Semi-Parametric Estimation of a Logit Model

Gordon C.R. Kemp *
University of Essex

Current Version: January 2000

* The research reported in this paper was undertaken while I was visiting the University of British Columbia. Address for correspondence: Dr. Gordon C.R. Kemp, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, U.K. Tel: +44 (1206) 872766; Fax: +44 (1206) 872724; Email: kempgcr@essex.ac.uk.

Abstract

In this paper, I develop an estimator for a semi-parametric logit model based on a kernel-wieghted average of pairwise conditional logit terms. Then I demonstrate consistency, asymptotic normality, and consistent asymptotic covariance matrix estimation for this estimator using results for sequences of U-statistic.

JEL Classification: C14, C25.

Keywords: Semi-parametric, logit model, kernel method, U-statistics.

1 Introduction

The logit model is one of the most widely used discrete choice models in econometrics for three main reasons. First, it is easy to estimate due to the functional form of the logistic distribution. Second, it can be motivated as a model of choice between alternatives with random utilities, where the randomness comes from independent draws from an Weibull distribution (McFadden 1974, McFadden 1976). Third, it gives rise to a linear log-odds ratio which makes the interpretation of the parameters very simple. In the present paper, I develop a method for estimating a semi-parametric logit model in which the log-odds ratio is partially linear:

$$\ln\left[\frac{\Pr(Y=1|X,Z)}{1-\Pr(Y=1|X,Z)}\right] \;=\; X'\beta_0 + g_0(Z). \tag{1.1}$$

In this model, the β_0 parameters retain the same interpretation as in the conventional logit model. Furthermore, this model could be derived from a random utilities model in which each of the utilities was partially linear (with the same split between variables) plus a draw from an extreme value distribution.

The method of estimation which I develop is based on eliminating the $g_0(\cdot)$ function rather than on simultaneously estimating it jointly with β_0. The procedure is based on the following observation. For any arbitrary pair of observations (i,j) with $i \neq j$, the logarithm of the probability that $(Y_i,Y_j) = (y_i,y_j)$ given $(Y_i+Y_j) = (y_i+y_j)$ and $(X_i,X_j,Z_i,Z_j) = (x_i,x_j,z_i,z_j)$ is:

$$p^*(y_i,y_j,x_i,x_j,z_i,z_j) \;=\; \left\{\frac{\exp[(y_i-y_j)\{(x_i-x_j)'\beta_0 + g_0(z_i)-g_0(z_j)\}]}{1+\exp[(y_i-y_j)\{(x_i-x_j)'\beta_0 + g_0(z_i)-g_0(z_j)\}]}\right\}^{|y_i-y_j|}. \tag{1.2}$$

When $[g_0(z_i)-g_0(z_j)]$ is small then the right-hand-side of (1.2) is approximately the same as:

$$p^0(y_i,y_j,x_i,x_j) \;=\; \left\{\frac{\exp[(y_i-y_j)(x_i-x_j)'\beta_0]}{1+\exp[(y_i-y_j)(x_i-x_j)'\beta_0]}\right\}^{|y_i-y_j|}, \tag{1.3}$$

which is familiar as a contribution to the conditional likelihood function used to eliminate fixed-effects in the fixed-effect panel data logit model (Chamberlain 1980). The estimator I propose for this semi-parametric logit model is based on maximizing a weighted sum of the logarithms of these approximate conditional likelihood terms where the average is over all distinct pairs:

$$\begin{aligned}\hat\beta_n = \; \arg\max_{\beta\in B} \binom{2}{n}^{-1} \sum_{i<j}^{n} & w_n(z_i,z_j)\cdot|y_i-y_j|\cdot\big[(y_i-y_j)(x_i-x_j)'\beta \\ & -\ln\{1+\exp[(y_i-y_j)(x_i-x_j)'\beta]\}\big].\end{aligned} \tag{1.4}$$

Here the $\{w_n(z_i,z_j)\}$ are based on a symmetric kernel in the difference between z_i and z_j with a bandwidth parameter which tends to zero as the sample size grows, and B is the parameter space.

This estimator is designed for the situation in which Z is a continuous variable and $g_0(z)$ is continuous in z. If Z were discrete then one could modify the basic idea as follows. First, partition the observations by the value of Z. Second, for each member of the partition construct a pseudo-conditional log-likelihood for that group of observations taken jointly rather than pairwise. Third, sum up these resulting continubutions and maximize the result. Note that this procedure does not require the use of kernels and hence does not require the choice of a bandwidth. In fact, this modified estimator can be viewed simply as a fixed-effects logit estimator where the groups are indixed by the value of Z.[1]

The idea behind this pairwise comparison estimator is not entirely new. Ahn and Powell (1993) use a similar method for estimating a censored slection model: first, they use nonparametric regression to estimate the selection variable, and second, they use a weighted pairwise difference estimator where the weights depend on the difference between the estimated selection variables. More recently, Honore, Kyriazidou and Udry (1997) have proposed a number of pairwise comparison estimators of the Type-3 Tobit model, although these do not involve the use of kernel methods.

The objective function which I use takes the form of a U-statistic which means that rather than working with standard laws of large numbers and central limit theorems I need to work with laws of large numbers and central limit theorems for designed for U-statistics. The key result which I use is a lemma on mean-square convergence of the first two terms in the Hoeffding decomposition; this is given as Lemma A.1 in Appendix A and is taken from Lemma A.3 of Ahn and Powell (1993).[2]

The layout of the paper is as follows. Section 2 presents the model and the estimator, and demonstrates the existence of the estimator. Section 3 proves the weak consistency of the estimator and Section 4 establishes its root-n asymptotic normality. Section 5 then proves the consistency of an asymptotic covariance matrix estimator. Section 6 concludes the paper.

[1]Of course, if Z can take a very large set of discrete values and takes none of them with high probability then on average one would need many data points to construct this second estimator.

[2]See also Lemma 2.1 of Lee (1988) and Lemma 3.1 of Powell, Stock and Stoker (1989).

2 Model and Estimator

This section of the paper deals with the existence of the estimator of the parameters of interest. To proceed I make a number of assumptions which are given below.

Assumptions

A1. *Semiparametric Logistic Model*
$\{W_i\}_{i=1}^{\infty}$ is a sequence of independently identically distributed (iid) sequence random vectors, such that $W_i = (Y_i, X_i, Z_i)'$ where $Y_i \in \{0, 1\}$, $X_i \in \mathbb{R}^p$ and $Z_i \in \mathbb{R}^k$ for each $i \in \mathbb{N}$. In addition, there exist:

(i) a non-stochastic vector $\beta_0 \in \mathbb{R}^p$; and

(ii) a non-stochastic measurable function $g_0(\cdot) : \mathbb{R}^k \to \mathbb{R}$;

such that a version of the conditional probability that $Y = 1$ given X and Z is:

$$p(X, Z) \equiv F_0[X'\beta_0 + g_0(Z)],$$

where $F_0(\cdot)$ is the logistic function, given by $F_0(\alpha) = e^{\alpha}(1 + e^{\alpha})^{-1}$.

A2. *Bandwidth Sequence: I*
There exists a sequence of strictly positive constants $\{\gamma_n\}_{n=1}^{\infty}$ (the bandwidth sequence).

A3. *Kernel Function: I*
There exists a bounded, real-valued, measurable function $K(\cdot) : \mathbb{R}^k \to \mathbb{R}$ (the kernel function) such that $K(u) = K(-u)$ for all $u \in \mathbb{R}^k$.

A4. *Compact Parameter Space*
The parameter space B is a compact subset of $\mathbb{R}^p$ with a non-empty interior, denoted int(B) such that $\beta_0 \in$ int(B).

Assumption A1 is a formal stamement of the semi-parametric logit model which I use in the paper. It would be possible to weaken this assumption by allowing the $\{(X_i, W_i)\}$ sequence to exhibit serial dependence but this would substantially complicate the proofs. Assumptions A2 and A3 specify the basic requirements on the bandwidth sequence and the kernel function which I use in defining the estimator. Note that since $l_{ij}^*(\beta)$ is symmetric in (i, j), the asumption that the Kernel function is symmetric is in effect made without loss of generality.

Assumption A4 is a technical assumption. I would like to define the estimator $\hat{\beta}$ to be the value of β which maximizes $Q_n(\beta; \mathcal{W}_n)$ over $\mathbb{R}^p$; however, $\mathbb{R}^p$ is not compact which leads to difficulties in ensuring that the estimator exists, is unique, and is measurable. To avoid these problems I assume that the parameter space is compact with a non-empty interior which contains β_0 as stated in Assumption A4. In practice I could weaken this somewhat by using a sequence $\{B_n\}_{n=1}^{\infty}$ of nested compact parameter spaces satisfying Assumption A4 and designed so that for any given $b \in \mathbb{R}^p$ there exists $\bar{n}(b) < \infty$ such that $b \in \text{int}(B_n)$ for all $n \geq \bar{n}(b)$. The estimator would be defined for all n but asymptotically the bounds on the parameter space would vanish. However, in the remainder of the paper I will continue with the assumption of a fixed compact parameter space.

The objective function is then given by:

$$Q_n(\beta; \mathcal{W}_n) \; = \; \binom{n}{2}^{-1} \sum_{i<j}^{n} q_n(\beta; W_i, W_j) \tag{2.1}$$

where $\mathcal{W}_n = (W_i)_{i=1}^n$, and where $\sum_{i<j}^n$ denotes the sum over all (i, j) pairs such that $1 \leq i < j \leq n$, and where:

$$q_n(\beta; W_i, W_j) \; = \; q_{n,ij}(\beta) \; \equiv \; l^*(\beta; Y_i, Y_j, X_i, X_i)\gamma_n^{-k} K\left(\frac{Z_i - Z_j}{\gamma_n}\right), \tag{2.2}$$

$$l^*(\beta; Y_i, Y_j, X_i, X_j) \; = \; l^*_{ij}(\beta) \; \equiv \; (Y_i - Y_j)^2 \ln F_0[(Y_i - Y_j)(X_i - X_j)'\beta]. \tag{2.3}$$

Since Y can only take the values 0 and 1 then $(Y_i - Y_j)^2 = |Y_i - Y_j|$ so that (2.1) is the same as the objective function as that given in the Introduction.

Theorem 2.1 (Existence). *Under Assumptions A1–A4 there exists a mapping:*

$$\hat{\beta}_n(\cdot) : \times_{i=1}^n (\{0, 1\} \times \mathbb{R}^p \times \mathbb{R}^k) \to B$$

with Borel measurable components such that:

$$Q_n(\hat{\beta}_n(\mathcal{W}_n); \mathcal{W}_n) \; = \; \sup_{\beta \in B} Q_n(\beta; \mathcal{W}_n). \tag{2.4}$$

Proof. See Appendix B.[3] □

This theorem only establishes existence of the estimator; it does not indicate the best way to compute the estimator. If the kernel were non-negative everywhere then the objective function would be globally concave (from the concavity of the logarithm of the logistic cdf). If in addition the parameter space was convex one could simply pick an arbitrary starting value and use any standard derivative-based algorithm in order to reach the global maximum. However, as I will show subsequently, in order to establish asyptotic normality when $k > 3$ (or to establish consistent asymptotic covariance matrix estimation when $k > 1$) it is necessary to use higher-order

[3]Assumption A1 is not actually necessary. All that is needed from this assumption for the existence of $\hat{\beta}_n$ is that the sequence $\{W_i\}_{i=1}^{\infty}$ is a sequence of random vectors with the dimensions and event space as given by the assumption; the iid aspect is not required.

kernels which are negative over some ranges. In such situations it is not obvious how best to proceed when implementing this estimator.

3 Consistency

In this section, I demonstrate that the consistency of the estimator $\hat{\beta}_n$ of β defined in Section 2 under the assumptions made in Section 2 supplemented by a set of additional assumptions.

Assumptions

B1. *Bandwidth Sequence: II*

The sequence $\{\gamma_n\}_{n=1}^{\infty}$ specified in Assumption A2 satisfies the following additional requirements:

(i) $\lim_{n\to\infty} \gamma_n = 0$; and

(ii) $\lim_{n\to\infty}(n\gamma_n^k)^{-1} = 0$.

B2. *Kernel Function: II*

The function $K(\cdot)$ specified in Assumption A3 satisfies the following additional requirements:

(i) $\int_{\mathbb{R}^k} K(u)\, du = 1$; and

(ii) $\int_{\mathbb{R}^k} K(u)^2\, du = L_1 < \infty$.

B3. *Conditional PDF of Z Given X: I*

There exists a version $f_{Z|X}(z|x)$ of the conditional pdf of Z given X such that:

(i) there exists $L_2 < \infty$ such that $f_{Z|X}(z|x) \leq L_2$ for all $z \in \mathbb{R}^k$ and $x \in \mathbb{R}^p$;

(ii) $f_{Z|X}(z|x)$ is continuous in (z, x) for all $z \in \mathbb{R}^k$ and $x \in \mathbb{R}^p$; and

(iii) $f_{Z|X}(z|x) > 0$ for all $z \in \mathbb{R}^k$ and $x \in \mathbb{R}^p$.

B4. *Semiparametric Component: I*

The function $g_0(z)$ specified in Assumption A1 is continuous for all $z \in \mathbb{R}^k$.

B5. *Moments of X: I*

(i) There exists $L_3 < \infty$ such that $E\{\|X_i\|^2\} \leq L_3$; and

(ii) for any fixed $\xi \in \mathbb{R}^p$, such that $\xi \neq 0$, and any scalar constant c, $\Pr(\xi' X_i = c) = 0$.

All of these assumptions are fairly straightforward. Assumption B1 states that the bandwidth sequence γ_n tends to 0 but not too rapidly; in particular, the higher the dimension of Z the slower the rate of convergence must be. Thi is closely related to the standard curse of dimensionality issue in kernel-based non-parametric estimation. Assumption B2 states that the kernel fucntion should be normalized to integrate to 1 and should be square integrable. It seems unlikely that these two assumptions can be weakened.

