

Estimation with Overidentifying Inequality Moment Conditions – Technical Appendix

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide*
University of Pennsylvania, CEPR and NBER

August 27, 2008

(Not intended to appear in print)

*Correspondence: H.R. Moon: Department of Economics, KAP 300, University Park Campus University of Southern California, Los Angeles, CA 90089. Email: moonr@usc.edu. F. Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104. Email: schorf@ssc.upenn.edu.

1 Consistency

For a precise definition of the notation see the main text.

1.1 Assumptions

Assumption 1 (a) $X_i, i = 1, \dots, n$ are strictly stationary on a probability space $(\Omega_n, \mathcal{F}_n, P_n)$;

(b) Θ is an m -dimensional compact subset of \mathbb{R}^m , where $m \leq h_1$, $\theta_{n,0} \rightarrow \theta_0$, $\theta_{n,0} \in \Theta \forall n$, and $\theta_0 \in \Theta$;

(c) $g(x, \theta)$ is continuous at each $\theta \in \Theta$ with probability one;

(d) $\mathbb{E}_n[g_1(X_i, \theta_{n,0})] = 0$, and $\inf_n \|\mathbb{E}_n[g_1(X_i, \theta)]\| > 0$ for $\theta \neq \theta_{n,0}$;

(e) $\nu_{n,0} \rightarrow \nu_0$ and $\sqrt{n}\nu_{n,0} \rightarrow u_0 \in [0, \infty]^{h_2}$;

(f) $\mathbb{E}_n[g(X_i, \theta_{n,0})g(X_i, \theta_{n,0})'] \rightarrow J$ is non-singular;

(g) $Z_n = O_p(1)$;

(h) $\mathbb{V} = \{\nu \in \mathbb{R}^{h_2} : \nu \geq 0 \text{ and } \|\nu\| \leq K\}$, $\nu_{n,0} \in \mathbb{V} \forall n$, and ν_0 lies in the interior of \mathbb{V} ;

(i) $\mathbb{E}_n \left[\sup_{\theta \in \Theta} \|g(X_i, \theta)\|^\alpha \right] \leq K < \infty$ for some $\alpha > 2$;

(j) for any θ and θ^* , $\|g(X_i, \theta) - g(X_i, \theta^*)\| \leq L(X_i)l(\|\theta - \theta^*\|)$, for some measurable function L of X_i such that $\sup_n \mathbb{E}_n(L(X_i)) < \infty$, and $l(y) \downarrow 0$ as $y \downarrow 0$.

1.2 Main Results

Theorem 1 Suppose that Assumption 1 is satisfied. Then $\hat{\theta}_n - \theta_{n,0} \xrightarrow{P} 0$ and $\hat{\nu}_n - \nu_{n,0} \xrightarrow{P} 0$. Moreover, $\hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n) \xrightarrow{P} 0$.

Throughout this appendix we are frequently using the following results. Notice that Assumption 1 implies that

$$\max_{1 \leq i \leq n} \sup_{\theta \in \Theta} \|g(X_i, \theta)\| = O_p(n^{1/\alpha}), \quad (1)$$

$$\frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} \|g(X_i, \theta)\|^\alpha = O_p(1), \quad (2)$$

$$\sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n \{g(X_i, \theta)g(X_i, \theta)' - \mathbb{E}[g(X_i, \theta)g(X_i, \theta)']\} \right\| = o_p(1). \quad (3)$$

According to Assumptions 1 and 2,

$$\mathbb{E}[g^{(1)}(X_i, \theta)] \text{ and } \mathbb{E}[g_j^{(2)}(X_i, \theta)] \quad (4)$$

are equicontinuous uniformly in θ , and

$$\begin{aligned} \sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n \{g^{(1)}(X_i, \theta) - \mathbb{E}[g^{(1)}(X_i, \theta)]\} \right\| &= o_p(1) \\ \sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n \{g_j^{(2)}(X_i, \theta) - \mathbb{E}[g_j^{(2)}(X_i, \theta)]\} \right\| &= o_p(1) \text{ for all } j = 1, \dots, h. \end{aligned} \quad (5)$$

(See, for instance, Andrews (1992)).

Proof of Theorem 1: We have to show that for any $\delta > 0$

$$\lim_{n \rightarrow \infty} P \left\{ \hat{\theta}_n \in \mathbb{B}(\theta_{n,0}, \delta), \hat{\nu}_n \in \mathbb{B}(\nu_{n,0}, \delta) \right\} = 1,$$

where

$$\mathbb{B}(\theta, \delta) = \{ \tilde{\theta} \in \Theta \mid \|\theta - \tilde{\theta}\| < \delta \}, \quad \mathbb{B}(\nu, \delta) = \{ \tilde{\nu} \in \mathbb{V} \mid \|\nu - \tilde{\nu}\| < \delta \}.$$

Define

$$\Theta_0^c = \Theta \cap \mathbb{B}(\theta_{n,0}, \delta)^c \quad \text{and} \quad N_0^c = \mathbb{V} \cap \mathbb{B}(\nu_{n,0}, \delta)^c.$$

To simplify the notation we omit the subscript n from the sets Θ_0^c and N_0^c . Recall that according to Assumption 1(i), the constant $\alpha > 2$ is such that $\mathbb{E}[\sup_{\theta \in \Theta} \|g(X_i, \theta)\|^\alpha] < K$. We show the following two statements are true: (i) For a given $\varepsilon, \delta > 0$ and ζ such that $\frac{1}{\alpha} < \zeta < \frac{1}{2}$, there exist positive constants η and κ and \bar{n} such that for $n \geq \bar{n}$

$$P \left\{ \bar{G}_n^*(\theta_{n,0}, \nu_{n,0}) \geq n^{-\zeta - \kappa} \eta \right\} < \frac{\varepsilon}{2}, \quad (6)$$

where

$$\bar{G}_n^*(\theta_{n,0}, \nu_{n,0}) = \max_{\lambda \in \tilde{\Lambda}_n(\theta_{n,0})} G_n^*(\theta_{n,0}, \nu_{n,0}, \lambda),$$

and (ii)

$$P \left\{ \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \bar{G}_n^*(\theta, \nu) \leq n^{-\zeta} \eta \right\} < \frac{\varepsilon}{2}. \quad (7)$$

