1

Local Asymptotic Normality of Ranks and Covariates in Transformation Models

P.J. Bickel[1]
and
Y. Ritov[2]

1.1 Introduction

Le Cam and Yang (1988) addressed broadly the following question: Given observations $X^{(n)} = (X_{1n}, \ldots, X_{nn})$ distributed according to $P_\theta^{(n)}$; $\theta \in R^k$ such that the family of probability measures $\{P_\theta^{(n)}\}$ has a locally asymptotically normal (LAN) structure at θ_0 and a statistic $Y^{(n)} = g_n(X^{(n)})$:

(i) When do the distributions of $Y^{(n)}$ also have an LAN structure at θ_0?

(ii) When is there no loss in information about θ in going from $X^{(n)}$ to $Y^{(n)}$?

In this paper we exhibit an important but rather complicated example to which the Le Cam-Yang methods may, after some work, be applied.

Semiparametric transformation models arise quite naturally in survival analysis. Specifically, let $(Z_1, T_1), (Z_2, T_2), \ldots, (Z_n, T_n)$ be independent and identically distributed with Z, a vector of covariates, having distribution H, and T real. We suppose there exists an *unknown strictly monotone* transformation $a_0 : R \to R$ such that, given $Z = z$, $a_0(T)$ is distributed with distribution function $F(\cdot, z, \theta)$ where $\{F(\cdot, z, \theta) : \theta \in \Theta \subset R^d\}$ is a regular parametric model. That is, the $F(\cdot, z, \theta)$ are dominated by μ with densities $f(\cdot, z, \theta)$ and the map $\theta \to \sqrt{f(\cdot, z, \theta)}$ is Hellinger differentiable. As is usual in these models we take μ to be Lebesgue measure and $a_0' > 0$. The most important special cases of these models are the regression models where $\theta = (\eta, \nu)$, and defining the distribution of T given Z structurally,

$$a_0(T) = \eta + \nu' Z + \epsilon,$$

[1]University of California, Berkeley, USA
[2]The Hebrew University of Jerusalem, Israel

where ϵ is independent of Z with fixed known distribution G_0. If G_0 is an extreme value distribution this is the Cox proportional hazards model. If G_0 is log Pareto this is the Clayton-Cuzick (1985) model which introduces a gamma distributed frailty into the Cox model. Finally, if G_0 is Gaussian this is the natural semiparametric extension of the Box-Cox model (see Doksum, 1987, for instance). If θ is fixed in any of these semiparametric models it is evident that a maximal invariant under the group of monotone transformations of T is the vector

$$\tilde{Z} \equiv (Z_{(1)}, \ldots, Z_{(n)}),$$

where $T_{(1)} < \ldots < T_{(n)}$ are the ordered T_i and $Z_{(i)}$ is the covariate of $T_{(i)}$, that is, $(Z_{(i)}, T_{(i)})$ $i = 1, \ldots n$ is the appropriate permutation of (Z_i, T_i), $i = 1, \ldots, n$. Knowing $\tilde{Z}$ is equivalent to knowing the ranks of the T_i and the corresponding $Z_1, \ldots, Z_n$. It is intuitively plausible that, asymptotically, the (marginal) likelihood of $\tilde{Z}$ which doesn't depend on a_0 can be used for inference about θ in the usual way, without any loss of information. That is, question (i) and (ii) can be answered affirmatively if $P_\theta^{(n)}$ is taken to be, in some sense, the least favorable family for estimation of θ_0. That is, we take $a_\theta(T)$ given $Z = z$ to have distribution $F_\theta(\cdot|z)$ with a_θ so chosen as to make this the hardest parametric submodel of our semiparametric model at θ_0 (in the sense of Stein (1956) – see also Bickel, Klaassen, Ritov, and Wellner, 1993, pages 153–175). The detailed construction is given in what follows.

Implementation of this approach for inference is well known and simple only for the Cox model where the likelihood of $\tilde{Z}$ is the Cox partial likelihood. In general, the likelihood is expressible only as an n fold integral. Thus if Λ_n is the log likelihood ratio of $\theta_0 + sn^{-1/2}$ vs θ_0, we have,

$$\begin{aligned} \Lambda_n(s) \quad = \quad & \log \int_{t_1 < \ldots < t_n} \cdots \int \prod_{j=1}^{n} f(t_j, Z_{(j)}, \theta_0 + sn^{-1/2})\, dt_j \\ & - \log \int_{t_1 < \ldots < t_n} \cdots \int \prod_{j=1}^{n} f(t_j, Z_{(j)}, \theta_0)\, dt_j \end{aligned} \qquad (1.1)$$

However, it is possible to use Monte Carlo methods in a subtle way, drawing on some of the information we develop for our result, to compute Λ_n accurately enough to use it for inference. Alternatively, the analytic approximation to Λ_n that we develop can be used more conveniently for this purpose also – see Bickel (1986). We shall pursue these approaches elsewhere.