Assumption B3 is in many respects the strongest and least justifiable of the assumptions made here: it imposes boundedness, continuity, and strict positivity everwhere of a version of the conditional pdf of Z given X. It seems highly plausible that all of its requirements can be weakened to some extent. Assumption B4 ensures that for any specified z, if z^* is close to z^* then $g(z^*)$ is close to $g(z)$; in effect, this motivates the use of the kernel method in this context as discussed in the Introduction. However, imposing this assumptions only makes sense if Z is continuous. If Z had a discrete distribution then, as discussed in the Introduction, we could construct an estimator of β which did not require the use of kernels at all. Nevertheless, even if one retains Assumption B3 it may still be possible to weaken Assumption B4 to some extent to allow for discontinuities. For example, one might require that the set of discontinuity points of $g_0(Z)$ is finite. The argument here would be that discontinuities in $g_0(Z)$ only matter for pairs (Z_i, Z_j) such that Z_i and Z_j are both close together and close to a discontinuity point. As the sample size grows while the bandwidth shrinks the combined influence of such pairs may wash out as being asymptotically irrelevant.

Assumption B5 guarantees the existence of the covariance matrix of X and also ensures that X does not with positive probability satisfy any linear restriction. This is important for ensuring global identifiction, though it may well be possible to weaken to simply requiring that the variance matrix of X is finite and non-singular.

In what follows, I will typically suppress the explicit dependence of Q_n on $\mathcal{W}_n$ and simply write $Q_n(\beta)$ in place of $Q_n(\beta; \mathcal{W}_n)$. The first step in the proof of consistency is to establish that the expectation of $Q_n(\beta)$ converges to a non-stochastic function $Q_0(\beta)$.

Lemma 3.1. *Under Assumptions A1–A4 and B1–B5:*

$$E_0[Q_n(\beta)] \;\rightarrow\; Q_0(\beta) \tag{3.1}$$

as $n \to \infty$*, where*

$$\begin{aligned} Q_0(\beta) = \; E_0 \big\{ & p(X_1, Z_2)\bar{p}(X_2, Z_2) f_{Z|X}(Z_2|X_1) \ln F_0[(X_1 - X_2)'\beta] \\ & + \; \bar{p}(X_1, Z_2) p(X_2, Z_2) f_{Z|X}(Z_2|X_1) \ln F_0[-(X_1 - X_2)'\beta] \big\} , \end{aligned} \tag{3.2}$$

where $\bar{p}(x, z) = 1 - p(x, z)$.

Proof. See Appendix B. □

Next I establish that $Q_n(\beta)$ minus its expectation converges pointwise in probability to zero.

Lemma 3.2. *Under Assumptions A1–A4 and B1–B5:*

$$Q_n(\beta) \; - \; E_0[Q_n(\beta)] \;\xrightarrow{\;p\;}\; 0, \tag{3.3}$$

as $n \to \infty$.

Proof. See Appendix B. □

Then I extend this pointwise convergence in probability to uniform convergence in probability on the compact set B.

Lemma 3.3. *Under Assumptions A1–A4 and B1–B5:*

$$\sup_{\beta \in B} \left| Q_n(\beta) \; - \; E_0[Q_n(\beta)] \right| \;\xrightarrow{\;p\;}\; 0, \tag{3.4}$$

as $n \to \infty$.

Proof. See Appendix B. □

Finally I establish that the function $Q_0(\beta)$ has a unique global maximum on B at $\beta = \beta_0$.

Lemma 3.4. *Under Assumptions A1–A4 and B1–B5:*

$$Q_0(\beta) \;\leq\; Q_0(\beta_0) \qquad \forall\, \beta \;\in\; \mathbb{R}^p \tag{3.5}$$

with equality if and only if $\beta = \beta_0$.

Proof. See Appendix B. □

These four lemmas taken jointly then establish the consistency of the estimator $\hat{\beta}_n$ as desired by the following theorem.

Theorem 3.1 (Consistency). *Under Assumptions A1–A4 and B1–B5:*

$$\hat{\beta}_n \;\xrightarrow{\;p\;}\; \beta_0,$$

as $n \to \infty$.

Proof. Theorem 2.1 and Lemmas 3.3 and 3.4 imply that the conditions of Theorem 4.1.1 of Amemiya (1985) are satisifed from which the result follows immediately. □

It may be possible to strengthen this result somewhat. First, it may be possible to drop Assumption A4, namely the compactness of the parameter space. The main role which this plays, apart from ensuring existence, lies in the proof of Lemma 3.3, i.e. the proof of the uniform convergence in probability of $Q_n(\beta)$ to $E_0[Q_n(\beta)]$ over the parameter space. If one could demonstrate that the probability that the unrestricted argmax of $Q_n(\beta)$ belongs to B tended to 1 then weak consistency of the unrestricted argmax would be easy to establish (provided that one handled issues of existence carefully). This in turn would be straightforward to establish if the objective function was almost surely globally concave, but as noted in Section 2 to ensure this would require that the kernel function was non-negative which has implications about the maximum values of k at which one can etablish asymptotic normality of the estimator and consistent asymptotic covariance matrix estimation. Second, it may be possible to establish almost sure consistency. This would require considerable modification to the method of proof used in this paper which is based on a stochastic expansion which converges in mean square (as given by Lemma A.1). Third, as noted earlier, it may be possible to prove Theorem 3.1 under weakened versions of Assumptions B3–B5.

4 Root-n Asympototic Normality

In this section of the paper I demonstrate the asymptotic normality of $\hat{\beta}_n$. As previously, I need to make some additional assumptions in order to establish the desired result.

Assumptions

C1. *Bandwidth Sequence: III*
There exists a strictly positive integer $t(k)$ such that the sequence $\{\gamma_n\}_{n=1}^{\infty}$ specified in Assumptions A2 and B1 satisfies the following additional requirement that:

$$\lim_{n\to\infty} n^{1/2}\gamma_n^{t(k)} = 0.$$

C2. *Kernel Function: III*
The function $K(\cdot)$ specified in Assumptions A3 and B2 satisfies the following additional requirements:

(i) There exists $L_4 < \infty$ such that $\int_{\mathbb{R}^k} \|u\|^{t(k)} \cdot |K(u)|\, du \leq L_4$; and
(ii) $\int_{\mathbb{R}^k} u^s K(u)\, du = 0$ for all $s = 1, \ldots, t(k) - 1$.

C3. *Conditional PDF of Z Given X: II*
The version $f_{Z|X}(z|x)$ of the conditional pdf of Z given X specified in Assumption B3 satisfies the following additional requirement: there exists $L_6 < \infty$ such that:

$$\left\| \frac{\partial^s f_{Z|X}(z|x)}{\partial z^s} \right\| \leq L_6,$$

for all $z \in \mathbb{R}^k$, $x \in \mathbb{R}^p$ and $s = 1, \ldots, t(k)$.

C4. *Semiparametric Component: II*
The function $g_0(z)$ specified in Assumptions A1 and B4 in addition satisfies the following additional requirement: there exists $L_5 < \infty$ such that:

$$\left\| \frac{\partial^s g_0(z)}{\partial z^s} \right\| \leq L_5,$$

for all $z \in \mathbb{R}^k$ and $s = 1, \ldots, t(k)$.

C5. *Moments of X: II*
There exists $L_7 < \infty$ such that $E\{\|X_i\|^4\} \leq L_7$ for all $i \in \mathbb{N}$.

Again all of these assumptions are fairly straightforward. Their main function is to ensure that the limiting normal distribution has a mean of zero as demostrated below in Lemma 4.3. In the proof of this Lemma I take a Taylor Series expansion of $E[\partial Q_n/\partial\beta]$ to order $t(k)-1$ and show that the leading terms are all equal to zero and that the remainder term is of order $o(\gamma_n^{t(k)})$. It follows that $n^{1/2}E[\partial Q_n/\partial\beta]$ is of order $o[n^{1/2}\gamma_n^{t(k)}]$ which by Assumption C1 tends to zero. The other assumptions are necessary in order to ensure that a Taylor series expansion of this order can be taken and to guarantee behaviour as described above of the terms in the expansion. If I weakened Assumption C1 so that $\lim_{n\to\infty} n\gamma_n^{2t(k)} = c_0$ for some $0 < c_0 < \infty$ then I would obtain asymptotic normality with a non-zero mean. Note that Assumption C1 (or this weaker assumption) combined with Assumption B1 implies that $t(k) > k/2$. This follows because by Assumptions B1(ii) and C1 then $\lim_{n\to\infty}\gamma_n^{2t(k)-k} = 0$ which is only consistent with $\lim_{n\to\infty}\gamma_n = 0$, from Assumptions B1(i), provided that $2t(k)-k > 0$.

A similar set of assumptions is used in the standard proof of asymptotic normality of the kernel regression estimator in order to establish a similar result. Nevertheless there are some differences between the kernel regression framework and the semi-parametric logit framework considered here. The most important points to note are as follows. First, I demostrate root-N asymptotic normality provided that $t(k) > k/2$, whereas in the kernel regression context the larger the value of the equivalent to $t(k)$ the faster is the rate of convergence of the estimator though it never gets to $N^{1/2}$. Second, in my framework asymptotic normality only holds if $t(k) > k/2$ while no such rquirement is necessary in the kernel regression framework.

Note that Assumption C2 requires that if $k \geq 4$ then $t(k) > 2$ so that $\int_{\mathbb{R}^k} uu'K(u)du = 0$ which implies that $K(\cdot)$ must be negative for some values of u in view of the requirement that $\int_{\mathbb{R}^k} K(u)du = 1$ by Assumption B2. As pointed out in Section 3 it is not then possible to ensure that the objective function will be globally concave. In contrast, if $k \leq 3$ then one can use a kernel which ensures that the objective function will be globally concave and still ensure asymptotic normality.[4]

Assumptions C3 and C4 can probably be weakened somewhat so that although derivatives to the relevant order do exist they need not be uniformly bounded by a constant but instead are bounded by suitable functions of (Z, X). However, such a weakening of these asumptions would certainly require a corresponding strengthening of the assumptions about the existence of moments of X and Z.

The proof of asympototic normality follows a fairly standard line of argument and hinges on a first-order condition expansion given by the following lemma.

[4]Yatchew (1997) has proposed an ingenious method of estimation for a partially linear regression model. In the simplest version the non-parametric component only depends on a scalar variable Z and Yatchew's procedure is as follows. First, re-order the data points by their values of Z. Second, run a regression in first differences of this re-ordered data. Yatchew also shows how to extend this procedure to a nearest neighbour differencing method when the non-parametric component depends on a vector variable. Interestingly, the maximal dimension of this vector variable is 3.

Lemma 4.1. *Under Assumptions A1–A4 and B1–B5:*

$$n^{1/2}\left[\frac{\partial Q_n}{\partial \beta}\bigg|_{\beta_0}\right] + \left[\frac{\partial^2 Q_n}{\partial \beta \partial \beta'}\bigg|_{\beta_n^*}\right][n^{1/2}(\hat{\beta}_n - \beta_0)] = o_p(1), \tag{4.1}$$

for some β_n^* belonging to the line segment joining $\hat{\beta}_n$ and β_0.[5]

Proof. See Appendix B. □

I then show that $n^{1/2}[(\partial Q_n/\partial \beta); \beta = \beta_0]$ converges to a multivariate normal with mean zero. I do this by first, showing that $n^{1/2}[\partial Q_n/\partial \beta]$ can be stochastically expanded in a fashion which will permit application of a central limit theorem (CLT).

Lemma 4.2. *Under Assumptions A1–A4 and B1–B5:*

$$n^{1/2}\left[\frac{\partial Q_n}{\partial \beta}\bigg|_{\beta_0}\right] = n^{1/2} r_n^e(\beta_0) + \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left[r_{n,i}^e(\beta_0) - r_n^e(\beta_0)\right] + o_p(1), \tag{4.2}$$

where

$$r_n^e(\beta) = E_0\left[r_{n,ij}(\beta);\right] = E_0\left[r_{n,i}^e(\beta)\right], \tag{4.3}$$

$$r_{n,i}^e(\beta) = E_0\left[r_{n,ij}(\beta)\,|W_i\right], \tag{4.4}$$

$$r_{n,ij}(\beta) = \left[\frac{\partial q_{n,ij}}{\partial \beta}\right] = \gamma_n^{-k} K\left(\frac{Z_i - Z_j}{\gamma_n}\right)\left[\frac{\partial l_{ij}^*}{\partial \beta}\right]. \tag{4.5}$$

Proof. See Appendix B. □

To ensure that $n^{1/2}[(\partial Q_n/\partial \beta); \beta = \beta_0]$ converges to a multivariate normal with mean zero I need to ensure that $n^{1/2} r_n^e(\beta_0)$ converges to zero.

Lemma 4.3. *Under Assumptions A1–A4, B1–B5 and C1–C5:*

$$n^{1/2} r_n^e(\beta_0) \to 0. \tag{4.6}$$

Proof. See Appendix B. □

Then I need to apply a CLT to the triangular array $\left\{ r_{n,i}^e(\beta_0) - r_n^e(\beta_0) \right\}$, indexed by $n = 1, 2 \ldots, \infty$ and $i = 1, 2, \ldots n$.

[5]Strictly speaking, β_n^* is different for each row in the second derivative matrix.

Lemma 4.4. *Under Assumptions A1–A4, B1–B5 and C1–C5:*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\, r^{e}_{n,i}(\beta) - r^{e}_{n}(\beta)\,\right] \;\stackrel{D}{\longrightarrow}\; N[0, A_1], \tag{4.7}$$

where A_1 is a finite symmetric positive definite matrix given by:

$$A_1 \;=\; V_0\left[E_0\left\{\left[Y_i\mu_1(W_i,W_j) - (1-Y_i)\mu_2(W_i,W_j)\right](\Delta_{ij}X) \mid W_i\right\}\right], \tag{4.8}$$

and:

$$\mu_1(W_i,W_j) \;\;\equiv\;\; F_0\left[-(\Delta_{ij}X)'\beta_0\right]\bar{p}(X_j,Z_i)f_{Z|X}(Z_i|X_j) \tag{4.9}$$

$$\mu_2(W_i,W_j) \;\;\equiv\;\; F_0\left[(\Delta_{ij}X)'\beta_0\right]p(X_j,Z_i)f_{Z|X}(Z_i|X_j). \tag{4.10}$$

Proof. See Appendix B. □

These three lemmas then imply that:

$$\eta_n \;=\; n^{1/2}\left[\left.\frac{\partial Q_n}{\partial\beta}\right|_{\beta_0}\right] \;\stackrel{D}{\longrightarrow}\; N[0, 4A_1], \tag{4.11}$$

which takes care of the first term on the right-hand-side of (B.22).

The second stage of the proof consists of establishing that $[\partial^2 Q_n/\partial\beta\partial\beta'; \beta^*_n]$ converges in probability to a non-singular matrix. I start by establishing that the expectation of $[\partial^2 Q_n/\partial\beta\partial\beta'; \beta_0]$ converges to a non-singular matrix A_2.

