j}K(w)dw$$
$$= \int_D w_{j_1}\cdots w_{j_i}\int\frac{\partial}{\partial w_j}K(w)dw_j dw_1\cdots dw_{j-1}dw_{j+1}\cdots dw_{\bar{m}}$$
$$= 0.$$

Also, for any $p \geq 1$, $j \notin \{j_1, \cdots, j_i\}$ and $1 \leq i + p - 1 \leq s^* - 1$,

$$
\begin{aligned}
&\int_D w_{j_1} w_{j_2} \cdots w_{j_i} w_j^p \frac{\partial}{\partial w_j} K(w) dw \\
&= \int_D w_{j_1} \cdots w_{j_i} \int w_j^p \frac{\partial}{\partial w_j} K(w) dw_j dw_1 \cdots dw_{j-1} dw_{j+1} \cdots dw_{\bar{m}} \\
&= -p \int_D w_{j_1} \cdots w_{j_i} w_j^{p-1} K(w) dw \\
&= 0.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&| \int_D \{ \frac{\partial}{\partial t} [g_n(t) E_{\theta_n}(c_n(z, z_i)|t, z_i)]|_{t=t_n(z_i)-a_n w} \\
&\qquad - \frac{\partial}{\partial t} [g_n(t) E_{\theta_n}(c_n(z, z_i)|t, z_i)]|_{t=t_n(z_i)} \} w' \frac{\partial}{\partial w} K(w) dw | \\
&\leq a_n^{s^*} \frac{1}{s^*!} \sup_{z_i \in Z_n} \sup_{w \in D} \sum_{j_1=1}^{\bar{m}} \cdots \sum_{j_{s^*}=1}^{\bar{m}} \| R_{n,j_1 \cdots j_{s^*}}(z_i, \bar{a}_n w) \| \int_D \| w_{j_1} \cdots w_{j_{s^*}} w \| \cdot \| \frac{\partial}{\partial w} K(w) \| dw \\
&= O(a_n^{s^*}).
\end{aligned}
$$

Q.E.D.

**Proof of Proposition 4.**

For any $\delta > 0$, by Markov inequality and Cauchy inequality,

$$
\begin{aligned}
&P_{\theta_n}( \frac{1}{\sqrt{n}} \sum_{i=1}^n I_{Z_n}(z_i) | C_{1,n}(z_1, \cdots, z_n; z_i, \theta_n) - C_1(z_i, \theta_n) | \\
&\qquad \cdot | C_{2,n}(z_1, \cdots, z_n; z_i, \theta_n) - C_2(z_i, \theta_n) | > \delta ) \\
&\leq \frac{1}{\delta} \frac{1}{\sqrt{n}} \sum_{i=1}^n E_{\theta_n}( I_{Z_n}(z_i) | C_{1,n}(z_1, \cdots, z_n; z_i, \theta_n) - C_1(z_i, \theta_n) | \\
&\qquad \cdot | C_{2,n}(z_1, \cdots, z_n; z_i, \theta_n) - C_2(z_i, \theta_n) | ) \\
&\leq \frac{1}{\delta} \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ E_{\theta_n}( I_{Z_n}(z_i) | C_{1,n}(z_1, \cdots, z_n; z_i, \theta_n) - C_1(z_i, \theta_n) |^2 ) \\
&\qquad \cdot E_{\theta_n}( I_{Z_n}(z_i) | C_{2,n}(z_1, \cdots, z_n; z_i, \theta_n) - C_2(z_i, \theta_n) |^2 ) \}^{\frac{1}{2}}.
\end{aligned}
$$

The moment conditions (1) and (2) imply that

$$
\begin{aligned}
&E_{\theta_n}( I_{Z_n}(z_i) | C_{j,n}(z_1, \cdots, z_n; z_i, \theta_n) - C_j(z_i, \theta_n) |^2 ) \\
&= E_{\theta_n}( I_{Z_n}(z_i) [ Var_{\theta_n}( C_{j,n}(z_1, \cdots, z_n; z_i, \theta_n) | z_i ) + ( E_{\theta_n}( C_{j,n}(z_1, \cdots, z_n; z_i, \theta_n) | z_i ) - C_j(z_i, \theta_n) )^2 ] ) \\
&\leq O( \frac{1}{n a_n^{r_j}} ) + O( a_n^{2 s_j} ), \quad j = 1, 2.
\end{aligned}
$$

Hence

$$P_{\theta_n}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_{Z_n}(z_i)|C_{1,n}(z_1,\cdots,z_n;z_i,\theta_n)-C_1(z_i,\theta_n)|$$
$$\cdot|C_{2,n}(z_1,\cdots,z_n;z_i,\theta_n)-C_2(z_i,\theta_n)|>\delta)$$
$$\leq \frac{1}{\delta}[O(\frac{1}{na_n^{r_1+r_2}})+O(a_n^{2s_1-r_2})+O(a_n^{2s_2-r_1})+O(na_n^{2(s_1+s_2)})]^{1/2}$$

which converges to zero.