The paper is organized as follows. In section 2 we introduce notation and the least favorable family a_θ and establish LAN directly under restrictive conditions. In section 3 we state our general result and show how it follows from Le Cam and Yang's Theorem 4.

1.2 LAN – the bounded case

For simplicity we state our results for the case θ one dimensional. There is no real loss of generality since proofs carry over easily. Here is some notation. We denote the distribution of Z by H, which we take as known and independent on θ (this is irrelevant). Then if $P_{(\theta,a)}$ is the distribution of (T, Z) and (θ_0, a_0) is the true state of Nature,

$$\frac{dP_{(\theta,a)}}{dP_{(\theta_0,a_0)}}(t,z) = \frac{a'}{a_0'}(t)\frac{f(a(t),z,\theta)}{f(a_0(t),z,\theta_0)}.$$

It is convenient to reparametrize the model. Let $U = F_Y \circ a_0(T)$, where

$$F_Y(t) = \int\int_{-\infty}^{t} f(s,z,\theta_0)dsdH(z)$$

is the marginal df of $a_0(T)$. Suppose F_Y is strictly increasing (this is inessential). Then the conditional density of $U|Z = z$ under $P_{(\theta,a_0)}$ is

$$g(u,z,\theta) \equiv f(F_Y^{-1}(u),z,\theta)/f_Y(F_Y^{-1}(u)) \tag{2.1}$$

where $f_Y = F_Y'$. So if $b \equiv F_Y \circ a \circ a_0^{-1} \circ F_Y^{-1}$ and $Q_{(\theta,b)}$ is the distribution of U under $P_{(\theta,a)}$

$$\frac{dQ_{(\theta,b)}}{dQ_{(\theta_0,b_0)}} = b'(u)\frac{g(b(u),z,\theta)}{g(u,z,\theta_0)}, \quad 0 < u < 1$$

the likelihood ratio for a transformation model where transformations are from $(0, 1)$ to $(0, 1)$ and $b_0 = F_Y \circ a_0 \circ a_0^{-1} \circ F_Y^{-1}$ is the identity. The distribution of the ranks of U under $Q_{(\theta,b)}$ is the same as the distribution of the ranks of T under $P_{(\theta,a)}$. We formulate our conditions in terms of derivatives of $\lambda \equiv \log g$ which hold in all cases we have mentioned when f is related to g by (2.1). For convenience in what follows, we let $\lambda_\theta = \frac{\partial\lambda}{\partial\theta}$, $\lambda_{u\theta} = \frac{\partial^2\lambda}{\partial u\partial\theta}$ and, in general, let subscripts denote differentiation with primes used for functions of u only. We will use the following assumptions;

A1 The function $\lambda(u, z, \theta)$ is twice differentiable in (θ, u) on $V(\theta_0)\times(0, 1)$ where $V(\theta_0)$ is a neighborhood of θ_0.

B: λ, λ_θ, λ_u, $\lambda_{\theta u}$, λ_{uu} are uniformly bounded in (u, z, θ) on $V(\theta_0)$.

B is an extremely restrictive assumption. It permits essentially only families, such that $g(u, z, \theta)$ is bounded away from 0 on $[0, 1]$ and in particular rules out all our examples. However, the argument here makes clear the essential computation which, in the next section, enables us to apply the Le Cam-Yang results to all our examples.