Lemma 4.5. *Under Assumptions A1–A4, B1–B5 and C1–C5:*

$$E_0\left[\left.\frac{\partial^2 Q_n}{\partial\beta\partial\beta'}\right|_{\beta_0}\right] \;\longrightarrow\; A_2, \tag{4.12}$$

where:

$$A_2 \;=\; E_0\left\{p(X_i,Z_j)\bar{p}(X_j,Z_j)f_{Z|X}(Z_j|X_i)(\Delta_{ij}X)(\Delta_{ij}X')\right\} \tag{4.13}$$

Proof. See Appendix B. □

Next I establish that $[\partial^2 Q_n/\partial\beta\partial\beta']$ converges pointwise in probability.

Lemma 4.6. *Under Assumptions A1–A4, B1–B5 and C1–C5:*

$$\left[\frac{\partial^2 Q_n}{\partial\beta\partial\beta'}\right] - E_0\left[\frac{\partial^2 Q_n}{\partial\beta\partial\beta'}\right] \xrightarrow{p} 0, \tag{4.14}$$

for each $\beta \in B$.

Proof. See Appendix B. □

Last I show that the difference between $[\partial^2 Q_n/\partial\beta\partial\beta'; \beta_n^*]$ and $[\partial^2 Q_n/\partial\beta\partial\beta'; \beta_0]$ converges in probability to zero.

Lemma 4.7. *Under Assumptions A1–A4, B1–B5 and C1–C5:*

$$\left[\left.\frac{\partial^2 Q_n}{\partial\beta\partial\beta'}\right|_{\beta_n^*}\right] - \left[\left.\frac{\partial^2 Q_n}{\partial\beta\partial\beta'}\right|_{\beta_0}\right] \xrightarrow{p} 0. \tag{4.15}$$

Proof. See Appendix B. □

These second three lemmas taken together then imply that $[\partial^2 Q_n/\partial\beta\partial\beta'; \beta_n^*]$ converges in probability to the non-singular matrix A_2 defined by (4.13). Combined with the asymptotic normality of $n^{1/2}[(\partial Q_n/\partial\beta); \beta = \beta_0]$ together with the result that (B.21) holds with probability tending to one and hence (B.22) holds with probability tending to one, I can then establish the following theorem.

Theorem 4.1. *Under Assumptions A1–A4, B1–B5 and C1–C5:*

$$n^{1/2}(\hat\beta_n - \beta_0) \xrightarrow{D} N[0, 4A_2^{-1}A_1A_2^{-1}], \tag{4.16}$$

where A_1 *and* A_2 *are defined in (4.8) and (4.13) respectively.*

Proof. Follows immediately from Theorems 2.1 and 3.1, and Lemmas 4.1–4.7. □

This estimator is only the simplest of a collection of possible estimators based on the elimination of approximate fixed effects. Thus one could take each triplet of data points (W_i, W_j, W_k) and compute a conditional likelihood contribution for this triplet. These contributions could then be averaged with weights that depend on the distances between Z_i, Z_j and Z_k, and would thus be generalizations of the kernel-based weights in the pairwise comparison estimator. It seems plausible that such an estimator would be asymptotically more efficient than the pairwise comparison estimator discussed in the present paper. A somewhat similar issue arises in the differencing estimator of the partially linear regression model proposed by Yatchew (1997) and discussed in footnote 4 above. Yatchew shows that asymptotic efficiency of the simple differencing estimator can be improved upon by using more complex differencing transformations; indeed, it is possible to construct a differencing based estimator whose asymptotic covariance matrix is arbitrarily close to the semi-parametric efficiency bound. In the present context a natural counterpart to a more complex difference transformation is a conditional likelihood contribution for than two data points. However, formulating an appropriate weighting scheme and analyzing the properties of such an estimator remains a topic for research.

5 Asympototic Covariance Matrix Estimation

As in the earlier sections of the paper, in order to proved the desired result, namely the consistency of a particular estimator of the asymptotic covariance matrix of $\hat{\beta}_n$ I need to strengthen the assumptions made previously.

Assumptions

D1. *Bandwidth Sequence: IV*
The bandwidth sequence $\{\gamma_n\}$ specified in Assumptions A2, B1 and C1 satisifes the additional requirement that γ_n satisfies $\lim_{n\to\infty}(n\gamma_n^{2k})^{-1} = 0$.

When combined with Assumptions A2, B1 and C1, Assumption D1 implies that $t(k) > k$ which then implies that Assumptions C2–C4 become much stronger. In particular, if $k \geq 2$ then Assumption C2 implies that the kernel function must be negative over some set and thus the objective function cannot be guaranteed to be globally concave.

So far I have not specified an asymptotic covariance matrix estimator. Theorem 4.1 demonstrates that the asymptotic covariance matrix of $\hat{\beta}_n$ is given as $\Sigma_0 = 4A_2^{-1}A_1A_2^{-1}$. This leads to an obvious strategy for estimating Σ_0, namely to construct consistent estimators $\hat{A}_{1,n}$ and $\hat{A}_{2,n}$ of A_1 and A_2 respectively and then use $\hat{\Sigma}_n = 4\hat{A}_{2,n}^{-1}\hat{A}_{1,n}\hat{A}_{2,n}^{-1}$.

Since $Q_n(\beta)$ is not a sample average of independent (or even just uncorrelated) terms, and hence neither is $[\partial Q_n(\beta)/\partial\beta]$, I cannot simply use an outer product of the gradient type estimator of A_1. Instead, I propose to use the following estimator of A_1:

$$\hat{A}_{1,n} \;=\; n^{-1}\sum_{i=1}^{n}\bar{r}_{n,i}(\hat{\beta}_n)\bar{r}_{n,i}(\hat{\beta}_n)', \tag{5.1}$$

where:

$$\bar{r}_{n,i}(\hat{\beta}_n) \;=\; (n-1)^{-1}\sum_{j=1}^{n}(1-\delta_{ij})r_{n,ij}(\hat{\beta}_n), \tag{5.2}$$

in which δ_{ij} equals 1 if $i = j$ and 0 otherwise. The motivation behind this is that A_1 is the limiting covariance matrix of $r_{n,i}^e(\beta_0) = E_0[r_{n,ij}(\beta_0)|W_i]$ by Lemmma 4.4. Heuristically, I am then estimating the individual $r_{n,i}^e(\beta_0)$ by their sample equivalents $\bar{r}_{n,i}(\hat{\beta}_n)$ and then using the outer product of the $\bar{r}_{n,i}(\hat{\beta}_n)$ to estimate A_1. Note that:

$$n^{-1}\sum_{i=1}^{n}\bar{r}_{n,i}(\hat{\beta}_n) \;=\; \binom{n}{2}^{-1}\sum_{i<j}^{n}r_{n,ij}(\hat{\beta}_n) \;=\; \left[\left.\frac{\partial Q_n}{\partial\beta\partial\beta'}\right|_{\hat{\beta}_n}\right], \tag{5.3}$$

which will equal zero with probability tending to one as shown in the proof of Lemma 4.1. Hence there is no need asymptotically to recenter the $\bar{r}_{n,i}(\hat{\beta}_n)$ when constructing their sample covariance matrix.

A little manipulation reveals that:

$$\hat{A}_{1,n} \;=\; \hat{A}_{11,n} + \hat{A}_{12,n} \tag{5.4}$$

where:

$$\hat{A}_{11,n} \;=\; n^{-1}(n-1)^{-2}\sum_{i\neq j}^{n} r_{n,ij}(\hat{\beta}_n) r_{n,ij}(\hat{\beta}_n)' \tag{5.5}$$

$$\hat{A}_{12,n} \;=\; n^{-1}(n-1)^{-2}\sum_{i\neq j\neq l}^{n} r_{n,ij}(\hat{\beta}_n) r_{n,il}(\hat{\beta}_n)'. \tag{5.6}$$

I then handle these two terms separately. First, I show that $\hat{A}_{11,n}$ converges in probability to zero.

Lemma 5.1. *Under Assumptions A1–A4, B1–B5, C1–C5, and D1 :*

$$\hat{A}_{11,n} \;\xrightarrow{p}\; 0. \tag{5.7}$$

Proof. See Appendix B. □

Second, I show that $\hat{A}_{12,n}$ converges in probability to A_1. As in previous proofs of consistency I do this in three steps.

Lemma 5.2. *Under Assumptions A1–A4, B1–B5, C1–C5, and D1 :*

$$E_0\left[n^{-1}(n-1)^{-2}\sum_{i\neq j\neq k}^{n} r_{n,ij}(\beta_0) r_{n,ik}(\beta_0)'\right] \;=\; A_{1,n} \;\longrightarrow\; A_1. \tag{5.8}$$

Proof. See Appendix B. □

Lemma 5.3. *Under Assumptions A1–A4, B1–B5, C1–C5, and D1 :*

$$\left[n^{-1}(n-1)^{-2}\sum_{i\neq j\neq k}^{n} r_{n,ij}(\beta_0) r_{n,ik}(\beta_0)'\right] - A_{1,n} \;\xrightarrow{p}\; 0. \tag{5.9}$$

Proof. See Appendix B. □

The main role which Assumption D1 plays is to ensure the validity of Lemma 5.3. The reason why such a strengthening of the earlier assumptions is needed is that to prove Lemma 5.3 I proceed by proving mean-suare convergence. The squares of terms of the form $[\xi_1' r_{n,ij}(\beta_0)][r_{n,ik}(\beta_0)'\xi_2]$ involve multiplicative factors of γ_n^{-4k}. Transforming the variables in a suitable fashion eliminates a factor of γ_n^{-2k} leaving multiplicative factors of γ_n^{-2k}. Lemma A.1 requires that the expectations of these squares are of order $o(n)$ in order to ensure mean-square convergence. Thus I require that $\gamma_n^{-2k} = o(n)$ which is equivalent to Assumption D1.

Lemma 5.4. *Under Assumptions A1–A4, B1–B5, C1–C5, and D1 :*

$$\hat{A}_{12,n} - \left[n^{-1}(n-1)^{-2} \sum_{i \neq j \neq k}^{n} r_{n,ij}(\beta_0) r_{n,ik}(\beta_0)' \right] \xrightarrow{p} 0. \tag{5.10}$$

Proof. See Appendix B. □

The obvious estimator of A_2 is simply:

$$\hat{A}_{2,n} = \left[\left. \frac{\partial^2 Q_n}{\partial\beta\partial\beta'} \right|_{\hat{\beta}_n} \right]. \tag{5.11}$$

This is easily shown to be consistent.

Lemma 5.5. *Under Assumptions A1–A4, B1–B5, C1–C5, and D1 :*

$$\hat{A}_{2,n} \xrightarrow{p} A_2. \tag{5.12}$$

Proof. See Appendix B. □

The consistency of $4\hat{A}_{2,n}^{-1}\hat{A}_{1,n}\hat{A}_{2,n}^{-1}$ as an estimator of the asymptotic covariance matrix of $\hat{\beta}_n$ then follows straightforwardly.

Theorem 5.1. *Under Assumptions A1–A4, B1–B5, C1–C5, and D1 :*

$$4\hat{A}_{2,n}^{-1}\hat{A}_{1,n}\hat{A}_{2,n}^{-1} \xrightarrow{p} 4A_2^{-1}A_1A^{-2}. \tag{5.13}$$

Proof. This follows from Lemmas 5.1–5.5. □

This is not the only estimator of Σ_0 which one can construct in the present context: clearly I could drop $\hat{A}_{11,n}$ and simply use $\hat{A}_{12,n}$ in place of $\hat{A}_{1,n}$. However, a drawback with this estimator of A_1, and hence of the resulting estimator of Σ_0, is thatthere is no obvious reason why this latter estimator would be non-negative definite whereas the estimator examined earlier will be non-negative definite by construction. Another point to note is that the scaling factor of n used in the definition of $\hat{A}_{1,n}$ is somewhat arbitrary: any positive scaling factor $h(n)$ such that $h(n)/n$ would give the same probability limit. In particular, a scaling factor of $(n-1)$ might be more appropriate.[6]

[6] This is of course a standard problem in asymptotic covariance matrix estimation.

6 Conclusions

In this paper I have formulated a semi-parametric logit model and developed an estimator of the parameters of interest based on maximizing a kernel weighted average of pairwise conditional log-likelihood contributions. I have then demonstrated the consistency and root-n asymptotic normality of this estimator together with the consistency of a particualr asymptotic covariance matrix estimator.

There are several possible directions for future research. First, as noted in several places in the paper it may well be possible to derive the same results under somewhat weaker conditions. Second, it seems likely that asymptotically more efficient estimators can be constructed by using an objective function which is a weighted average of triplet comparison or even higher order comparison conditional likelihood contributions. Third, for low values of k, the number of variables determining the non-parametric component it may be possible to apply a variant of Yatchew's differencing estimator for the partially linear regression model. The idea here (for the case where $k = 1$) would be to order the data by the values of z and then construct an objective functi0on which was just the sum of the conditional log-likelihood contributions from the observations which are adjacent in the re-ordered data set. Fourth, the method of kernel weighting applied in this paper uses a pre-specified non-random bandwidth sequence. It would clearly be desirable to be able to use a data-dependent bandwidth sequence, chosen perhaps by some type of cross-validation scheme though it is not obvious what how to formulate such a scheme in the present context.

If Z is a scalar continuous random variable then we can adapt the approach of "Yatchew (1997)" (Economics Letters) & "Yatchew (1998)" (Journal of Economic Literature) to construct a simpler though related estimator as follows. First, order the data by increasing value of Z. The costruct the object function using only adjacent pairs of observations in this ordering with no weights. Establishing the asymptotic properties of this estimator reamins a topic for future research.

Need to add more here

A Supplementary Lemmas

In this Appendix, I state a supplementary lemma on sequences of U-statistics which will be used extensively in the proofs of results of the paper.