Q.E.D.

**Proof of Proposition 5:**

Let $U_n = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Phi_n(z_i,z_j,a_n,\theta_n)$. Define the symmetric function,

$$\Psi_n(z_i,z_j,a_n,\theta_n)=\frac{1}{2}(\Phi_n(z_i,z_j,a_n,\theta_n)+\Phi_n(z_j,z_i,a_n,\theta_n)).$$

$U_n$ can be rewritten in the standard $U$ statistic form:

$$U_n=\frac{1}{C_2^n}\sum_{i=1}^{n}\sum_{j>i}\Psi_n(z_i,z_j,a_n,\theta_n)$$

where $C_2^n=n(n-1)/2$. As

$$E_{\theta_n}(nU_n^2)=n\cdot Var_{\theta_n}(U_n)+n\cdot E_{\theta_n}{}^2(\Psi_n(z_1,z_2,a_n,\theta_n))$$
$$=n\cdot Var_{\theta_n}(U_n)+O(na_n^{2s})$$

and $na_n^{2s}\to 0$, it remains to show that $n\cdot var(U_n)$ goes to zero. From Hoeffding[1948],

$$n\cdot Var_{\theta_n}(U_n)=\frac{2}{(n-1)}[2(n-2)Var_{\theta_n}(E_{\theta_n}[\Psi_n(z_1,z_2,a_n,\theta_n)|z_1])+Var_{\theta_n}(\Psi_n(z_1,z_2,a_n,\theta_n))]$$
$$=\frac{4(n-2)}{(n-1)}Var_{\theta_n}(E_{\theta_n}[\Psi_n(z_1,z_2,a_n,\theta_n)|z_1])+\frac{2}{n-1}Var_{\theta_n}(\Psi_n(z_1,z_2,a_n,\theta_n)).$$

As

$$Var_{\theta_n}(E_{\theta_n}[\Psi_n(z_1,z_2,a_n,\theta_n)|z_1])$$
$$=E_{\theta_n}\{E_{\theta_n}{}^2[\Psi_n(z_1,z_2,a_n,\theta_n)|z_1]\}-E_{\theta_n}{}^2(\Psi_n(z_1,z_2,a_n,\theta_n))$$
$$=E_{\theta_n}\{E_{\theta_n}{}^2[\Psi_n(z_1,z_2,a_n,\theta_n)|z_1]\}-O(a_n^{2s}),$$

and

$$E_{\theta_n}{}^2[\Psi_n(z_1,z_2,a_n,\theta_n)|z_1]f_{\theta_n}(z_1)\leq\frac{M}{4}(h_1(z_1)+h_2(z_1))^2,$$

it follows by Lebesque dominated convergence theorem that

$$\lim_{n\to\infty}Var_{\theta_n}(E_{\theta_n}[\Psi_n(z_1,z_2,a_n,\theta_n)|z_1])$$
$$\leq\frac{M}{4}\int_{-\infty}^{\infty}[\lim_{n\to\infty}E_{\theta_n}(\Phi_n(z_1,z_2,a_n,\theta_n)|z_1)+\lim_{n\to\infty}E_{\theta_n}(\Phi_n(z_2,z_1,a_n,\theta_n)|z_1)]^2dz_1$$
$$=0.$$

Finally,

$$
\begin{aligned}
&\frac{2}{(n-1)} Var_{\theta_n}(\Psi_n(z_1, z_2, a_n, \theta_n)) \\
&\leq \frac{1}{2(n-1)} \{[Var_{\theta_n}(\Phi_n(z_1, z_2, a_n, \theta_n))]^{1/2} + [Var_{\theta_n}(\Phi_n(z_1, z_2, a_n, \theta_n))]^{1/2}\}^2 \\
&= O(\frac{1}{na_n^r})
\end{aligned}
$$

which converges to zero.

Q.E.D.