Formal derivation of least favorable b_θ

Without loss of generality let $\theta_0 = 0$, but without any presumption that 0 corresponds to independency of Z and T. Let

$$b_\theta(u) = u + \theta\Delta(u) \tag{2.2}$$

where $\Delta \in \mathcal{D}$, and $\mathcal{D} \equiv \{\Delta : \Delta, \Delta'$ and Δ'' are all bounded on $[0,1]$ and $\Delta(0) = \Delta(1) = 0\}$. Then for $|\theta| < \epsilon$, $\epsilon > 0$ the b_θ are transformations of u that depend on θ and b_0 is the identity. Under A1 and B the model $\{Q_{(\theta,b_\theta)} : |\theta| < \epsilon\}$ is regular and the score function at θ is,

$$v_\Delta(u,z,\theta) \equiv \lambda_\theta(u,z,\theta) + \Delta'(u) + \Delta(u)\lambda_u(u,z,\theta).$$

This follows since the map $\theta \to \sqrt{q_{(\theta,b_\theta)}}$ where $q_{(\theta,b)} \equiv dQ_{(\theta,b)}/dQ_{(0,b_0)}$, is pointwise differentiable by A1 and therefore Hellinger differentiable by B.

By standard theory, see for example Bickel et al., if the Fisher information $\int v_\Delta^2(u,z,0)g(u,z,0)dudH(z)$ is minimized over the closure in $L_2(Q_{(0,b_0)})$ of $\{v_\Delta : \Delta \in \mathcal{D}\}$ by v_{Δ_0} with $\Delta_0 \in \mathcal{D}$ then

$$Ev_{\Delta_0}(U,Z,0)(\Delta'(U) + \Delta(U)\lambda_u(U,Z,0)) = 0 \tag{2.3}$$

for all $\Delta \in \mathcal{D}$. Furthermore, Klaassen (1992) shows that, under regularity conditions, (2.3) holds if and only if $\Delta_0 \in \mathcal{D}$ satisfies the Sturm-Liouville equation

$$\Delta_0''(u) - \alpha(u)\Delta_0(u) + \gamma(u) = 0 \tag{2.4}$$

for $0 < u < 1$ subject to the boundary conditions $\Delta_0(0) = \Delta_0(1) = 0$, where

$$\alpha(u) = -E\{\lambda_{uu}(u,Z,0)|U = u\} \tag{2.5}$$

$$\gamma(u) = E\{\lambda_{\theta u}(u,Z,0)|U = u\}. \tag{2.6}$$

Expectations here and in what follows are under $\theta = 0$. Equation (2.4) is equivalent to,

$$E\{\frac{\partial}{\partial u}v_{\Delta_0}(u,Z,0)|U = u\} = 0 \tag{2.7}$$

given that, as one would expect from $\int g(u,z,0)dH(z) \equiv 1$,

$$E(\lambda_u(u,Z,0)|U = u) = 0. \tag{2.8}$$

It is easy to see that A1 and B guarantee the validity of (2.3), (2.4) and (2.7), as well as the boundedness of Δ_0, Δ_0'', see Bickel (1986) and Klaassen (1992). Then $b_\theta^0(u) \equiv u + \theta\Delta_0(u)$ are least favorable. Let $\Lambda_n(s)$ be the log likelihood ratio of the ranks as defined by (1.1) . Note that

$$\Lambda_n(s) = \log E\{\exp L_n(sn^{-1/2})|\tilde{Z}\} \tag{2.9}$$

where

$$L_n(\theta) = \sum_{i=1}^{n} \log \frac{q_{(\theta, b_\theta^0)}}{q_{(0,b_0)}}(U_i, Z_i)$$

the log likelihood ratio of (U_i, Z_i) $i = 1, \ldots, n$ for $Q_{(\theta, b_\theta^0)}$ and again, expectation is under θ_0. We have noted earlier that the $\{Q_{(\theta, b_\theta^0)} : |\theta| < \epsilon\}$ family is LAN at $\theta = 0$, and, in fact,

$$L_n(sn^{-1/2}) = sn^{-1/2} \sum_{i=1}^{n} v_{\Delta_0}(U_i, Z_i, 0) - \frac{s^2}{2} E v_{\Delta_0}^2 (U_1, Z_1, 0) + o_p(1) \quad (2.10)$$

We now establish,

Theorem 1 *Under A1 and B,*

$$\Lambda_n(s) = L_n(sn^{-1/2}) + o_p(1)$$

which establishes our claim in the bounded case.
Key Lemma: *If $w(u,z)$ is twice differentiable in u and,*

(i) $w(U,Z) \in L_2(Q_{(0,b_0)})$, $Ew(U,Z) = 0$

(ii) $w_u(U,Z) \in L_2(Q_{(0,b_0)})$, $E\{w_u(u,Z)|U = u) = 0$ *for all* u.