Lemma A.1. [from Lemma A.3 of Ahn and Powell (1993); see also Lemma 2.1 of Lee (1988) and Lemma 3.1 of Powell et al. (1989)]. *Let $\{\nu_i\}_{i=1}^{\infty}$ be a sequence of iid random vectors, and consider an m-th order vector U-statistic of the form:*

$$U_n \;=\; \binom{n}{m}^{-1} \sum_c a_n(\nu_{i_1}, \dots \nu_{i_m}) \tag{A.1}$$

where the sum is over the $\binom{n}{m}$ combinations of m distinct elements $\{i_1, \dots, i_m\}$ from the set $\{1, \dots, n\}$ and, without loss of generality, the sequence of functions $a_n(\nu_{i_1}, \dots \nu_{i_m})$ are all taken to be symmetric in their m arguments. Also define the 'projection':

$$U_n^* \;=\; \theta_n + \frac{m}{n} \sum_{i=1}^{n} \psi_n(\nu_{i_1}), \tag{A.2}$$

as the sum of the first two terms in the Hoeffding decomposition of U_n, where:

$$\psi_n(\nu_{i_1}) \;\equiv\; E[\{a_n(\nu_i, \dots \nu_{i_m}) - \theta_n\}|\nu_i], \qquad \theta_n \;\equiv\; E[a_n(\nu_i, \dots \nu_{i_m})]. \tag{A.3}$$

With these definitions, suppose that the sequence of functions $\{a_n(\cdot)\}$ satisifes:

$$E[\, \|a_n(\nu_i, \dots \nu_{i_m})\|^2 \,] \;=\; o(n), \tag{A.4}$$

then:

(i) $U_n = \theta_n + o_p(1)$,

(ii) $U_n = U_n^* + o_p(n^{-1/2})$.

Note that there appears to be a misprint in Ahn and Powell (1993) whose statement of this lemma replaces $o(n)$ by $O(n)$ in their equivalent to (A.4).

B Proofs of Main Results

Proof of Theorem 2.1

It is clear that $Q_n(\beta; \mathcal{W}_n)$ is continuous in $\beta \in \mathbb{R}^p$ given $\mathcal{W}_n$, and measuable with respect to $\mathcal{W}_n$ given β. The desired result then follows by Lemma xxx of Jennrich (1969). $\square$

Proof of Lemma 3.1

In this and the subsequent proofs it is convenient to define the following notation. First, the operator Δ_{ij} applied to a variable such as X is defined by $\Delta_{ij} X \equiv (X_i - X_j)$, and, second, $\Gamma_3 \equiv \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^p$ and $\Gamma_4 \equiv \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^p \times \mathbb{R}^p$. Since the $\{W_i\}$ are iid by Assumption A1 it follows that:

$$E_0[Q_n(\beta)] \;=\; E_0[q_{n,ij}(\beta)] \;=\; E_0[E_0\{q_{n,ij}(\beta)|(X_i, X_j, Z_i, Z_j)\}]. \tag{B.1}$$

But:

$$E_0[q_{n,ij}(\beta)|(X_i, X_j, Z_i, Z_j)] \;=\; \gamma_n^{-k} K\left(\frac{\Delta_{ij} Z}{\gamma_n}\right) E_0\left\{l_{ij}^*(\beta)|(X_i, X_j, Z_i, Z_j)\right\}, \tag{B.2}$$

while:

$$\begin{aligned} E_0[l_{ij}^*(\beta)|(X_i, X_j, Z_i, Z_j)] = \; & p(X_i, Z_i)\bar{p}(X_j, Z_j) \ln F_0[(\Delta_{ij} X)'\beta] \\ & + \; \bar{p}(X_i, Z_i) p(X_j, Z_j) \ln F_0[-(\Delta_{ij} X)'\beta], \end{aligned} \tag{B.3}$$

from which it follows that:

$$\begin{aligned} & E_0[Q_n(\beta)] \\ & \quad = \; E_0\left\{\gamma_n^{-k} K\left(\frac{Z_1 - Z_2}{\gamma_n}\right) p(X_1, Z_1)\bar{p}(X_2, Z_2) \ln F_0[(X_1 - X_2)'\beta]\right\} \\ & \qquad\quad + E_0\left\{\gamma_n^{-k} K\left(\frac{Z_1 - Z_2}{\gamma_n}\right) \bar{p}(X_1, Z_1) p(X_2, Z_2) \ln F_0[-(X_1 - X_2)'\beta]\right\} \\ & \quad = \; E_0^*\left\{p(X_1, Z_2 + \gamma_n U)\bar{p}(X_2, Z_2) f_{Z|X}(Z_2 + \gamma_n U|X_1) \ln F_0[(X_1 - X_2)'\beta]\right\} \\ & \qquad + E_0^*\left\{\bar{p}(X_1, Z_2 + \gamma_n U) p(X_2, Z_2) f_{Z|X}(Z_2 + \gamma_n U|X_1) \ln F_0[-(X_1 - X_2)'\beta]\right\}, \end{aligned} \tag{B.4}$$

by transformation of variables from (Z_1, Z_2) to (U, Z_2) with $U = (Z_1 - Z_2)/\gamma_n$, and where:

$$E_0^*[h(X_1, X_2, Z_2, U)] \;\equiv\; \int_{\Gamma_4} h(x_1, x_2, z_2, u) K(u) f_X(x_1) f_{Z,X}(z_2, x_2)\, du\, dz_2\, dx_1 dx_2.$$

It is easy to demonstrate that $|\ln F_0(\alpha)| \leq |\alpha| + \ln 2$. Furthermore, since $0 \leq p(x,z) \leq 1$ by Assumption A1 and $f_{Z|X}(z|x) \leq L_2$ for all $(x,z) \in \mathbb{R}^p \times \mathbb{R}^k$ by Assumption B3 it follows that:

$$\begin{aligned} & E_0^* \left\{ \left\| p(X_1, Z_2 + \gamma_n U)\bar{p}(X_2, Z_2) f_{Z|X}(Z_2 + \gamma_n U|X_1) \ln F_0[(X_1 - X_2)'\beta] \right\| \right\} \\ & \qquad \leq \; L_2 E_0^* \left\{ \|X_1 - X_2\| \cdot \|\beta\| + \ln 2 \right\} \; < \; \infty. \end{aligned} \tag{B.5}$$

Furthermore, it is clear that:

$$\begin{aligned} & \lim_{n\to\infty} p(x_1, z_2 + \gamma_n u)\bar{p}(x_2, z_2) f_{Z|X}(z_2 + \gamma_n u|x_1) \ln F_0[(x_1 - x_2)'\beta] \\ & \quad = \; p(x_1, z_2)\bar{p}(x_2, z_2) f_{Z|X}(z_2|x_1) \ln F_0[(x_1 - x_2)'\beta], \end{aligned} \tag{B.6}$$

for all $(u, z_2, x_1, x_2) \in \Gamma_4$, in view of the continuity of $g_0(z)$ and $f_{Z|X}(z|x)$ from Assumptions B3(ii) and B4. Hence it follows by the dominated convergence theorem that:

$$\begin{aligned} & \lim_{n\to\infty} E_0^* \left\{ p(X_1, Z_2 + \gamma_n U)\bar{p}(X_2, Z_2) f_{Z|X}(Z_2 + \gamma_n U|X_1) \ln F_0[(X_1 - X_2)'\beta] \right\} \\ & \quad = \; E_0^* \left\{ p(X_1, Z_2)\bar{p}(X_2, Z_2) f_{Z|X}(Z_2|X_1) \ln F_0[(X_1 - X_2)'\beta] \right\} \\ & \quad = \; E_0 \left\{ p(X_1, Z_2)\bar{p}(X_2, Z_2) f_{Z|X}(Z_2|X_1) \ln F_0[(X_1 - X_2)'\beta] \right\}. \end{aligned} \tag{B.7}$$

By the same logic it follows that:

$$\begin{aligned} & \lim_{n\to\infty} E_0^* \left\{ \bar{p}(X_1, Z_2 + \gamma_n U) p(X_2, Z_2) f_{Z|X}(Z_2 + \gamma_n U|X_1) \ln F_0[-(X_1 - X_2)'\beta] \right\} \\ & \quad = \; E_0 \left\{ \bar{p}(X_1, Z_2) p(X_2, Z_2) f_{Z|X}(Z_2|X_1) \ln F_0[-(X_1 - X_2)'\beta] \right\}. \end{aligned} \tag{B.8}$$

Combining (B.7) and (B.8) then gives:

$$\begin{aligned} & \lim_{n\to\infty} E_0[Q_n(\beta)] \; = \; Q_0(\beta) \\ & \quad \equiv \; E_0 \left\{ p(X_1, Z_2)\bar{p}(X_2, Z_2) f_{Z|X}(Z_2|X_1) \ln F_0[(X_1 - X_2)'\beta] \right. \\ & \qquad + \; \left. \bar{p}(X_1, Z_2) p(X_2, Z_2) f_{Z|X}(Z_2|X_1) \ln F_0[-(X_1 - X_2)'\beta] \right\}, \end{aligned} \tag{B.9}$$

which establishes the desired result. □

Proof of Lemma 3.2

Now consider:

$$\begin{aligned} q_{n,ij}(\beta)^2 = & \; \gamma_n^{-2k} K \left(\frac{\Delta_{ij} Z}{\gamma_n} \right)^2 l_{ij}(\beta)^2 \\ \leq & \; \gamma_n^{-2k} K \left(\frac{\Delta_{ij} Z}{\gamma_n} \right)^2 \left\{ \|X_1 - X_2\| \cdot \|\beta\| + \ln 2 \right\}^2 \\ \leq & \; 2\gamma_n^{-2k} K \left(\frac{\Delta_{ij} Z}{\gamma_n} \right)^2 \left\{ \|X_1 - X_2\|^2 \cdot \|\beta\|^2 + [\ln 2]^2 \right\}. \end{aligned} \tag{B.10}$$

It follows that:

$$
\begin{aligned}
E_0\left[q_{n,ij}(\beta)^2\right] &\leq\; 2\gamma_n^{-k}E_0^*\left\{K(u)f_{Z|X}(Z_2|X_1)\left(\|X_1-X_2\|^2\cdot\|\beta\|^2+[\ln 2]^2\right)\right\}\\
&=\; 2\gamma_n^{-k}\left\{\int_{\mathbb{R}^k}k(u)^2du\right\}E_0\left\{f_{Z|X}(Z_2|X_1)\left(\|X_1-X_2\|^2\cdot\|\beta\|^2+[\ln 2]^2\right)\right\}\\
&=\; O(\gamma_n^{-k}) \;=\; o(n),
\end{aligned}
\tag{B.11}
$$

since $n\gamma_n^k \rightarrow \infty$ by Assumption B1(ii). But then it follows by Lemma A.1 that $Q_n(\beta) = E_0[Q_n(\beta)] + o_p(1)$ which establishes the desired result. □

Proof of Lemma 3.3

To prove uniform convergence in probability I first establish stochastic equicontinuity. Observe that for any $\beta^*, \beta^{**} \in B$:

$$
|Q_n(\beta^*) - Q_n(\beta^{**})| \;\leq\; \binom{n}{2}^{-1}\sum_{i<j}^{n}\left|q_{n,ij}(\beta^*) - q_{n,ij}(\beta^{**})\right|. \tag{B.12}
$$

(2.2) and (2.3) then imply that:

$$
\begin{aligned}
|q_{n,ij}(\beta^*) - q_{n,ij}(\beta^{**})| \;=\; \gamma_n^{-k}\left|K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)\right|&\\
\cdot\left|\ln F_0[(\Delta_{ij}Y)(\Delta_{ij}X)'\beta^*] - \ln F_0[(\Delta_{ij}Y)(\Delta_{ij}X)'\beta^{**}]\right|&.
\end{aligned}
\tag{B.13}
$$

But $[d\ln F_0(\alpha)/d\alpha] = [1 - F_0(\alpha)] = F_0(-\alpha)$, $|F_0(-\alpha)| \leq 1$ and $|\Delta_{ij}Y| \leq 1$ so that by the mean value theorem:

$$
|\ln F_0[(\Delta_{ij}Y)(\Delta_{ij}X)'\beta^*] - \ln F_0[(\Delta_{ij}Y)(\Delta_{ij}X)'\beta^{**}]| \;\leq\; |(\Delta_{ij}X)'(\beta^* - \beta^{**})|. \tag{B.14}
$$

Now $|(\Delta_{ij}X)'(\beta^* - \beta^{**})| \leq \|X_i - X_j\| \cdot \|\beta^* - \beta^{**}\|$ so that:

$$
\begin{aligned}
|Q_n(\beta^*) - Q_n(\beta^{**})| \;&\leq\; \binom{n}{2}^{-1}\sum_{i<j}^{n}\gamma_n^{-k}\left|K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)\right|\cdot\|X_i - X_j\|\cdot\|\beta^* - \beta^{**}\|\\
&=\; H_{1,n}\cdot\|\beta^* - \beta^{**}\|.
\end{aligned}
\tag{B.15}
$$

Then by the same arguments as used in the proof of Lemma 3.1 to establish (B.5) and (B.9) it follows that:

$$
E_0\left[\gamma_n^{-k}\left|K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)\right|\cdot\|X_i - X_j\|\right] \;\longrightarrow\; E_0\left\{\|X_1 - X_2\|\cdot f_{Z|X}(Z_2|X_1)\right\}, \tag{B.16}
$$

and hence that:

$$E_0[H_{1,n}] \longrightarrow\; E_0\left\{\|X_1 - X_2\| \cdot f_{Z|X}(Z_2|X_1)\right\} \;\leq\; 2L_2 L_3^{1/2} \;<\; \infty, \tag{B.17}$$

by Assumptions B3(i) and B5(i). But then $H_{1,n} = O_p(1)$ by the Markov inequality since $H_{1,n} \geq 0$. Thus $Q_n(\beta)$ satifies the conditions of Theorem 21.20 of Davidson (1994) and hence is stochastically equicontinuous on B. But it is clear from (B.9) that $Q_0(\beta)$ is a continuous function of β and hence is uniformly continuous in $\beta \in B$ since B is compact. It therefore follows that the sequence $\{Q_n(\beta) - Q_0(\beta)\}_{n=1}^{\infty}$ is stochastically equicontinuous on B which combined with Lemmas 3.1 and 3.2 implies that:

$$\sup_{\beta \in B} |Q_n(\beta) - Q_0(\beta)| \;\xrightarrow{\;p\;}\; 0, \tag{B.18}$$

by Theorem 21.9 of Davidson (1994). This establishes the desired result. □

Proof of Lemma 3.4

It is easy to show that for each $(z_2, x_1, x_2) \in \Gamma_3$:

$$\begin{aligned} & p(x_1,z_2)\bar{p}(x_2,z_2)\ln F_0[(x_1-x_2)'\beta] + \bar{p}(x_1,z_2)p(x_2,z_2)\ln F_0[-(x_1-x_2)'\beta] \\ & \quad = \; [p(x_1,z_2)\bar{p}(x_2,z_2) + \bar{p}(x_1,z_2)p(x_2,z_2)]^{-1} \\ & \qquad \times \{F_0[(x_1-x_2)'\beta_0]\ln F_0[(x_1-x_2)'\beta] + F[-(x_1-x_2)'\beta_0]\ln F_0[-(x_1-x_2)'\beta]\}\,. \end{aligned} \tag{B.19}$$