**Proof of Proposition 6:**

$$
\begin{aligned}
&nP(T_{n1} + \delta_n \leq z_{ni} \leq \min_{j \in \{1, \cdots, n\} \setminus \{i\}} z_{nj} + \delta_n) \\
&= n \int_{T_{n1}+\delta_n}^{T_{n2}} [1 - G_n(z - \delta_n)]^{n-1} g_n(z) dz \\
&= n \int_{T_{n1}}^{T_{n2}-\delta_n} [1 - G_n(u)]^{n-1} g_n(u + \delta_n) du.
\end{aligned}
$$

Since $g_n(z)$ is uniformly bounded from above and away from zero, there exists constants $c_1 > 0$ and $c_2 > 0$ such that $c_2 \leq g_n(z) \leq c_1$. It follows that for all $u \in [T_{n1}, T_{n2} - \delta_n]$, $g_n(u + \delta_n) \leq \frac{c_1}{c_2} g_n(u)$. Hence

$$
\begin{aligned}
&n \int_{T_{n1}}^{T_{n2}-\delta_n} [1 - G_n(u)]^{n-1} g_n(u + \delta_n) du \\
&\leq \frac{c_1}{c_2} \int_{T_{n1}}^{T_{n2}-\delta_n} n[1 - G_n(u)]^{n-1} g_n(u) du \\
&\leq \frac{c_1}{c_2} \int_{T_{n1}}^{T_{n2}} n[1 - G_n(u)]^{n-1} dG_n(u) \\
&= \frac{c_1}{c_2} \int_0^1 nx^{n-1} dx \\
&= \frac{c_1}{c_2}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&nP\left(\max_{j\in\{1,\cdots,n\}\setminus\{i\}} z_{nj} - \delta_n \leq z_{ni} \leq T_{n2} - \delta_n\right) \\
&= n\int_{T_{n1}}^{T_{n2}-\delta_n} [G_n(z+\delta_n)]^{n-1} g_n(z)dz \\
&= n\int_{T_{n1}+\delta_n}^{T_{n2}} [G_n(u)]^{n-1} g_n(u-\delta_n)du \\
&\leq \frac{c_1}{c_2}\int_{T_{n1}+\delta_n}^{T_{n2}} n[G_n(u)]^{n-1} g_n(u)du \\
&\leq \frac{c_1}{c_2}\int_0^1 nx^{n-1}dx \\
&= \frac{c_1}{c_2}.
\end{aligned}
$$

Q.E.D.

## Appendix 2.

1. DERIVATION OF (6.33) AND (6.34):

Because $E_\theta(x|I_1 = 1, y - x\gamma, x\alpha) = E_\theta(x|x\alpha)$, it implies that $\frac{\partial}{\partial\epsilon}E_\theta(x|I_1 = 1, \epsilon, x\alpha) = 0$. The equation (6.29) can be rewritten as

$$\begin{aligned}
&\frac{\partial C(z_i,\theta)}{\partial\theta_k}\\
&= \Big[-tr\nabla E_\theta(x\frac{\partial\alpha(\theta)}{\partial\theta_k}|x_i\alpha)\cdot E_\theta(I_1|y_i - x_i\gamma, x_i\alpha)\\
&\quad + E_\theta([(x - x_i)\frac{\partial\gamma(\theta)}{\partial\theta_k}, (x_i - x)\frac{\partial\alpha(\theta)}{\partial\theta_k}]|x_i\alpha)\bar{\nabla}E_\theta(I_1|y_i - x_i\gamma, x_i\alpha)\Big]g(y_i - x_i\gamma, x_i\alpha)\\
&\quad + E_\theta([(x - x_i)\frac{\partial\gamma(\theta)}{\partial\theta_k}, (x_i - x)\frac{\partial\alpha(\theta)}{\partial\theta_k}]|x_i\alpha)E_\theta(I_1|y_i - x_i\gamma, x_i\alpha)\bar{\nabla}g(y_i - x_i\gamma, x_i\alpha)\\
&= -tr\nabla E_\theta(x\frac{\partial\alpha(\theta)}{\partial\theta_k}|x_i\alpha)\cdot E_\theta(I_1|y_i - x_i\gamma, x_i\alpha)g(y_i - x_i\gamma, x_i\alpha)\\
&\quad + E_\theta([(x - x_i)\frac{\partial\gamma(\theta)}{\partial\theta_k}, (x_i - x)\frac{\partial\alpha(\theta)}{\partial\theta_k}]|x_i\alpha)\cdot\bar{\nabla}\{E_\theta(I_1|y_i - x_i\gamma, x_i\alpha)g(y_i - x_i\gamma, x_i\alpha)\}
\end{aligned} \tag{A.2.1}$$