(iii) $\sup_{u,z} |w_{uu}(u,z)| < \infty$

then, if the $Z_{(i)}$ are the concomitants of the order statistics $U_{(i)}$ as in the introduction,

$$\sum_{i=1}^{n} w(U_i, Z_i) = \sum_{i=1}^{n} w(\frac{i}{n+1}, Z_{(i)}) + O_P(1).$$

Proof. Write, expanding around $(U_{(1)}, \ldots, U_{(n)})$,

$$\begin{aligned} \sum_{i=1}^{n} w(U_i, Z_i) &= \sum_{i=1}^{n} w(U_{(i)}, Z_{(i)}) \qquad (2.11) \\ &= \sum_{i=1}^{n} w(\frac{i}{n+1}, Z_{(i)}) + \sum_{i=1}^{n} \{w_u(U_{(i)}, Z_{(i)})(U_{(i)} - \frac{i}{n+1}) \\ &\quad - \frac{1}{2} w_{uu}(U_{(i)}^*, Z_{(i)})(U_{(i)} - \frac{i}{n+1})^2\} \end{aligned}$$

where $|U_{(i)}^* - \frac{i}{n+1}| \leq |U_{(i)} - \frac{i}{n+1}|$ for all i. Now,

$$\begin{aligned} &E\{[\sum_{i=1}^{n} w_u(U_{(i)}, Z_{(i)})(U_{(i)} - \frac{i}{n+1})]^2 | U_{(1)}, \ldots, U_{(n)}\} \\ &= \sum_{i=1}^{n} E\{w_u^2(U,Z)|U = U_{(i)}\}(U_{(i)} - \frac{i}{n+1})^2 = O_P(1) \end{aligned}$$

Here we use (ii) for the first identity and also

$$\sum_{i=1}^{n} E\{w_u^2(U_i, Z_i)|U_i\} = O_P(n).$$

The third term in (2.11) is $O_P(1)$ by (iii). □

The theorem follows readily from the key lemma by appealing to Theorem 4 of Le Cam and Yang applied to the distinguished statistics $n^{-1/2}\sum_{i=1}^{n} v_{\Delta_0}(\frac{i}{n+1}, Z_{(i)}, 0)$. Alternatively, we can argue directly. Write

$$\begin{aligned}\Lambda_n(s) &= sn^{-1/2}\sum_{i=1}^{n} v_{\Delta_0}(\frac{i}{n+1}, Z_{(i)}, 0) - \frac{s^2}{2} E v_{\Delta_0}^2(U_1, Z_1, 0) \\ &\quad + \log E\{\frac{A_n}{B_n}|\tilde{Z}\}\end{aligned}$$

where

$$\begin{aligned}A_n &= \exp L_n(s), \\ B_n &= \exp\{sn^{-1/2}\sum_{i=1}^{n} v_{\Delta_0}(\frac{i}{n+1}, Z_{(i)}, 0) - \frac{s^2}{2} E v_{\Delta_0}^2(U_1, Z_1, 0)\}.\end{aligned}$$

Then

$$|E\{|\frac{A_n}{B_n} - 1|\tilde{Z}\}| \leq E\{\frac{A_n}{B_n} - 1|1(\frac{A_n}{B_n} \leq M)|\tilde{Z}\}$$

$$+\frac{1}{B_n} E\{A_n 1(\frac{A_n}{B_n} > M)|\tilde{Z}\} + P\{\frac{A_n}{B_n} > M|\tilde{Z}\}$$

But A_n is uniformly integrable by LAN and by the key lemma $B_n^{-1} = O_P(1)$ and $A_n/B_n = 1 + o_p(1)$. The theorem follows. □

1.3 LAN – the general case

To encompass the examples of Section 1 we need to replace condition B. We do so with

A2: $\lambda, \lambda_\theta, \lambda_u, \lambda_{\theta u}, \lambda_{uu}$ are uniformly bounded in (u, z, θ) for $\epsilon \leq u \leq 1-\epsilon$, $\theta \in V(\theta_0)$, all $\epsilon > 0$.

and

A3: $\lambda_\theta(U, Z, 0) \in L_2(Q_{0,b_0})$. The functions $\lambda_{uu}(U, Z, 0)$, $\lambda_{\theta u}(U, Z, 0)$, and $\lambda_\theta(U, Z, 0)\lambda_u(U, Z, 0)$ are all in $L_1(Q_{(0,b_0)})$. Further, γ and α given by (2.5) and (2.7) are continuous on $(0, 1)$ and satisfy: $\int_0^1 \alpha(t)t(1-t)\,dt < \infty$ and $\sup_{t\in(0,1)} t^{3/2}(1-t)^{3/2}|\gamma(t)| < \infty$.