In addition, it is easy to show that:

$$F_0(\alpha_0)\ln F_0(\alpha) + F_0(-\alpha_0)\ln F_0(-\alpha), \tag{B.20}$$

treated as a function of α given α_0, achieves a strict maximum at $\alpha = \alpha_0$ which combined with (B.9), (B.19) and Asumption B5(ii) implies that $Q_0(\beta)$ achieves a strict maximum at $\beta = \beta_0$, thus establishing the desired result. □

Proof of Lemma 4.1

First, observe that the objective function $Q_n(\beta)$ is almost surely continuously differentiable to an arbitrary order with respect to β by inspection of (2.1)–(2.3). This combined with the consistency of $\hat{\beta}_n$ from Theorem 3.1 and the property that $\beta_0 \in \text{int}(B)$ from Assumption A4 implies that $\hat{\beta}_n$ will satisfy the following first-order condition with probability tending to unity:

$$\Pr\left\{ n^{1/2}\left[\left.\frac{\partial Q_n}{\partial \beta}\right|_{\hat{\beta}_n}\right] \;=\; 0\right\} \;\longrightarrow\; 1, \tag{B.21}$$

from which it follows immediately that:

$$n^{1/2}\left[\frac{\partial Q_n}{\partial\beta}\bigg|_{\hat{\beta}_n}\right] \;=\; o_p(1). \tag{B.22}$$

Taking a first-order Taylor series expansion of $n^{1/2}[(\partial Q_n/\partial\beta);\beta = \hat{\beta}_n]$ around $\beta = \beta_0$ and substituting this into (B.22) gives the desired result. Note that the value of β_n^* is different for each row of the second-derivative matrix. □

Proof of Lemma 4.2

Observe that $[\partial Q_n/\partial\beta]$ is itself a U-statistic given by:

$$\left[\frac{\partial Q_n}{\partial\beta}\right] \;=\; \binom{n}{2}^{-1}\sum_{i<j}^{n} r_{n,ij}(\beta), \tag{B.23}$$

where, $r_{n,ij}(\beta)$ is defined in (4.5). This implies that:

$$\|r_{n,ij}(\beta)\|^2 \;\le\; \gamma_n^{-2k}K\left(\frac{Z_i - Z_j}{\gamma_n}\right)^2\cdot\left\|\frac{\partial l_{ij}^*}{\partial\beta}\right\|^2. \tag{B.24}$$

But differentiating $l_{ij}^*(\beta)$ with respect to β gives:

$$\left[\frac{\partial l_{ij}^*}{\partial\beta}\right] \;=\; F_0\left[-(Y_i - Y_j)(X_i - X_j)'\beta\right](Y_i - Y_j)(X_i - X_j), \tag{B.25}$$

using $[\partial\ln F_0(\alpha)/\partial\alpha] = F_0(-\alpha)$, and $(Y_i - Y_j)^3 = (Y_i - Y_j)$. Hence:

$$\left\|\frac{\partial l_{ij}^*}{\partial\beta}\right\|^2 \;\le\; \|X_i - X_j\|^2 \;\le\; \|X_i\|^2 + \|X_j\|^2, \tag{B.26}$$

since $F_0(-\alpha)^2 \le 1$ for all α and $(Y_i - Y_j)^2 \le 1$, so that:

$$\|r_{n,ij}(\beta)\|^2 \;\le\; \gamma_n^{-2k}K\left(\frac{Z_i - Z_j}{\gamma_n}\right)^2\cdot\left(\|X_i\|^2 + \|X_j\|^2\right). \tag{B.27}$$

Since $K(\cdot)$ is symmetric from Assumption A2(i) it follows that:

$$E_0\left(\gamma_n^k\cdot\|r_{n,ij}(\beta)\|^2\right) \;\le\; 2E_0\left[\gamma_n^{-k}K\left(\frac{Z_i - Z_j}{\gamma_n}\right)^2\cdot\|X_i\|^2\right]. \tag{B.28}$$

But now observe that:

$$
\begin{aligned}
& E_0\left[\gamma_n^{-k} K\left(\frac{Z_i-Z_j}{\gamma_n}\right)^2 \cdot \|X_i\|^2\right] \\
& \quad = \int_{\Gamma_3} \gamma_n^{-k} K\left(\frac{z_1-z_2}{\gamma_n}\right)^2 \cdot \|x_1\|^2 \cdot f_{Z,X}(z_1,x_1) f_Z(z_2) dz_1 dz_2 dx_1 \\
& \quad = \int_{\Gamma_3} K(u)^2 \cdot \|x_1\|^2 \cdot f_{Z|X}(z_2+\gamma_n u|x_1) f_X(x_1) f_Z(z_2) dz_1 dz_2 dx_1, \qquad \text{(B.29)}
\end{aligned}
$$

where we have transformed from (z_1, x_1, z_2) to (u_1, x_1, z_2) with $u = (z_1 - z_2)/\gamma_n$. But $f_{Z|X}(z_2 + \gamma_n u|x_1) \le L_2 < \infty$ by Assumption B3(ii) and $\int_{\mathbb{R}^k} K(u)^2\, du = L_1 < \infty$ by Assumption B2(iii) so that:

$$
\begin{aligned}
& E_0\left[\gamma_n^{-k} K\left(\frac{Z_i-Z_j}{\gamma_n}\right)^2 \cdot \|X_i\|^2\right] \\
& \quad \le L_2 \cdot \int_{\Gamma_3} K(u)^2 \cdot \|x_1\|^2 \cdot f_X(x_1) f_Z(z_2) dz_2 du dx_1 \\
& \quad \le L_2 \cdot \int_{\mathbb{R}^k} f_Z(z_2) dz_2 \cdot \int_{\mathbb{R}^k} K(u)^2 du \cdot \int_{\mathbb{R}^p} \|x_1\|^2 f_X(x_1) dx_1 \;=\; L_1 L_2 L_3 \;<\; \infty. \qquad \text{(B.30)}
\end{aligned}
$$

Thus $E_0\left(\|r_{n,ij}(\beta)\|^2\right) = O(\gamma_n^{-k}) = o(n)$ and hence by Lemma A.1:

$$
\left[\frac{\partial Q_n}{\partial \beta}\right] \;=\; E_0\left[\, r_{n,ij}(\beta)\,\right] + \frac{2}{n}\sum_{i=1}^{n} \left\{ E_0\left[\, r_{n,ij}(\beta)\,|W_i\,\right] - E_0\left[\, r_{n,ij}(\beta)\,\right]\right\} + o_p(n^{-1/2}), \qquad \text{(B.31)}
$$

from which the desired result follows immediately. □

Proof of Lemma 4.3

From (4.5) and (B.25) it is clear that:

$$
r_{n,ij}(\beta) \;=\; \gamma_n^{-k} K\left(\frac{\Delta_{ij} Z}{\gamma_n}\right) F_0\left[-(\Delta_{ij}Y)(\Delta_{ij}X)'\beta\right](\Delta_{ij}Y)(\Delta_{ij}X). \qquad \text{(B.32)}
$$

At this point it is convenient to define:

$$
\begin{aligned}
q_{n,ij}^0(\beta) &= \gamma_n^{-k} K\left(\frac{\Delta_{ij} Z}{\gamma_n}\right) l_{ij}^0(\beta), \qquad &\text{(B.33)} \\
l_{ij}^0(\beta) &= \ln F_0\left[(\Delta_{ij}Y)\left\{(\Delta_{ij}X)'\beta + \Delta_{ij} g_0(Z)\right\}\right], \qquad &\text{(B.34)}
\end{aligned}
$$

where $\Delta_{ij} g_0(Z) = g_0(Z_i) - g_0(Z_j)$. Differentiating $q^0_{n,ij}(\beta)$ with respect to β gives:

$$r^0_{n,ij}(\beta) \;=\; \left[\frac{\partial q^0_{n,ij}}{\partial \beta}\right] \;=\; \gamma_n^{-k} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)\left[\frac{\partial l^0_{ij}}{\partial \beta}\right], \tag{B.35}$$

where:

$$\left[\frac{\partial l^0_{ij}}{\partial \beta}\right] \;=\; F_0\left[-(\Delta_{ij}Y)\left\{(\Delta_{ij}X)'\beta + \Delta_{ij} g_0(Z)\right\}\right](\Delta_{ij}Y)(\Delta_{ij}X). \tag{B.36}$$

Now it is straightforward to show that:

$$E_0\left[\left(\left.\frac{\partial l^0_{ij}}{\partial \beta}\right|_{\beta_0}\right)|(X_i, X_j, Z_i, Z_j)\right] \;=\; 0, \tag{B.37}$$

which combined with (B.33) implies that:

$$E_0\left[r^0_{n,ij}(\beta_0)\right] \;=\; E_0\left[\left.\frac{\partial q^0_{n,ij}}{\partial \beta}\right|_{\beta_0}\right] \;=\; 0, \tag{B.38}$$

and hence that:

$$\begin{aligned} r^e_n(\beta_0) =&\; E_0\left[r_{n,ij}(\beta_0) - r^0_{n,ij}(\beta_0)\right] \\ =&\; E_0\left\{\gamma_n^{-k} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)\left[\left(\left.\frac{\partial l^*_{ij}}{\partial \beta}\right|_{\beta_0}\right) - \left(\left.\frac{\partial l^0_{ij}}{\partial \beta}\right|_{\beta_0}\right)\right]\right\}, \end{aligned} \tag{B.39}$$

where from (B.25) and (B.36):

$$\begin{aligned} \left[\left(\frac{\partial l^*_{ij}}{\partial \beta}\right) - \left(\frac{\partial l^0_{ij}}{\partial \beta}\right)\right] =&\; \{F_0[(\Delta_{ij}Y)\{(\Delta_{ij}X)'\beta + \Delta_{ij}g_0(Z)\}] - F_0[(\Delta_{ij}Y)(\Delta_{ij}X)'\beta]\} \\ &\times (\Delta_{ij}Y)(\Delta_{ij}X). \end{aligned} \tag{B.40}$$

But then the expectation of (B.40) conditional on (X_i, X_j, Z_i, Z_j) is;

$$\begin{aligned} &E_0\left\{\left[\left(\left.\frac{\partial l^*_{ij}}{\partial \beta}\right|_{\beta_0}\right) - \left(\left.\frac{\partial l^0_{ij}}{\partial \beta}\right|_{\beta_0}\right)\right]|(X_i, X_j, Z_i, Z_j)\right\} \\ &\quad= \{F_0[(\Delta_{ij}X)'\beta_0 + \Delta_{ij}g_0(Z)] - F_0[(\Delta_{ij}X)'\beta_0]\}(\Delta_{ij}X)p(X_i, Z_i)\bar{p}(X_j, Z_j) \\ &\qquad+ \{F_0[(\Delta_{ji}X)'\beta_0 + \Delta_{ji}g_0(Z)] - F_0[(\Delta_{ji}X)'\beta_0]\}(\Delta_{ij}X)\bar{p}(X_i, Z_i)p(X_j, Z_j) \\ &\quad\equiv m_0(X_i, X_j, Z_i, Z_j), \end{aligned} \tag{B.41}$$

which is arbitrarily differentiable in all its arguments, so that:

$$r_n^e(\beta_0) = \int_{\Gamma_4} \gamma_n^{-k} K\left(\frac{z_1 - z_2}{\gamma_n}\right) m_1(x_1, x_2, z_1, z_2) \times f_X(x_1) f_{Z,X}(z_2, x_2) dz_1 dz_2 dx_1 dx_2, \tag{B.42}$$

where $m_1(x_1, x_2, z_1, z_2) = m_0(x_1, x_2, z_1, z_2) f_{Z|X}(z_1|x_1)$.

Now consider a Taylor Series expansion of $m_1(x_i, x_j, z_i, z_j)$ in z_i around $z_i = z_j$ up to order $t(k)$, where $t(k)$ is smallest strictly positive integer such that $t > k/2$ as specified in Assumption C2; this gives:

$$\begin{aligned} m_1(x_i, x_j, z_i, z_j) &= \sum_{s=1}^{t(k)} \left[\frac{m_1^{(s)}(x_i, x_j, z_j, z_j)}{s!}\right] \circ (\Delta_{ij} z)^s \\ &\quad + \left[\frac{m_1^{(t(k)+1)}(x_i, x_j, z^*, z_j)}{(t(k)+1)!}\right] \circ (\Delta_{ij} z)^{t(k)+1}, \end{aligned} \tag{B.43}$$

where $m_{0,ij}^{(s)}(z)$ denotes the sth-order derivative of $m_{0,ij}(z_i)$ with respect to z_i evaluated at $z_i = z$, and $z^* = \lambda z_i + (1-\lambda) z_j$ for some $0 \le \lambda \le 1$, and noting that $m_1(x_i, x_j, z_j, z_j) = 0$. Furthermore, all finite order derivatives of $F_0(\alpha)$ with respect to α are uniformly bounded and thus Assumptions C4 and C3 imply that for each $s = 1, \ldots, t(k)$, $\|m_1^{(s)}(x_i, x_j, z_i, z_j)\|$ is bounded above by a linear function of $\|x_i - x_j\|$ for all (z_i, z_j). But then for $s = 1, \ldots, t(k) - 1$:

$$\begin{aligned} &\int_{\Gamma_4} \gamma_n^{-k} K\left(\frac{\Delta_{12} z}{\gamma_n}\right) m_1^{(s)}(x_1, x_2, z_2, z_2) \circ (\Delta_{12} z)^s f_X(x_1) f_{Z,X}(z_2, x_2) dz_1 dz_2 dx_1 dx_2 \\ &= \gamma_n^s \int_{\Gamma_4} u^s K(u) m_1^{(s)}(x_1, x_2, z_2, z_2) f_X(x_1) f_{Z,X}(z_2, x_2) du\, dz_2 dx_1 dx_2 \\ &= \gamma_n^s \int_{\Gamma_3} m_1^{(s)}(x_1, x_2, z_2, z_2) f_X(x_1) f_{Z,X}(z_2, x_2) dz_2 dx_1 dx_2 \circ \int_{\mathbb{R}^k} u^s K(u) du \\ &= 0, \end{aligned} \tag{B.44}$$

by Asumption C2(ii). In addition:

$$\begin{aligned} &\left| \int_{\Gamma_4} \gamma_n^{-k} K\left(\frac{\Delta_{12} z}{\gamma_n}\right) (\Delta_{12} z)^{t(k)} m_1^{(t(k))}(x_1, x_2, z^*, z_2) f_X(x_1) f_{Z,X}(z_2, x_2) dz_1 dz_2 dx_1 dx_2 \right| \\ &\le \gamma_n^{t(k)+1} \int_{\Gamma_4} \|u\|^{t(k)} |K(u)| \cdot \|m_1^{(t(k))}(x_1, x_2, z^*, z_2)\| f_X(x_1) f_{Z,X}(z_2, x_2) du\, dz_2 dx_1 dx_2 \\ &= O(\gamma_n^{t(k)}), \end{aligned} \tag{B.45}$$

since $\int_{\mathbb{R}^k} \|u\|^{t(k)}|K(u)|du < \infty$ by Assumption C2(i), and $\|m_1^{t(k)}(x_1, x_2, z^*, z_2)\|$ is uniformly bounded above by a linear function of $\|x_1 - x_2\|$. But then (B.42), (B.43), (B.44) and (B.45) together imply that $r_n^e(\beta_0) = O(\gamma_n^{t(k)})$ and hence $n^{1/2} r_n^e(\beta_0) = O(n^{1/2}\gamma_n^{t(k)}) = o(1)$ by Assumption C1. This establishes the desired result. □