Similarly, the equation (6.27) can be rewritten as

$$\begin{aligned}
&\frac{\partial A_l(x_i,\theta)}{\partial\theta_k}\\
&= -tr\nabla E_\theta(x\frac{\partial\alpha(\theta)}{\partial\theta_k}|x_i\alpha)\cdot E_\theta(I_l|x_i\alpha)p(x_i\alpha) + (x_i - E_\theta(x|x_i\alpha))\frac{\partial\alpha(\theta)}{\partial\theta_k}\cdot\nabla\{E_\theta(I_1|x_i\alpha)p(x_i\alpha)\}.
\end{aligned} \tag{A.2.2}$$

Since $E_\theta(I_1|y-x\gamma, x\alpha)g(y-x\gamma, x\alpha) = g(y-x\gamma, x\alpha|I_1 = 1)E_\theta(I_1)$ and $E_\theta(I_1|x\alpha)p(x\alpha) = p(x\alpha|I_1 = 1)E_\theta(I_1)$, it follows that

$$\begin{aligned}
&\bar{\nabla}'\ln\{E_\theta(I_1|y - x\gamma, x\alpha)g(y - x\gamma, x\alpha)\} - (0, \nabla'\ln\{E_\theta(I_1|x\alpha)p(x\alpha)\})\\
&= \bar{\nabla}'\ln g(y - x\gamma, x\alpha|I_1 = 1) - (0, \bar{\nabla}'\ln p(x\alpha|I_1 = 1))\\
&= \bar{\nabla}'\ln f(y - x\gamma|I_1 = 1, x\alpha).
\end{aligned}$$

Using these relations,

$$\begin{aligned}
&\frac{1}{C(z_i,\theta)}\frac{\partial C(z_i,\theta)}{\partial\theta_k} - \frac{1}{A_1(x_i,\theta)}\frac{\partial A_1(x_i,\theta)}{\partial\theta_k}\\
&= E_\theta([(x - x_i)\frac{\partial\gamma(\theta)}{\partial\theta_k}, (x_i - x)\frac{\partial\alpha(\theta)}{\partial\theta_k}]|x_i\alpha)\cdot\bar{\nabla}\ln\{E_\theta(I_1|y_i - x_i\gamma, x_i\alpha)g(y_i - x_i\gamma, x_i\alpha)\}\\
&\quad - (x_i - E_\theta(x|x_i\alpha))\frac{\partial\alpha(\theta)}{\partial\theta_k}\cdot\nabla\ln\{E_\theta(I_1|x_i\alpha)p(x_i\alpha)\}\\
&= \left\{E_\theta\left(\left[\frac{\partial x\gamma}{\partial\theta}, -\frac{\partial x\alpha}{\partial\theta}\right]|x_i\alpha\right) - [\frac{\partial x_i\gamma}{\partial\theta}, -\frac{\partial x_i\alpha}{\partial\theta}]\right\}\bar{\nabla}\ln f(y_i - x_i\gamma|I_{1i} = 1, x_i\alpha)
\end{aligned}$$

which is (6.33). Similarly, the equation(6.34) follows from (6.24) and (A.2.2).

2. DERIVATION OF (7.23)-(7.26):

Because $I_{Z_{\theta_n}}(z_j)$ and $C(z_j,\theta_n)$ are functions of $(y_j - x_j\gamma_n, x_j\alpha_n)$ and are continuous everywhere except the boundary points of the support, Proposition 3 implies that