It follows (see below) that (2.4) has a unique solution Δ_0 which is bounded and differentiable. Therefore there exists $\epsilon > 0$ such that b_θ given by (2.2) is a transformation for $|\theta| < \epsilon$. We require:

A4: $v_{\Delta_0}(U, Z, 0) \in L_2(Q_{(0,b_0)})$ and the family $\{Q_{(\theta,b_\theta)} : |\theta| < \epsilon\}$ is regular (LAN) at $\theta = 0$. That is, $L_n(sn^{-1/2})$ obeys (2.10).

Finally, we require

A5: $\lambda_u(U, Z, 0) \in L_1(Q_{0,b_0})$ and (2.8) holds.

Note: Klaassen (1992) shows that A1-A5 hold for the Clayton-Cuzick and normal (generalized Box-Cox) transformation models. We prove

Theorem 2 *Under A1-A5 the conclusion of theorem 1 continues to hold.*

It is possible under the conditions of Klaassen (1992) to extend our direct argument. However, under the general conditions A1-A5 it is much easier to appeal to theorem 4 of Le Cam and Yang. By A3, A4 and (2.9) we can consider LAN for the ranks and covariates in the context of the parametric model $\{Q_{(\theta,b_\theta)}\}$. By the Le Cam-Yang theorem we need only exhibit statistics $T_n(\tilde{Z})$ such that

$$T_n(\tilde{Z}) = n^{-1/2} \sum_{i=1}^{n} v_{\Delta_0}(U_i, Z_i, 0) + o_p(1).$$

Let $s_m : [0, 1] \to [0, 1]$ such that $s_m \in C^\infty$ and

$$s_m(u) = \begin{cases} 1, & \frac{1}{m} \leq u \leq 1 - \frac{1}{m} \\ 0, & 0 \leq u \leq \frac{1}{2m}, 1 - \frac{1}{2m} \leq u \leq 1 \end{cases}$$

Consider the Sturm-Liouville equation on $[0, 1]$

$$\Delta''(u) - \alpha_m(u)\Delta(u) + \gamma_m(u) = 0 \tag{3.1}$$

subject to $\Delta(0) = \Delta(1) = 0$, where

$$\begin{aligned} \alpha_m(u) &= \alpha(u)s_m(u) \\ \gamma_m(u) &= \gamma(u)s_m(u) + E\{\lambda_\theta(u, Z)|U = u\}s'_m(u). \end{aligned}$$

As discussed in Bickel (1986), Klaassen (1992) the solution Δ_{om} to (3.1) is unique and solves uniquely the integral equation,

$$\Delta(u) + \int_0^1 K(u, s)\alpha_m(s)\Delta(s)ds - \int_0^1 K(u, s)\gamma_m(s)ds = 0 \tag{3.2}$$

where $K(u,s) = s \wedge u - su$. Note that $\alpha > 0$ and let

$$\psi_{om}(u) \equiv \sqrt{\alpha_m(u)}\Delta_{om}(u) \quad (3.3)$$

$$r_m(u) \equiv \sqrt{\alpha_m(u)}\int_0^1 K(u,v)\gamma_m(v)dv \quad (3.4)$$

Then, equivalently,

$$L_m\psi_{om} = r_m \quad (3.5)$$

where $L_m : L_2(0,1) \to L_2(0,1)$ is the operator $I + K_m$, I is the identity and K_m is the bounded self adjoint operator,

$$K_m(\chi)(u) = \int_0^1 \alpha_m^{1/2}(s)K(s,u)\alpha_m^{1/2}(u)\chi(s)ds.$$

The operators L_m have minimal eigenvalue ≥ 1 so that $\|L_m^{-1}\| \leq 1$.