Proof of Lemma 4.4

The approach which I adopt is to show that for each fixed $\phi \in \mathbb{R}^p$ with $\phi \neq 0$, the triangular array $\left\{ \phi'[r_{n,i}^e(\beta_0) - r_n^e(\beta_0)] \right\}$, indexed by $n = 1, 2 \ldots, \infty$ and $i = 1, 2, \ldots n$ satisifes the conditions of Liapunov's CLT in the form given by Theorem 2.4.2 of Bierens (1994). For convenience in what follows I will set $\tilde{\rho}_{n,i} = \phi' \left[r_{n,i}^e(\beta_0) - r_n^e(\beta_0) \right]$ and $\rho_{n,i} = \phi' r_{n,i}^e(\beta_0)$.

First, consider the behavior of:

$$\sigma_n \; = \; n^{-1} \sum_{i=1}^{n} V_0(\tilde{\rho}_{n,i}), \tag{B.46}$$

as $n \to \infty$. Clearly, for each fixed n, the $\{\tilde{\rho}_{n,i}\}_{i=1}^n$ are iid with mean zero and therefore $\sigma_n = V_0(\tilde{\rho}_{n,i}) = E_0\left[\tilde{\rho}_{n,i}^2\right] = E_0\left[\{\phi' r_{n,i}^e(\beta_0)\}^2\right] - [\phi' r_n^e(\beta_0)]^2$. From Lemma 4.3 it is clear that $\lim_{n\to\infty} [\phi' r_n^e(\beta_0)]^2 = 0$ and thus $\lim_{n\to\infty} \sigma_n = \lim_{n\to\infty} E_0\left[\{\phi' r_{n,i}^e(\beta_0)\}^2\right] = \lim_{n\to\infty} E_0\left[\rho_{n,i}^2\right]$. Now from (4.3) and (4.5) it follows that:

$$\begin{aligned} r_{n,i}^e(\beta) &= \; E_0\left[\left\{\gamma_n^{-k} K\left(\frac{Z_i - Z_j}{\gamma_n}\right)\left(\frac{\partial l_{ij}^*(\beta)}{\partial \beta}\right)\right\} |W_i\right] \\ &= \; E_0\left\{\gamma_n^{-k} K\left(\frac{Z_i - Z_j}{\gamma_n}\right) E_0\left[\left(\frac{\partial l_{ij}^*(\beta)}{\partial \beta}\right) \mid (W_i, X_j, Z_j)\right] |W_i\right\}. \end{aligned} \tag{B.47}$$

From (B.25) it follows that:

$$\begin{aligned} E_0\left[\left(\frac{\partial l_{ij}^*(\beta)}{\partial \beta}\right) \mid (W_i, X_j, Z_j)\right] \; &= \; Y_i F_0\left[(X_j - X_i)'\beta\right] \bar{p}(X_j, Z_j)(X_i - X_j) \\ &\quad - (1 - Y_i) F_0\left[(X_i - X_j)'\beta\right] p(X_j, Z_j)(X_i - X_j). \end{aligned} \tag{B.48}$$

Straightforward manipulations then show that:

$$\begin{aligned}
&E_0\left[\gamma_n^{-k}K\left(\frac{Z_i-Z_j}{\gamma_n}\right)Y_iF_0\left\{(X_j-X_i)'\beta\right\}\bar{p}(X_j,Z_j)(X_i-X_j)\mid W_i\right]\\
&\quad= Y_i\int_{\mathbb{R}^k\times\mathbb{R}^p}\gamma_n^{-k}K\left(\frac{Z_i-z_j}{\gamma_n}\right)F_0\left\{(x_j-X_i)'\beta\right\}\bar{p}(x_j,z_j)(X_i-x_j)\\
&\qquad\qquad\qquad\times f_{Z|X}(z_j|x_j)f_X(x_j)dz_jdx_j.\\
&\quad= Y_i\int_{\mathbb{R}^k\times\mathbb{R}^p}K(u)F_0\left\{(x_j-X_i)'\beta\right\}\bar{p}(x_j,Z_i-\gamma_nu)(X_i-x_j)\\
&\qquad\qquad\qquad\times f_{Z|X}(Z_i-\gamma_nu|x_j)f_X(x_j)du\,dx_j.\\
&\quad= Y_i\int_{\mathbb{R}^k\times\mathbb{R}^p}F_0\left\{(x_j-X_i)'\beta\right\}\bar{p}(x_j,Z_i-\gamma_nu)f_{Z|X}(Z_i-\gamma_nu|x_j)(X_i-x_j)\\
&\qquad\qquad\qquad\times K(u)f_X(x_j)du\,dx_j,
\end{aligned}\tag{B.49}$$

while:

$$\begin{aligned}
&E_0\left[\gamma_n^{-k}K\left(\frac{Z_i-Z_j}{\gamma_n}\right)(1-Y_i)F_0\left\{(X_i-X_j)'\beta\right\}p(X_j,Z_j)(X_i-X_j)\mid W_i\right]\\
&\quad= (1-Y_i)\int_{\mathbb{R}^k\times\mathbb{R}^p}F_0\left\{(X_i-x_j)'\beta\right\}p(x_j,Z_i-\gamma_nu)f_{Z|X}(Z_i-\gamma_nu|x_j)(X_i-x_j)\\
&\qquad\qquad\qquad\times K(u)f_X(x_j)du\,dx_j.
\end{aligned}\tag{B.50}$$

But by the mean value theorem and Assumptions C4 and C3:

$$\begin{aligned}
&|\bar{p}(x_j,Z_i-\gamma_nu)f_{Z|X}(Z_i-\gamma_nu|x_j)\;-\;\bar{p}(x_j,Z_i)f_{Z|X}(Z_i|x_j)|\\
&\qquad= \left\|\gamma_nu'\left(\left.\frac{\partial\{\bar{p}(x_j,z)f_{Z|X}(z|x_j)\}}{\partial z}\right|_{z=Z^*}\right)\right\|\\
&\qquad\le \gamma_n\cdot\|u\|\cdot\left\|\left(\left.\frac{\partial\{\bar{p}(x_j,z)f_{Z|X}(z|x_j)\}}{\partial z}\right|_{z=Z^*}\right)\right\|\;\le\;\gamma_n\cdot M_1\cdot\|u\|,
\end{aligned}\tag{B.51}$$

where $Z^*=\lambda Z_i+(1-\lambda)(Z_i-\gamma_nu)=Z_i-(1-\lambda)\gamma_nu$ for some $0\le\lambda\le 1$, and where M_1 is a finite constant. Likewise I obtain:

$$|p(x_j,Z_i-\gamma_nu)f_{Z|X}(Z_i-\gamma_nu|x_j)\;-\;p(x_j,Z_i)f_{Z|X}(Z_i|x_j)|\;\le\;\gamma_n\cdot M_2\cdot\|u\|,\tag{B.52}$$

where M_2 is also a finite constant. Since $|Y_i|,|F_0(\alpha)|\le 1$, taken together these results imply that:

$$\left\|r^e_{n,i}(\beta) - \left\{\int_{\mathbb{R}^k\times\mathbb{R}^p} \left[Y_i F_0\left\{(x_j - X_i)'\beta\right\}\bar{p}(x_j, Z_i) - (1-Y_i)F_0\left\{(X_i - x_j)'\beta\right\}p(x_j, Z_i)\right]\right.\right.$$

$$\left.\left. \cdot f_{Z|X}(Z_i|x_j)(X_i - x_j)K(u)f_X(x_j)du\, dx_j\right\}\right\|,$$

$$\leq \int_{\mathbb{R}^k\times\mathbb{R}^p} \gamma_n \cdot (M_1 + M_2)\cdot \|u\|\cdot\|X_i - x_j\|\cdot|K(u)|f_X(x_j)du\, dx_j, \tag{B.53}$$

and hence:

$$\begin{aligned}\left\|r^e_{n,i}(\beta) - r^e_{0,i}(\beta)\right\| &\leq \gamma_n(M_1+M_2)\left\{\int_{\mathbb{R}^k}\|u\|\cdot|K(u)|du\right\}\left\{\|X_i\| + E\left(\|X_j\|\right)\right\}\\ &= \gamma_n\left(M_3 + M_4\cdot\|X_i\|\right),\end{aligned} \tag{B.54}$$

where:

$$\begin{aligned}r^e_{0,i}(\beta) = \; & E_0\left\{\left[Y_iF_0\left\{-(\Delta_{ij}X)'\beta\right\}\bar{p}(X_j, Z_i) - (1-Y_i)F_0\left\{(\Delta_{ij}X)'\beta\right\}p(X_j,Z_i)\right]\right.\\ & \left.\times f_{Z|X}(Z_i|X_j)(\Delta_{ij}X)|W_i\right\},\end{aligned} \tag{B.55}$$

and M_3 and M_4 are finite constants, from which it follows that:

$$|\rho_{n,i} - \rho_{0,i}| \leq \gamma_n\|\phi\|\cdot(M_3 + M_4\cdot\|X_i\|) = \gamma_n h_1\left(\|X_i\|\right), \tag{B.56}$$

where $\rho_{0,i} = \phi' r^e_{0,i}$. But then:

$$|\rho_{0,i}| - \gamma_n h_1\left(\|X_i\|\right) \leq |\rho_{n,i}| \leq |\rho_{0,i}| + \gamma_n h_1\left(\|X_i\|\right). \tag{B.57}$$

But from (B.55) it is clear that:

$$|\rho_{0,i}| \leq \|\phi\|\cdot(M_5 + M_6\cdot\|X_i\|) = h_2\left(\|X_i\|\right), \tag{B.58}$$

where M_5 and M_6 are finite constants. Since $H_1(\cdot), h_2(\cdot)$ are linear functions and since $E_0\left(\|X_i\|^4\right) < \infty$ by Asumption C5 and $\gamma_n \to 0$ by Assumption A1(i), then (B.57) implies that:

$$\lim_{n\to\infty} E_0[\rho^2_{n,i}] = \lim_{n\to\infty} E_0[\rho^2_{0,i}] < \infty, \tag{B.59}$$

$$\lim_{n\to\infty} E_0[\rho^4_{n,i}] = \lim_{n\to\infty} E_0[\rho^4_{0,i}] < \infty. \tag{B.60}$$

Now define $\mu_1(W_i, W_j)$ and $\mu_1(W_i, W_j)$ as in (4.9) and (4.10) so that:

$$\rho_{0,i} = E_0\left\{\left[Y_i\mu_1(W_i,W_j) - (1-Y_i)\mu_2(W_i,W_j)\right](\Delta_{ij}X)'\phi \mid W_i\right\}. \tag{B.61}$$

Clearly, from Assumptions A1 and B3(iii) it follows that $\Pr\{\mu_s(W_i, W_j) > 0 \,|\, W_i\} = 1$ for $s = 1, 2$. But since $\Pr\{(\Delta_{ij}X)'\phi \neq 0 \,|\, W_i\} = 1$ from Assumption B5 then:

$$V_0\left[\rho_{0,i} \mid (X_i, Z_i)\right] \;>\; 0, \tag{B.62}$$

since $V_0[Y_i \,|\, (X_i, Z_i)] > 0$. This then implies that $V_0(\rho_{0,i}) > 0$ and hence $E_0[\rho_{0,i}^2] > 0$. But then the conditions of Theorem 2.4.2 of Bierens (1994) are satisfied (with $\delta = 2$) giving:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\rho}_{n,i} \;\xrightarrow{D}\; N[0, \sigma_0^2], \tag{B.63}$$

where $\sigma_0^2 = \lim_{n\to\infty} E_0[\rho_{0,i}^2]$. But since $\phi \neq 0$ was fixed and arbitrary, then:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[r_{n,i}^e(\beta_0) - r_n^e(\beta_0)\right] \;\xrightarrow{D}\; N[0, A_1], \tag{B.64}$$

where:

$$A_1 \;=\; V_0\left[E_0\left\{\left[Y_i\mu_1(W_i, W_j) - (1 - Y_i)\mu_2(W_i, W_j)\right](\Delta_{ij}X) \mid W_i\right\}\right], \tag{B.65}$$

which establishes the desired result. □

Proof of Lemma 4.5

Differentiation of Q_n with respect to β reveals that:

$$\left[\frac{\partial Q_n}{\partial\beta\partial\beta'}\right] \;=\; -\binom{n}{2}^{-1} \sum_{i<j}^{n} \omega_{n,ij}(\beta), \tag{B.66}$$

where:

$$\omega_{n,ij}(\beta) \;=\; \gamma_n^{-k} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)(\Delta_{ij}Y)^2(\Delta_{ij}X\Delta_{ij}X')f_0[(\Delta_{ij}Y)(\Delta_{ij}X'\beta)], \tag{B.67}$$

in which $f_0(\alpha) \equiv e^{\alpha}(1 + e^{\alpha})^{-2}$ is the pdf of the logistic distribution evaluated at α. From the iid property of the W_i it follows that the $\omega_{n,ij}(\beta)$ are identically distributed (though not independently distributed). Hence:

$$\begin{aligned} E_0\left[\frac{\partial Q_n}{\partial\beta\partial\beta'}\right] &= -E_0[\omega_{n,ij}(\beta)] \\ &= E_0\left\{\gamma_n^{-k} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)(\Delta_{ij}Y)^2(\Delta_{ij}X\Delta_{ij}X')f_0[(\Delta_{ij}Y)(\Delta_{ij}X)'\beta]\right\}. \end{aligned} \tag{B.68}$$