$$
\begin{aligned}
&\lim_{n\to\infty} E_{\theta_n}(T_{1,k}(z_j,I_j,z_i,a_n,\theta_n)|z_i)\\
&= -g(y_i - x_i\gamma_o, x_i\alpha_o)\cdot tr\bar{\nabla}\left\{E_{\theta_o}[I_1(\frac{\partial x_i\gamma_o}{\partial\theta_k} - \frac{\partial x\gamma_o}{\partial\theta_k}, \frac{\partial x\alpha_o}{\partial\theta_k} - \frac{\partial x_i\alpha_o}{\partial\theta_k})|y_i - x_i\gamma_o, x_i\alpha_o]\frac{1}{C(z_i,\theta_o)}\right\}\\
&\quad - E_{\theta_o}[I_1(\frac{\partial x_i\gamma_o}{\partial\theta_k} - \frac{\partial x\gamma_o}{\partial\theta_k}, \frac{\partial x\alpha_o}{\partial\theta_k} - \frac{\partial x_i\alpha_o}{\partial\theta_k})|y_i - x_i\gamma_o, x_i\alpha_o]\frac{1}{C(z_i,\theta_o)}\cdot\bar{\nabla}g(y_i - x_i\gamma_o, x_i\alpha_o)\\
&= -g(y_i - x_i\gamma_o, x_i\alpha_o)\cdot tr\bar{\nabla}\left\{E_{\theta_o}[(\frac{\partial x_i\gamma_o}{\partial\theta_k} - \frac{\partial x\gamma_o}{\partial\theta_k}, \frac{\partial x\alpha_o}{\partial\theta_k} - \frac{\partial x_i\alpha_o}{\partial\theta_k})|x_i\alpha_o]\frac{1}{g(y_i - x_i\gamma_o, x_i\alpha_o)}\right\}\\
&\quad - E_{\theta_o}[(\frac{\partial x_i\gamma_o}{\partial\theta_k} - \frac{\partial x\gamma_o}{\partial\theta_k}, \frac{\partial x\alpha_o}{\partial\theta_k} - \frac{\partial x_i\alpha_o}{\partial\theta_k})|x_i\alpha_o]\frac{\bar{\nabla}g(y_i - x_i\gamma_o, x_i\alpha_o)}{g(y_i - x_i\gamma_o, x_i\alpha_o)}\\
&= -tr\nabla E_{\theta_o}[(\frac{\partial x\alpha_o}{\partial\theta_k} - \frac{\partial x_i\alpha_o}{\partial\theta_k})|x_i\alpha_o]\\
&\quad + \frac{1}{g(y_i - x_i\gamma_o, x_i\alpha_o)}E_{\theta_o}[(\frac{\partial x_i\gamma_o}{\partial\theta_k} - \frac{\partial x\gamma_o}{\partial\theta_k}, \frac{\partial x\alpha_o}{\partial\theta_k} - \frac{\partial x_i\alpha_o}{\partial\theta_k})|x_i\alpha_o]\cdot\bar{\nabla}g(y_i - x_i\gamma_o, x_i\alpha_o)\\
&\quad - E_{\theta_o}[(\frac{\partial x_i\gamma_o}{\partial\theta_k} - \frac{\partial x\gamma_o}{\partial\theta_k}, \frac{\partial x\alpha_o}{\partial\theta_k} - \frac{\partial x_i\alpha_o}{\partial\theta_k})|x_i\alpha_o]\frac{\bar{\nabla}g(y_i - x_i\gamma_o, x_i\alpha_o)}{g(y_i - x_i\gamma_o, x_i\alpha_o)}\\
&= -tr\nabla E_{\theta_o}[\frac{\partial x\alpha_o}{\partial\theta_k}|x_i\alpha_o]
\end{aligned}
$$

which is (7.23). Similarly,

$$
\begin{aligned}
&\lim_{n\to\infty} E_{\theta_n}(T_{3,k}(x_j,I_j,x_i,a_n,\theta_n)|x_i)\\
&= -p(x_i\alpha_o)\cdot tr\nabla\{E_{\theta_o}[I_1(x - x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\frac{1}{A_1(x_i,\theta_o)}\}\\
&\quad - E_{\theta_o}[I_1(x - x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\frac{1}{A_1(x_i,\theta_o)}\cdot\nabla p(x_i\alpha_o)\\
&= -p(x_i\alpha_o)\cdot tr\nabla\{E_{\theta_o}[(x - x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\frac{1}{p(x_i\alpha_o)}\} - E_{\theta_o}[(x - x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\frac{1}{p(x_i\alpha_o)}\cdot\nabla p(x_i\alpha_o)\\
&= -tr\nabla E_{\theta_o}(\frac{\partial x\alpha_o}{\partial\theta_k}|x_i\alpha_o) + \frac{1}{p(x_i\alpha_o)}E_{\theta_o}[(x - x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\nabla p(x_i\alpha_o)\\
&\quad - E_{\theta_o}[(x - x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\frac{\nabla p(x_i\alpha_o)}{p(x_i\alpha_o)}\\
&= -tr\nabla E_{\theta_o}(\frac{\partial x\alpha_o}{\partial\theta_k}|x_i\alpha_o) \qquad a.e.
\end{aligned}
$$