Lemma 3 *If Δ_{om}, ψ_{om} are defined by (3.1), (3.5) and $\psi_0 \equiv \alpha^{1/2}\Delta_0$, then,*

$$\int (\psi_{om} - \psi_o)^2(u)du \to 0 \quad (3.6)$$

$$\sup_m \|\Delta_{om}\|_\infty < \infty \quad (3.7)$$

and, for every $\epsilon > 0$,

$$\sup\{|\Delta_{om}(u) - \Delta_o(u)| : \epsilon \leq u \leq 1-\epsilon\} \to 0 \quad (3.8)$$

$$\sup\{|\Delta'_{om}(u) - \Delta'_o(u)| : \epsilon \leq u \leq 1-\epsilon\} \to 0 \quad (3.9)$$

Proof. If $L \leftrightarrow \alpha$,

$$\|L_m - L\|^2 \quad (3.10)$$

$$\leq \int_0^1\int_0^1 K^2(s,u)((\alpha_m(s)\alpha_m(u))^{1/2} - (\alpha(s)\alpha(u))^{1/2})^2 dsdu \quad (3.11)$$

The integrand converges to 0 and

$$\int_0^1\int_0^1 K^2(s,u)|\alpha_m(s)||\alpha_m(u)|dsdu$$
$$\leq \int_0^1\int_0^1 K^2(s,u)|\alpha(s)||\alpha(u)|dsdu$$

for all m. So (3.6) follows by Vitali's theorem. Note that

$$\int_0^1 \int_0^1 K^2(s,u)|\alpha(s)|\alpha(u)|dsdu \le (\int_0^1 s(1-s)|\alpha(s)|ds)^2 < \infty.$$

Since $\|L_m^{-1}\| \le 1$,

$$\|L_m^{-1} - L^{-1}\| \to 0$$

Next, let $r_m(\cdot) = r_{1m}(\cdot) + r_{2m}(\cdot) + r_{3m}(\cdot)$ where

$$\begin{aligned}
r_{1m}(u) &= \alpha_m^{1/2}(u)\int_0^1 K(u,v)s_m(v)g(v)\ dv \\
r_{2m}(u) &= \alpha_m^{1/2}(u)\int\int\{(\lambda_{\theta u}(v,z,0)+\lambda_\theta\lambda_u(v,z,0)) \\
&\qquad K(u,v)(1-s_m(v))g(v,z,0)\ dv\ dH(z) \\
r_{3m}(u) &= \alpha_m^{1/2}(u)\int\int \lambda_\theta(v,z,0)(1(v\le u)-u)\}(1-s_m(v)) \\
&\qquad g(v,z,0)\ dv\ dH(z)
\end{aligned}$$

Now, by A3, $r_{1m} \to r$ a.e. where $r = \sqrt{\alpha(u)}\int_0^1 K(u,s)\gamma(s)\ \ ds$, and, $r_{2m}, r_{3m} \to 0$ a.e. Further, by A3

$$\begin{aligned}
|r_{1m}(u)| &\le \alpha^{1/2}(u)\int_0^1 K(u,v)|s_m(v)||\gamma(v)|\ dv & (3.12)\\
&\le C\sqrt{\alpha(u)}\left(\int_0^u \frac{1-u}{v^{1/2}(1-v)^{3/2}}\ dv + \int_u^1 \frac{u}{(1-v)^{1/2}v^{3/2}}\ dv\right) \\
&= 4C(\alpha(u)u(1-u))^{1/2} & (3.13)
\end{aligned}$$

We conclude that $\int(r_{1m}-r)^2 \to 0$. Since $K(u,v) \le u(1-u)$, it follows from A3 that $\int r_{2m}^2 \to 0$. Finally, $|r_{3m}| < (\alpha(u)u(1-u))^{1/2}\|\lambda_\theta\|_2$ and we obtain that

$$\int (r_m - r)^2 \to 0. \qquad (3.14)$$

Finally, conclude by (3.11), (3.14),

$$L_m^{-1} r_m \to L^{-1} r$$

in L_2 and (3.6) follows. Next, (3.7) follows from (3.2), since then

$$\|\Delta_m\|_\infty \le \sup_u(\int_0^1 K^2(u,s)|\alpha_m(s)|ds)^{1/2}(\int_0^1 \psi_{om}^2(s)ds)^{1/2}$$

$$+\int_0^1 s(1-s)|\gamma_m(s)|ds$$

Finally, (3.8) and (3.9) follow from (3.6), (3.7) and (3.1) since

$$\begin{aligned}\sup\{|\Delta''_{om}(t)| : \epsilon \le t \le 1-\epsilon\} \quad &\le \quad \sup\{|\alpha(t)| : \epsilon \le t \le 1-\epsilon\}\|\Delta_{om}\|_\infty \\ &\quad + \sup\{|\gamma(t)| : \epsilon \le t \le 1-\epsilon\}.\end{aligned}$$

so that the families $\{\Delta'_m(t)\}$, $\{\Delta_m(\cdot)\}$ are uniformly bounded and equicontinuous on $[\epsilon, 1-\epsilon]$. □