Now define:

$$\begin{aligned} m_2(\beta; X_i, X_j, Z_i, Z_j) \ &\equiv \ E_0 \left\{ (\Delta_{ij} Y)^2 f_0[(\Delta_{ij} Y)(\Delta_{ij} X'\beta)] | (X_i, X_j, Z_i, Z_j) \right\} \\ &= \ f_0[(\Delta_{ij} X)'\beta] p(X_i, Z_i) \bar{p}(X_j, Z_j) + f_0[-(\Delta_{ij} X)'\beta] \bar{p}(X_i, Z_i) p(X_j, Z_j) \\ &= \ f_0[(\Delta_{ij} X)'\beta] \left[p(X_i, Z_i) \bar{p}(X_j, Z_j) + \bar{p}(X_i, Z_i) p(X_j, Z_j) \right], \end{aligned} \tag{B.69}$$

and observe that $m_2(\beta; X_i, X_j, Z_i, Z_j)$ is clearly continuous in all its arguments and lies in the range $[0, 1/2]$. Then I have that:

$$\begin{aligned} E_0 \left[\omega_{n,ij}(\beta) \right] \ &= \ E_0 \left\{ \gamma_n^{-k} K \left(\frac{\Delta_{ij} Z}{\gamma_n} \right) m_2(\beta; X_i, X_j, Z_i, Z_j) \right\} \\ &= \ \int_{\Gamma_4} \gamma_n^{-k} K \left(\frac{\Delta_{12} z}{\gamma_n} \right) m_2(\beta; x_1, x_2, z_1, z_2)(\Delta_{12} x)(\Delta_{12} x') \\ &\qquad \times f_{Z,X}(z_1, x_1) f_{Z,X}(z_2, x_2) \, dz_1 \, dz_2 \, dx_1 \, dx_2, \\ &= \ \int_{\Gamma_4} m_2(\beta; x_1, x_2, z_2 + \gamma_n u, z_2) f_{Z|X}(z_2 + \gamma_n u | x_1)(\Delta_{12} x)(\Delta_{12} x') \\ &\qquad \times K(u) f_X(x_1) f_{Z,X}(z_2, x_2) \, du \, dz_2 \, dx_1 \, dx_2. \end{aligned} \tag{B.70}$$

Since $f_{Z|X}(z_2 + \gamma_n u | x_1)$ is bounded in absolute value, $0 < p(x, z) < 1$, $0 < f_0(\alpha) \leq (1/4)$ for all α, and $E_0 \left(\|X_i\|^2 \right) < \infty$, it follows by the dominated convergence theorem that:

$$E_0 \left[\omega_{n,ij}(\beta) \right] \ \to \ A_2^*(\beta) \ \equiv \ E_0 \left\{ m_2(\beta; X_1, X_2, Z_2, Z_2) f_{Z|X}(Z_2 | X_1)(\Delta_{12} X)(\Delta_{12} X)' \right\}. \tag{B.71}$$

Evaluating this at $\beta = \beta_0$ gives:

$$\begin{aligned} \left[\frac{\partial Q_n}{\partial \beta \partial \beta'} \right] \ &\to \ A_2 \ = \ E_0 \left\{ m_2(\beta_0; X_1, X_2, Z_2, Z_2) f_{Z|X}(Z_2 | X_1)(\Delta_{12} X)(\Delta_{12} X') \right\} \\ &= \ E_0 \left\{ \left[p(X_1, Z_1) \bar{p}(X_2, Z_2) + \bar{p}(X_1, Z_1) p(X_2, Z_2) \right] \right. \\ &\qquad \left. \times \ f_0[-(\Delta_{12}) X'\beta] f_{Z|X}(Z_2 | X_1)(\Delta_{12} X)(\Delta_{12} X)' \right\}. \end{aligned} \tag{B.72}$$

Clearly $m_2(\beta_0; X_1, X_2, Z_2, Z_2)$ is almost surely strictly positive, and $f_{Z|X}(Z_2 | X_1)$ is almost surely strictly positive by Assumption B3(iii). In addition, $E_0[(\Delta_{12} X)(\Delta_{12} X)']$ is non-singular in view of Assumption B5. Together these imply that A_2 is non-singular which establishes the desired result. □

Proof of Lemma 4.6

Observe that (B.67) implies that:

$$\begin{aligned}\|\omega_{n,ij}(\beta)\|^2 &= \text{tr}[\omega_{n,ij}(\beta)\omega_{n,ij}(\beta)'] \\ &= \gamma_n^{-2}K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2(\Delta_{ij}Y)^4\|\Delta_{ij}X\|^4 f_0[\Delta_{ij}Y(\Delta_{ij}X'\beta)]^2 \\ &\leq (1/2)\gamma_n^{-2}K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2(\|X_i\|^4+\|X_j\|^4), \end{aligned} \quad \text{(B.73)}$$

since the pdf of the logistic distribution is bounded above by $(1/4)$, $(\Delta_{ij}Y)^4$ is bounded above by 1, and $\|\Delta_{ij}X\|^4 \leq (2\|X_i\|^2+2\|X_j\|^2)^2 \leq 8(\|X_i\|^4+\|X_j\|^4)$. Hence,

$$E_0\left(\|\omega_{n,ij}(\beta)\|^2\right) \;\leq\; \gamma_n^{-1}E_0\left[\gamma_n^{-1}K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2\cdot\|X_i\|^4\right], \quad \text{(B.74)}$$

using the iid property of the W_i and the symmetry of $K(\cdot)$ from Assumptions A1 and A3. But then by parallel arguments to those used in the proof of Lemma 4.2 to establish (B.30) it follows that:

$$E_0\left[\gamma_n^{-1}K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2\cdot\|X_i\|^4\right] \;\leq\; L_1L_2L_7<\infty, \quad \text{(B.75)}$$

and hence that:

$$E_0\left(\|\omega_{n,ij}(\beta)\|^2\right) \;=\; O(\gamma_n^{-k})=o(n). \quad \text{(B.76)}$$

Lemma A.1 then implies that:

$$\left[\frac{\partial^2 Q_n}{\partial\beta\partial\beta'}\right] \;=\; E_0\left[\frac{\partial^2 Q_n}{\partial\beta\partial\beta'}\right] \;+\; o_p(1) = \; E_0\left[\omega_{n,ij}(\beta)\right] \;+\; o_p(1), \quad \text{(B.77)}$$

which establishes the desired result. □

Proof of Lemma 4.7

Let ξ_1,ξ_2 be two arbitrary non-stochastic $(p\times 1)$ vectors and consider the behaviour of:

$$C_n(\xi_1,\xi_2) \;=\; -\binom{n}{2}^{-1}\sum_{i<j}^{n}\xi_1'\left[\omega_{n,ij}(\beta_n^*)-\omega_{n,ij}(\beta_0)\right]\xi_2. \quad \text{(B.78)}$$

Applying the mean value theorem to each term gives:

$$\xi_1' \left[\omega_{n,ij}(\beta_n^*) - \omega_{n,ij}(\beta_0)\right] \xi_2 \;=\; \left(\frac{\partial[\xi_1'\omega_{n,ij}(\beta)\xi_2]}{\partial\beta} \bigg|_{\beta_{n,ij}^{**}} \right)' (\beta_n^* - \beta_0), \tag{B.79}$$

where $\beta_{n,ij}^{**}$ is a convex combination of β_n^* and β_0, whose coefficients depend on the term in question. Substituting this into (B.78) gives:

$$C_n(\xi_1, \xi_2) \;=\; -\binom{n}{2}^{-1} \sum_{i<j}^{n} \left(\frac{\partial[\xi_1'\omega_{n,ij}(\beta)\xi_2]}{\partial\beta} \bigg|_{\beta_{n,ij}^{**}} \right)' (\beta_n^* - \beta_0), \tag{B.80}$$

and hence:

$$|C_n(\xi_1, \xi_2)| \;\leq\; \binom{n}{2}^{-1} \sum_{i<j}^{n} \left\| \left(\frac{\partial[\xi_1'\omega_{n,ij}(\beta)\xi_2]}{\partial\beta} \bigg|_{\beta_{n,ij}^{**}} \right) \right\| \cdot \|\beta_n^* - \beta_0\|. \tag{B.81}$$

Now diferentiation of (B.67) reveals that:

$$\begin{aligned} \left[\frac{\partial[\xi_1'\omega_{n,ij}(\beta)\xi_2]}{\partial\beta} \right] \;=\; & \gamma_n^{-k} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right) (\Delta_{ij}Y)^3 (\Delta_{ij}X)[(\Delta_{ij}X)'\xi_1][(\Delta_{ij}X)'\xi_2] \\ & \times f_0^{(1)}[(\Delta_{ij}Y)(\Delta_{ij}X)'\beta], \end{aligned} \tag{B.82}$$

where $f_0^{(1)}(\cdot)$ denotes the first derivative of the logistic pdf. Hence:

$$\left\| \left(\frac{\partial[\xi_1'\omega_{n,ij}(\beta)\xi_2]}{\partial\beta} \right) \right\| \;\leq\; \gamma_n^{-k} \left| K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right) \right| \cdot \|\Delta_{ij}X\|^3 \cdot \|\xi_1\| \cdot \|\xi_2\|, \tag{B.83}$$

for all β, which substituted into (B.81) implies that:

$$\begin{aligned} |C_n(\xi_1, \xi_2)| \;\leq\; & \binom{n}{2}^{-1} \sum_{i<j}^{n} \gamma_n^{-k} \left| K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right) \right| \cdot \|\Delta_{ij}X\|^3 \cdot \|\xi_2\| \cdot \|\xi_2\| \cdot \|\beta_n^* - \beta_0\| \\ =\; & H_{2,n} \cdot \|\xi_1\| \cdot \|\xi_2\| \cdot \|\beta_n^* - \beta_0\|. \end{aligned} \tag{B.84}$$

It is clear that:

$$E_0[H_{2,n}] \;=\; E_0 \left\{ \gamma_n^{-k} \left| K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right) \right| \cdot \|\Delta_{ij}X\|^3 \right\}, \tag{B.85}$$

converges to a finite limit as $n \to \infty$ following the same line of argument used in the proof of Lemma 3.3 to establish (B.17) in view of the finiteness of $E\{\|X_i\|^3\}$ implied by Assumption C5. Since $H_{2,n}$ is non-negative it follows by the Markov inequality that $H_{2,n} = O_p(1)$.

Then since β_n^* lies on the line segment joining β_0 and $\hat{\beta}_n$ it follows that $\|\beta_n^* - \beta_0\| \leq \|\hat{\beta}_n - \beta_0\|$ and thus that:

$$|C_{2,n}(\xi)| \;\leq\; H_{2,n} \cdot \|\xi_1\| \cdot \|\xi_2\| \cdot \|\hat{\beta}_n - \beta_0\| \;=\; O_p(1) \times O(1) \times O(1) \times o_p(1) \;=\; o_p(1), \tag{B.86}$$

since $\xi_1 = O(1)$ and $\xi_2 = O(1)$ by assumption and since $\|\hat{\beta}_n - \beta_0\| = o_p(1)$ by Theorem 3.1. This establishes the desired result. □

Proof of Lemma 5.1

As in the proof of Lemma 4.7, let ξ_1, ξ_2 be two arbitrary non-stochastic $(p \times 1)$ vectors and consider the behaviour of:

$$\begin{aligned} a_{11,n}(\xi_1,\xi_2) &= \xi_1' \hat{A}_{11,n} \xi_2 \\ &= n^{-1}(n-1)^{-2} \sum_{i \neq j}^{n} [r_{n,ij}(\hat{\beta}_n)'\xi_1][r_{n,ij}(\hat{\beta}_n)'\xi_2] \\ &= n^{-1}(n-1)^{-2} \sum_{i \neq j}^{n} \gamma_n^{-2k} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2 (\Delta_{ij}Y)^2 \\ &\qquad \cdot F_0[-(\Delta_{ij}Y)(\Delta_{ij}X)'\hat{\beta}_n]^2 \cdot [(\Delta_{ij}X)'\xi_1] \cdot [(\Delta_{ij}X)'\xi_2], \end{aligned} \tag{B.87}$$

from which it is clear that:

$$\begin{aligned} |a_{11,n}(\xi_1,\xi_2)| &\leq n^{-1}(n-1)^{-2} \sum_{i \neq j}^{n} \gamma_n^{-2k} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2 (\Delta_{ij}Y)^2 \\ &\qquad \cdot F_0[-(\Delta_{ij}Y)(\Delta_{ij}X)'\hat{\beta}_n]^2 \cdot \|\Delta_{ij}X\|^2 \cdot \|\xi_1\| \cdot \|\xi_2\| \\ &\leq n^{-1}(n-1)^{-2} \sum_{i \neq j}^{n} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2 \|\Delta_{ij}X\|^2 \cdot \|\xi_1\| \cdot \|\xi_2\| \\ &= H_{3,n} \cdot \|\xi_1\| \cdot \|\xi_2\|. \end{aligned} \tag{B.88}$$

But then (B.30) from the proof of Lemma 4.2, which uses a subset of the assumptions currently made, implies that:

$$E_0[H_{3,n}] = (n-1)^{-1} E_0\left[\gamma_n^{-2k} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2 \|\Delta_{ij}X\|^2\right] \leq (n-1)^{-1}\gamma_n^{-k} L_1 L_2 L_3, \tag{B.89}$$

which tends to zero. Since $H_{3,n}$ is non-negative it follows by the Markov inequality that $H_{3,n} \xrightarrow{p} 0$. But this implies that $|a_{11,n}(\xi_1,\xi_2)| \xrightarrow{p} 0$ and hence that $\hat{A}_{11,n} \xrightarrow{p} 0$ since ξ_1 and ξ_2 were assumed fixed. This establishes the desired result. □