which is (7.25). On the other hand, Proposition 2 implies that

$$\begin{aligned}
&\lim_{n\to\infty} E_{\theta_n}(T_{2,k}(z_j, I_j, z_i, a_n, \theta_n)|z_i) \\
&= E_{\theta_o}[I_1 \frac{\partial C(z,\theta_o)}{\partial \theta_k}|y_i - x_i\gamma_o, x_i\alpha_o] \left(\frac{1}{C(z_i,\theta_o)}\right)^2 g(y_i - x_i\gamma_o, x_i\alpha_o) \\
&= E_{\theta_o}[\frac{\partial C(z,\theta_o)}{\partial \theta_k}|I_{1i} = 1, y_i - x_i\gamma_o, x_i\alpha_o]\frac{1}{C(z_i,\theta_o)} \qquad a.e.
\end{aligned}$$

With relation (A.2.1),

$$\begin{aligned}
&E_{\theta_o}[\frac{\partial C(z,\theta_o)}{\partial \theta_k}|I_{1i} = 1, y_i - x_i\gamma_o, x_i\alpha_o] \\
&= -tr\nabla E_{\theta_o}(x\frac{\partial \alpha(\theta_o)}{\partial \theta_k}|x_i\alpha_o)\cdot E_{\theta_o}(I_1|y_i - x_i\gamma_o, x_i\alpha_o)g(y_i - x_i\gamma_o, x_i\alpha_o) \\
&\quad + \left\{E_{\theta_o}([x\frac{\partial \gamma(\theta_o)}{\partial \theta_k}, -x\frac{\partial \alpha(\theta_o)}{\partial \theta_k}]|x_i\alpha_o) - E_{\theta_o}([x\frac{\partial \gamma(\theta_o)}{\partial \theta_k}, -x\frac{\partial \alpha(\theta_o)}{\partial \theta_k}]|I_{1i} = 1, y_i - x_i\gamma_o, x_i\alpha_o)\right\} \\
&\quad \cdot \bar{\nabla}\{E_{\theta_o}(I_1|y_i - x_i\gamma_o, x_i\alpha_o)g(y_i - x_i\gamma_o, x_i\alpha_o)\} \\
&= -tr\nabla E_{\theta_o}(x\frac{\partial \alpha(\theta_o)}{\partial \theta_k}|x_i\alpha_o)\cdot E_{\theta_o}(I_1|y_i - x_i\gamma_o, x_i\alpha_o)g(y_i - x_i\gamma_o, x_i\alpha_o) \qquad a.e.
\end{aligned}$$

Hence

$$\begin{aligned}
&\lim_{n\to\infty} E_{\theta_n}(T_{2,k}(z_j, I_j, z_i, a_n, \theta_n)|z_i) \\
&= -tr\nabla E_{\theta_o}(x|x_i\alpha_o)\frac{\partial \alpha(\theta_o)}{\partial \theta_k} \qquad a.e.
\end{aligned}$$

which is (7.24). Relation (A.2.2) implies that

$$\begin{aligned}
&E_{\theta_o}(\frac{\partial A_1(x,\theta_o)}{\partial \theta_k}|I_{1i} = 1, x_i\alpha_o) \\
&= -tr\nabla E_{\theta_o}(x\frac{\partial \alpha(\theta_o)}{\partial \theta_k}|x_i\alpha_o)\cdot E_{\theta_o}(I_l|x_i\alpha_o)p(x_i\alpha_o) \\
&\quad + (E_{\theta_o}(x|I_{1i} = 1, x_i\alpha_o) - E_{\theta_o}(x|x_i\alpha_o))\frac{\partial \alpha(\theta_o)}{\partial \theta_k}\cdot \nabla\{E_{\theta_o}(I_1|x_i\alpha_o)p(x_i\alpha_o)\} \\
&= -tr\nabla E_{\theta_o}(x\frac{\partial \alpha(\theta_o)}{\partial \theta_k}|x_i\alpha_o)\cdot E_{\theta_o}(I_l|x_i\alpha_o)p(x_i\alpha_o).
\end{aligned}$$