Proof of Theorem 2: Let

$$\begin{aligned} h_m(u,z,0) \quad = \quad & \lambda_\theta(u,z,0)s_m(u) - \Delta'_{om}(u) - \lambda_u(u,z)s_m(u)\Delta_{om}(u) \\ & - E\lambda_\theta(U,Z,0)s_m(U) + E\Delta'_{om}(U) + E\lambda_u(U,Z)\Delta_{om}(U). \end{aligned}$$

By construction, for each m, h_m and all its derivatives with respect to u are bounded. Furthermore,

$$\begin{aligned} &\frac{\partial h_m}{\partial u}(u,z,0) \\ &= \quad \lambda_{\theta u}(u,z,0)s_m(u) + \lambda_\theta(u,z,0)s'_m(u) - \Delta''_{om}(u) \\ &\qquad - \lambda_{uu}(u,z)s_m(u)\Delta_{om}(u) - \lambda_u(u,z)(s'_m(u)\Delta_{om}(u) + s_m(u)\Delta'_{om}(u)). \end{aligned}$$

By (2.8) and (3.1),

$$E\{\frac{\partial h_m}{\partial u}(U,Z,0)|U=u\} = 0.$$

Hence the key lemma applies and

$$n^{-1/2}\sum_{i=1}^n h_m(U_i,Z_i,0) = n^{-1/2}\sum_{i=1}^n h_m(\frac{i}{n+1}, Z_{(i)}, 0) + o_p(1).$$

To complete the proof of the theorem, we need only show that for some sequence $m \to \infty$

$$n^{-1/2}\sum_{i=1}^n (h_{m_n}(U_i,Z_i,0) - v_{\Delta_0}(U_i,Z_i,0)) = o_p(1)$$

or equivalently that, as $m \to \infty$

$$\begin{aligned} E\Big(&\ell_\theta(U,Z,0)(1-s_m(U)) - (\Delta'_{om} - \Delta'_o)(U) \\ &- \lambda_u(U,Z,0)(s_m\Delta_{om}(U) - \Delta_o(U))\Big)^2 \to 0 \end{aligned} \tag{3.15}$$

Now by (2.3)

$$\begin{aligned} &E(\ell_\theta(U,Z,0) - \Delta'_0(U) - \lambda_u(U,Z,0)\Delta_0(U))^2 \\ &= \quad E\ell^2_\theta(U,Z,0) - E(\Delta'_0(U) + \lambda_u(U,Z,0)\Delta_o(U))^2. \end{aligned} \tag{3.16}$$

On the other hand, a tedious calculation using integration by parts shows that,

$$\begin{aligned} & E(\ell_\theta(U,Z,0)s_m(U) - \Delta'_{om}(U) - \lambda_u(U,Z,0)\Delta_{om}(U)s_m(U))^2 \qquad (3.17) \\ = \quad & E\ell_\theta^2(U,Z,0)s_m^2(U) - E(\Delta'_{om}(U) - \lambda_u(U,Z,0)\Delta_{om}(U)s_m(U))^2 \\ & - 2\{E\ell_\theta(U,Z,0)\ell_u(U,Z,0)s_m(U)(1-s_m(U)) \\ & - E\ell_u^2(U,Z,0)\Delta_m^2(U)s_m(U)(1-s_m(U))\} \end{aligned}$$

The last two terms in (3.17) tend to 0 by dominated convergence in view of (3.7) and (3.16). Finally

$$\Delta'_{om}(U) - \lambda_u(U,Z,0)s_m(U) \xrightarrow{P} \Delta'_0(U) - \lambda_u(U,Z,0)\Delta_0(U)$$

by (3.8) and (3.9) so that by Fatou's theorem and (3.16), (3.17),

$$\begin{aligned} & \underline{\lim}_m E(\ell_\theta(U,Z,0)s_m(U) - \Delta'_{om}(U) - \lambda_u(U,Z)s_m(U)\Delta_{om}(U))^2 \quad (3.18) \\ \leq \quad & E(\ell_\theta(U,Z,0) - \Delta'_0(U) - \lambda_u(U,Z)\Delta_0(U))^2. \end{aligned}$$

Then, (3.15) and the theorem follow from (3.18).