Proof of Lemma 5.2

Observe that:

$$E_0\left[n^{-1}(n-1)^{-2} \sum_{i \neq j \neq l}^{n} r_{n,ij}(\beta_0) r_{n,il}(\beta_0)'\right] = \left(\frac{n-2}{n-1}\right) E_0\left[r_{n,ij}(\beta_0) r_{n,ik}(\beta_0)'\right] \tag{B.90}$$

Now note that:

$$\text{Cov}_0\left[r_{n,ij}(\beta_0), r_{n,ik}(\beta_0)'\right] \;=\; V_0\Big[E_0\big\{r_{n,ij}(\beta_0)|W_i\big\}\Big], \tag{B.91}$$

since W_i, W_j and W_l are iid; hence:

$$\begin{aligned} E_0\left[r_{n,ij}(\beta_0)r_{n,ik}(\beta_0)'\right] \;&=\; V_0\Big[E_0\big\{r_{n,ij}(\beta_0)|W_i\big\}\Big] + E_0\big\{r_{n,ij}(\beta_0)\big\}E_0\big\{r_{n,ij}(\beta_0)\big\}' \\ &=\; V_0\big[r^e_{n,i}(\beta_0)\big] + r^e_n(\beta_0)r^e_n(\beta_0)'. \end{aligned} \tag{B.92}$$

Lemma 4.3 implies that $r^e_n(\beta_0)$ converges to zero, and the proof of Lemma 4.4 establishes that $V_0\big[r^e_{n,i}(\beta_0)\big]$ converges to A_1. Since $(n-2)/(n-1)$ converges to one this establishes the desired result. □

Proof of Lemma 5.3

First observe that:

$$n^{-1}(n-1)^{-2}\sum_{i\neq j\neq l}^{n} r_{n,ij}(\beta)r_{n,il}(\beta)' \;=\; \binom{n}{3}^{-1}\sum_{i<j<l}\psi_{n,ijl}(\beta) \tag{B.93}$$

where:

$$\begin{aligned} \psi_{n,ijl}(\beta) \;=\; & r_{n,ij}(\beta)r_{n,il}(\beta)' + r_{n,il}(\beta)r_{n,ij}(\beta)' + r_{n,ji}(\beta)r_{n,jl}(\beta)' \\ & + r_{n,jl}(\beta)r_{n,ji}(\beta)' + r_{n,li}(\beta)r_{n,lj}(\beta)' + r_{n,lj}(\beta)r_{n,li}(\beta)' \end{aligned} \tag{B.94}$$

which is clearly symmetric in the indices (i,j,l), so that the right-hand-side expression in (B.93) is a third-order symmetric U-statistic function to which it is possible to apply Lemma A.1. Furthermore, the six terms in the right-hand-side expression in (B.94) have an exchangeable joint distribution and hence have identical means and cross-covariances. Thus:

$$E_0\big[\psi_{n,ijl}(\beta)\big] \;=\; 6E_0\big[r_{n,ij}(\beta)r_{n,il}(\beta)'\big], \quad V_0\big[\xi_i'\psi_{n,ijl}(\beta)\xi_2\big] \;\leq\; 36V_0\big[\xi_1'r_{n,il}(\beta)r_{n,ij}(\beta)'\xi_2\big], \tag{B.95}$$

for any fixed $(p\times 1)$ vectors ξ_1 and ξ_2 and all β. Now consider the behaviour of $E_0\big[\{\xi_1'r_{n,ij}(\beta)\}^2\{r_{n,il}(\beta)'\xi_2\}^2\big]$. Observe that:

$$\begin{aligned} \{\xi_1'r_{n,ij}(\beta)\}^2\{\xi_2'r_{n,il}(\beta)\}^2 \;=\; & \gamma_n^{-4k}K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2 K\left(\frac{\Delta_{il}Z}{\gamma_n}\right)^2(\Delta_{ij}Y)^2(\Delta_{il}Y)^2 \\ & F_0[-(\Delta_{ij}Y)(\Delta_{ij}X)'\beta]^2F_0[-(\Delta_{il}Y)(\Delta_{il}X)'\beta]^2[(\Delta_{ij}X)'\xi_1]^2[(\Delta_{il}X)'\xi_2]^2. \end{aligned} \tag{B.96}$$

Now define:

$$\begin{aligned} m_4(\beta;X_i,X_j,X_l,Z_i,Z_j,Z_l) \;=\; & E_0\big\{(\Delta_{ij}Y)^2(\Delta_{il}Y)^2F_0[-(\Delta_{ij}Y)(\Delta_{ij}X)'\beta]^2 \\ & F_0[-(\Delta_{il}Y)(\Delta_{il}X)'\beta]^2|(X_i,X_j,X_l,Z_i,Z_j,Z_l)\big\}, \end{aligned} \tag{B.97}$$

and observe that $m_4(\cdot)$ is continuous in all its arguments and lies in the range $[0,2]$; then:

$$\begin{aligned}
&E_0\big[\{\xi_1' r_{n,ij}(\beta)\}^2\{\xi_2' r_{n,il}(\beta)\}^2\big] \\
&\quad= E_0\big\{\gamma_n^{-4k} K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)^2 K\left(\frac{\Delta_{il}Z}{\gamma_n}\right)^2 \\
&\qquad\qquad m_4(\beta; X_i, X_j, X_l, Z_i, Z_j, Z_l)[(\Delta_{ij}X)'\xi_1]^2[(\Delta_{il}X)'\xi_2]^2 \\
&\quad= \gamma_n^{-2k} E_0^*\big\{K(U_j)K(U_l)m_4(\beta; X_i, X_j, X_l, Z_i, Z_i - \gamma_n U_j, Z_i - \gamma_n U_l) \\
&\qquad\qquad f_{Z|X}(Z_i - \gamma_n U_j|X_j) f_{Z|X}(Z_i - \gamma_n U_l|X_l)[(\Delta_{ij}X)'\xi_1]^2[(\Delta_{il}X)'\xi_2]^2\big\} \\
&\quad\leq 2\gamma_n^{-2k} L_1^2 L_2^2 E_0^*\big\{\|\Delta_{ij}X\|^2 \cdot \|\Delta_{il}X\|^2\big\} \;=\; o(n),
\end{aligned} \tag{B.98}$$

by Assumptions C5 and D1. But this implies that $E_0[\{\xi_1'\psi_{n,ijl}(\beta)\xi_2\}^2] = o(n)$ so that:

$$n^{-1}(n-1)^{-2} \sum_{i\neq j\neq l}^{n} r_{n,ij}(\beta) r_{n,il}(\beta)' \;=\; E_0[\xi_1'\psi_{n,ijl}(\beta)\xi_2] + o_p(1), \tag{B.99}$$

by application of Lemma A.1, and since ξ_1 and ξ_2 were arbitrary fixed vectors this establishes the deisred result. □

Proof of Lemma 5.4

Again let ξ_1 and ξ_2 be arbitrary fixed $(p \times 1)$ vectors and consider:

$$\begin{aligned}
&\Big| n^{-1}(n-1)^{-2} \sum_{i\neq j\neq l}^{n} \xi_1'\big\{r_{n,ij}(\hat\beta_n) r_{n,il}(\hat\beta_n)' - r_{n,ij}(\beta_0) r_{n,il}(\beta_0)'\big\}'\xi_2 \Big| \\
&\qquad\qquad \leq n^{-1}(n-1)^{-2} \sum_{i\neq j\neq l}^{n} \big|\xi_1'\big\{r_{n,ij}(\hat\beta_n) r_{n,il}(\hat\beta_n)' - r_{n,ij}(\beta_0) r_{n,il}(\beta_0)'\big\}'\xi_2\big|.
\end{aligned} \tag{B.100}$$

By application of the mean value theorem to each term I have that:

$$\begin{aligned}
&\xi_1'\big\{r_{n,ij}(\hat\beta_n) r_{n,il}(\hat\beta_n)' - r_{n,ij}(\beta_0) r_{n,il}(\beta_0)'\big\}'\xi_2 \\
&\quad= -\Big\{\big[\omega_{n,ij}(\beta^*_{n,ijl})\xi_1\big]\big[r_{n,ij}(\beta^*_{n,ijl})'\xi_2\big] + \big[\omega_{n,il}(\beta^*_{n,ijl})\xi_2\big]\big[r_{n,il}(\beta^*_{n,ijl})'\xi_1\big]\Big\}(\hat\beta_n - \beta_0),
\end{aligned} \tag{B.101}$$

where $\beta^*_{n,ijl}$ lies on the line segment joining $\hat\beta_n$ and β_0. Hence:

$$\begin{aligned}
&n^{-1}(n-1)^{-2} \sum_{i\neq j\neq l}^{n} \big|\xi_1'\big\{r_{n,ij}(\hat\beta_n) r_{n,il}(\hat\beta_n)' - r_{n,ij}(\beta_0) r_{n,il}(\beta_0)'\big\}'\xi_2\big| \\
&\qquad\qquad \leq H_{4,n} \cdot \|\xi_1\| \cdot \|\xi_2\| \cdot \|\hat\beta_n - \beta_0\|,
\end{aligned} \tag{B.102}$$

where:

$$H_{4,n} \;=\; n^{-1}(n-1)^{-2} \sum_{i \neq j \neq l}^{n} \Big\{ \|\omega_{n,ij}(\beta^*_{n,ijl})\| \cdot \|r_{n,ij}(\beta^*_{n,ijl})\| + \|\omega_{n,il}(\beta^*_{n,ijl})\| \cdot \|r_{n,il}(\beta^*_{n,ijl})\| \Big\}. \tag{B.103}$$

But from (B.67) it is clear that:

$$\|\omega_{n,ij}(\beta)\xi\| \;\leq\; \gamma_n^{-k} \left| K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)\right| \cdot \|\Delta_{ij}X\|^2 \cdot \|\xi\|, \tag{B.104}$$

for all β while from (4.5) it is clear that:

$$|r_{n,ij}(\beta)'\xi| \;\leq\; \gamma_n^{-k} \left| K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)\right| \cdot \|\Delta_{ij}X\| \cdot \|\xi\|. \tag{B.105}$$

From these it follows that:

$$\begin{aligned} &\|\omega_{n,ij}(\beta^*_{n,ijl})\| \cdot \|r_{n,ij}(\beta^*_{n,ijl})\| + \|\omega_{n,il}(\beta^*_{n,ijl})\| \cdot \|r_{n,il}(\beta^*_{n,ijl})\| \\ &\quad \leq\; \gamma_n^{-2k} \left| K\left(\frac{\Delta_{ij}Z}{\gamma_n}\right)\right| \cdot \left| K\left(\frac{\Delta_{il}Z}{\gamma_n}\right)\right| \cdot \|\Delta_{ij}X\| \cdot \|\Delta_{il}X\| \cdot \big(\|\Delta_{ij}X\| + \|\Delta_{il}X\|\big) \cdot \|\xi_1\| \cdot \|\xi_2\|. \end{aligned} \tag{B.106}$$

But then by the same line of arguments as used in the proof of (B.89) from Lemma 5.1 it follows that the expectation of the right-hand-side is uniformly bounded for all n and hence is $O_p(1)$. But then $H_{4,n} = O_p(1)$ and thus $H_{4,n} \cdot \|\xi_1\| \cdot \|\xi_2\| \cdot \|\hat{\beta}_n - \beta_0\| = o_p(1)$ since $\hat{\beta}_n \xrightarrow{p} \beta_0$ by Lemma 3.1 and since ξ_1 and ξ_2 are fixed $(p \times 1)$ vectors. In view of (B.100), this establishes that:

$$\left| n^{-1}(n-1)^{-2} \sum_{i \neq j \neq l}^{n} \xi_1' \big\{ r_{n,ij}(\hat{\beta}_n) r_{n,il}(\hat{\beta}_n)' - r_{n,ij}(\beta_0) r_{n,il}(\beta_0)' \big\}' \xi_2 \right| \;=\; o_p(1), \tag{B.107}$$

which in turn establishes the desired result since ξ_1 and ξ_2 are arbitrary. □

Proof of Lemma 5.5

First observe that following exactly the same line of reasoning as used in the proof of Lemma 4.7:

$$\left[\frac{\partial^2 Q_n}{\partial\beta\partial\beta'}\bigg|_{\hat{\beta}_n} \right] - \left[\frac{\partial^2 Q_n}{\partial\beta\partial\beta'}\bigg|_{\beta_0} \right] \;\xrightarrow{p}\; 0, \tag{B.108}$$

since $\|\hat{\beta}_n - \beta_0\| = o_p(1)$ just as $\|\beta^*_n - \beta_0\| = o_p(1)$. The desired result follows immediately from (B.108) combined with Lemmas 4.5 and 4.6. □

References

Ahn, H., and J.L. Powell (1993) ‘Semiparametric estimation of censored selection models with a nonparametric selection mechanism.’ *Journal of Econometrics* 58, 3–29

Amemiya, T. (1985) *Advanced Econometrics* (Oxford: Basil Blackwell)

Bierens, H.J. (1994) *Topics in Advanced Econometrics: Estimation, testing, and specification of cross-section and time series models* (Cambridge: Cambridge University Press)

Chamberlain, G. (1980) ‘Analysis of covariance with qualitative data.’ *The Review of Economic Studies* 47, 225–238

Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians* (Oxford: Oxford University Press)

Honore, B.E., E. Kyriazidou, and C. Udry (1997) ‘Estimation of type 3 tobit models using symmetric trimming and pairwise comparisons.’ *Journal of Econometrics* 76(1-2), 107–128

Jennrich, R.I. (1969) ‘Asymptotic properties of nonlinear least squares estimators.’ *Annals of Mathematical Statistics* 40, 633–643

Lee, B. (1988) ‘Nonparametric tests using a kernel estimation method.’ Doctoral Dissertation. Department of Economics, University of Wisconsin, Madison, WI

McFadden, D.L. (1974) ‘Conditional logit analysis of qualitative choice behavior.’ In *Frontiers in Econometrics*, ed. P. Zarembka Academic Press New York, NY pp. 105–142

____ (1976) ‘Quantal choice analysis: A survey.’ *Annals of Economic and Social Measurement* 5(4), 363–390

Powell, J.L., J.H. Stock, and T.M. Stoker (1989) ‘Semiparametric estimation of weighted average derivatives.’ *Econometrica* 57, 1403–1430

Yatchew, A. (1997) ‘An elementary estimator of the partial linear model.’ *Economics Letters* 57, 135–143