Hence it follows that

$$\begin{aligned}
&\lim_{n\to\infty} E_{\theta_n}(T_{4,k}(x_j, I_j, x_i, a_n, \theta_n)|x_i) \\
&= E_{\theta_o}(I_1\frac{\partial A_1(x,\theta_o)}{\partial \theta}|x_i\alpha_o)\left(\frac{1}{A_1(x_i,\theta_o)}\right)^2 p(x_i\alpha_o) \\
&= E_{\theta_o}(\frac{\partial A_1(x,\theta_o)}{\partial \theta}|I_{1i} = 1, x_i\alpha_o)\frac{1}{A_1(x_i,\theta_o)} \\
&= -tr\nabla E_{\theta_o}(x|x_i\alpha_o)\frac{\partial \alpha(\theta_o)}{\partial \theta_k} \qquad a.e.
\end{aligned}$$

which is (7.26).

3. DERIVATION OF (7.33)-(7.36):

The expressions of $W_1$ and $W_2$ are similar to the expressions $T_3$ and $T_4$ respectively. Following similar analysis for $T_3$ and $T_4$ above, (7.33) and (7.34) hold. Proposition 3 implies that

$$
\begin{aligned}
&\lim_{n\to\infty} E_{\theta_n}(W_{3,k}(x_j, I_{lj}, x_i, a_n, \theta_n)|x_i) \\
&= -p(x_i\alpha_o)\cdot tr\nabla\{E_{\theta_o}[I_l(x-x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\frac{1}{p(x_i\alpha_o)}\} - E_{\theta_o}[I_l(x-x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\frac{1}{p(x_i\alpha_o)}\nabla p(x_i\alpha_o) \\
&= -p(x_i\alpha_o)\Big\{tr\nabla E_{\theta_o}[(x-x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\frac{E_{\theta_o}(I_l|x_i\alpha_o)}{p(x_i\alpha_o)} \\
&\quad + E_{\theta_o}[(x-x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\left(\frac{1}{p(x_i\alpha_o)}\nabla E_{\theta_o}(I_l|x_i\alpha_o) - \frac{E_{\theta_o}(I_l|x_i\alpha_o)}{p^2(x_i\alpha_o)}\nabla p(x_i\alpha_o)\right)\Big\} \\
&\quad - E_{\theta_o}[(x-x_i)|x_i\alpha_o]\frac{\partial\alpha_o}{\partial\theta_k}\frac{E_{\theta_o}(I_l|x_i\alpha_o)}{p(x_i\alpha_o)}\nabla p(x_i\alpha_o) \\
&= -tr\nabla E_{\theta_o}(x|x_i\alpha_o)\frac{\partial\alpha_o}{\partial\theta_k}E_{\theta_o}(I_l|x_i\alpha_o) - [E_{\theta_o}(x|x_i\alpha_o) - x_i]\frac{\partial\alpha_o}{\partial\theta_k}\nabla E_{\theta_o}(I_l|x_i\alpha_o) \quad a.e.
\end{aligned}
$$

which is (7.35). Finally, Proposition 2 and (6.26) imply that

$$
\begin{aligned}
&\lim_{n\to\infty} E_{\theta_n}(W_{4,k}(x_j, I_{lj}, x_i, a_n, \theta_n)|x_i) \\
&= E_{\theta_o}(I_l\frac{\partial B(x,\theta_o)}{\partial\theta}|x_i\alpha_o)\left(\frac{1}{B(x_i,\theta_o)}\right)^2 p(x_i\alpha_o) \\
&= E_{\theta_o}(\frac{\partial B(x,\theta_o)}{\partial\theta}|I_{li}=1, x_i\alpha_o)\frac{E_{\theta_o}(I_l|x_i\alpha_o)}{p(x_i\alpha_o)} \\
&= \Big\{-tr\nabla E_{\theta_o}(x|x_i\alpha_o)\frac{\partial\alpha_o}{\partial\theta_k}p(x_i\alpha_o) + (E_{\theta_o}(x|I_{li}=1, x_i\alpha_o) - E_{\theta_o}(x|x_i\alpha_o))\frac{\partial\alpha_o}{\partial\theta_k}\nabla p(x_i\alpha_o)\Big\} \\
&\qquad \cdot \frac{E_{\theta_o}(I_l|x_i\alpha_o)}{p(x_i\alpha_o)} \\
&= -tr\nabla E_{\theta_o}(x|x_i\alpha_o)\frac{\partial\alpha_o}{\partial\theta_k}E_{\theta_o}(I_l|x_i\alpha_o) \quad a.e.
\end{aligned}
$$

which is (7.36)

## References


1. Bickel, P.J.(1982), "On Adaptive Estimation", *The Annals of Statistics* 10, pp. 647-671.
2. Bierens, H.J.(1985), "Kernel Estimators of Regression Functions". In T.W. Bewley (ed.), *Advances in Econometrics Fifth World Congress*, vol.1, pp.99-144. New York: Cambridge University Press.
3. Chamberlain, G.(1986), "Asymptotic Efficiency in Semi-parametric Models with Censoring", *Journal of Econometrics* 32, pp. 189-218.
4. Cosslett, S.(1983), "Distribution-free Maximum Likelihood Estimator of the Binary Choice Model", *Econometrica* 51, pp.765-782.
5. Cosslett, S.(1984), "Distribution-free Estimation of A Regression Model with Sample Selectivity", forthcoming in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, edited by W.A. Barnett, J. Powell and G. Tauchen.
6. Dubin, J.A. and D.L. McFadden (1984), "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption", *Econometrica* 52, pp.345-362.
7. Gallant, A.R. and D.W. Nychka (1987), "Semi-nonparametric Maximum Likelihood Estimation", *Econometrica* 55, pp. 363-390.
8. Gronau, R.(1974), "Wage Comparisons-A Selectivity Bias", *Journal of Political Economy* 82, pp. 1119-1143.
9. Heckman, J.(1974), "Shadow Prices, Market Wages, and Labor Supply", *Econometrica* 42, pp. 679-694.
10. Hoeffding, W.(1948), "A Class of Statistics with Asymptotically Normal Distribution", *Annals of Mathematical Statistics* 19, pp. 293-325.
11. Ichimura, H.(1987), "Estimation of Single Index Models", Ph.D. Thesis, Department of Economics, M.I.T.
12. Ichimura, H. and L.F. Lee (1988), "Semiparametric Estimation of Multiple Index Models: Single Equation Estimation", forthcoming in *Nonparametric and Semiparametric Methods in Econometrics and statistics*, edited by W.A. Barnett, J. Powell and G.Tauchen.
13. Klein, R.W. and R.H. Spady (1987), "An Efficient Semiparametric Estimator for Discrete Choice Models", Economics Research Group, Bell Communications Research, Morristown, N.J.
14. LeCam, L.(1960), "Locally Asymptotically Normal Families of Distributions", *Univ. California Publ. Statist.* 3, pp. 37-98.
15. Lee, L.F.(1982), "Some Approaches to the Correction of Selectivity Bias", *Review of Economic Studies* 49, pp. 355-372.
16. Lee, L.F.(1989),"Semiparametric Maximum Profile Likelihood Estimation of Polytomous and Sequential Choice Models", Discussion Paper no. 253, Center for Economic Research, U. of Minnesota.
17. Manski, C.F.(1984), "Adaptive Estimation of Non-Linear Regression Models", *Econometric Reviews* 3, pp.145-194.
18. McFadden, D.(1974), "Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka, ed., *Frontiers in Econometrics*, New York, Academic Press.
19. Newey, W.K.(1988), "Two Step Estimation of Sample Selection Models", manuscript, Department of Economics, Princeton University.
20. Powell, J.L.(1987), "Semiparametric Estimation of Bivariate Latent Variable Models", Discussion paper no. 8704, Social Systems Research Institute, U. of Wisconsin.
21. Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling(1986), *Numerical Recipes*, Cambridge University Press, Cambridge, Mass.

22. Robinson, P.M.(1988), "Root-N-Consistent Semiparametric Regression", *Econometrica* 56, pp. 931-954.
23. Roussas, G.(1972), *Contiguity of Probability Measures: Applications in Statistics*, Cambridge University Press.
24. Ruud, P.A.(1986), "Consistent Estimation of Limited Dependent Variable Models Despite Misspecification of Distribution", *Journal of Econometrics, Annals* 32, pp. 157-187.
25. Schick, A.(1986), "On Asymptotically Efficient Estimation in Semiparametric Models", *The Annals of Statistics* 14, pp. 1139-1151.
26. Severini, T.A. and W.H. Wong(1987), "Profile Likelihood and Semiparametric Models", unpublished manuscript, Department of Statistics, University of Chicago.
27. Silverman, B.W.(1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall Ltd., New York.
28. Stone, C.(1975), "Adaptive Maximum Likelihood Estimation of A Location Parameter", *Annals of Statistics* 3, pp. 267-284.