Then, from (6) and (7) we deduce that there exists an $\eta > 0$ such that for $n \geq \bar{n}$:

$$\begin{aligned} & P \left\{ \hat{\theta}_n \in \mathbb{B}(\theta_{n,0}, \delta), \hat{\nu}_n \in \mathbb{B}(\nu_{n,0}, \delta) \right\} \\ & \geq P \left\{ \bar{G}_n^*(\theta_{n,0}, \nu_{n,0}) < n^{-\zeta - \kappa} \eta, \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \bar{G}_n^*(\theta, \nu) > n^{-\zeta} \eta \right\} \geq 1 - \varepsilon. \end{aligned}$$

Proof of (i): By Lemma 2 $\bar{G}_n^*(\theta_{n,0}, \nu_{n,0}) \leq O_p(1/n)$. Choose $\kappa > 0$ such that $\zeta + \kappa < 1$.

Then

$$n^{\zeta + \kappa} \bar{G}_n^*(\theta_{n,0}, \nu_{n,0}) \leq O_p(n^{\zeta + \kappa - 1}) = o_p(1)$$

as required.

Proof of (ii): To obtain a lower bound for $\bar{G}_n^*(\theta, \nu)$ we will evaluate the function $G_n^*(\theta, \nu, \lambda)$ at $\lambda = n^{-\zeta} u(\theta, \nu)$, where the function $u(\theta, \nu)$ is defined as

$$u(\theta, \nu) = \begin{cases} 0 & \text{if } \theta = \theta_{n,0}, \nu = \nu_{n,0} \\ \frac{\mathbb{E}[g(X_i, \theta)] - M' \nu}{\|\mathbb{E}[g(X_i, \theta)] - M' \nu\|} & \text{otherwise} \end{cases}$$

such that $\|u(\theta, \nu)\| \leq 1$.

Moreover, we truncate the function $g(x, \theta)$ as follows. Since $\alpha > 2$, we can choose a positive constant ξ such that

$$\frac{1}{\alpha^2} < \xi < \frac{1}{2\alpha}.$$

Let

$$\mathcal{X}_n = \left\{ x : \sup_{\theta \in \Theta} \|g(x, \theta)\| \leq n^\xi \right\} \quad \text{and} \quad g_n(x, \theta) = I \{x \in \mathcal{X}_n\} g(x, \theta).$$

We then replace the terms

$$\ln(1 + \lambda' g(x, \theta)) - \lambda' M' \nu$$

in the definition of the objective function $G_n^*(\theta, \nu, \lambda)$ by

$$q_n(x, \theta, \nu) = \ln(1 + n^{-\zeta} u'(\theta, \nu) g_n(x, \theta)) - n^{-\zeta} u'(\theta, \nu) M' \nu.$$

In what follows, we deduce the required result for (ii) by showing that

$$\text{(ii)-(a):} \quad \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \frac{1}{n} \sum_{i=1}^n q_n(X_i, \theta, \nu) \leq \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \bar{G}_n^*(\theta, \nu) + o_p(n^{-\zeta})$$

and

$$\text{(ii)-(b):} \quad P \left\{ \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \frac{1}{n} \sum_{i=1}^n q_n(X_i, \theta, \nu) < n^{-\zeta} \eta \right\} \leq \frac{\varepsilon}{2}.$$

Proof of (ii)-(a): Notice that $n^{-\zeta} u(\theta, \nu) \in \Lambda_n^\zeta \subset \cap_{\theta \in \Theta} \hat{\Lambda}_n(\theta)$ w.p.a.1 by Lemma 1. Then, by Lemma 4 and by the definition of $\hat{\lambda}_n(\theta, \nu)$,

$$\begin{aligned} & \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \frac{1}{n} \sum_{i=1}^n q_n(X_i, \theta, \nu) \\ &= \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \left[\frac{1}{n} \sum_{i=1}^n \ln(1 + n^{-\zeta} u'(\theta, \nu) g(X_i, \theta)) - n^{-\zeta} u'(\theta, \nu) M' \nu \right] + o_p(n^{-\zeta}) \\ &\leq \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \left[\frac{1}{n} \sum_{i=1}^n \ln(1 + \hat{\lambda}'_n(\theta, \nu) g(X_i, \theta)) - \hat{\lambda}'_n(\theta, \nu) M' \nu \right] + o_p(n^{-\zeta}) \\ &= \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \bar{G}_n^*(\theta, \nu) + o_p(n^{-\zeta}), \end{aligned}$$

as required.

Proof of (ii)-(b): A second-order Taylor expansion of q_n around $u(\theta, \nu) = 0$ yields

$$n^\zeta q_n(x, \theta, \nu) = u(\theta, \nu)' (g_n(x, \theta) - M' \nu) - \frac{1}{2} \frac{n^{-\zeta} u'(\theta, \nu) g_n(x, \theta) g_n(x, \theta)' u(\theta, \nu)}{(1 + n^{-\zeta} u'_*(\theta, \nu) g_n(x, \theta))^2}, \quad (8)$$

where $u_*(\theta, \nu)$ lies between zero and $u(\theta, \nu)$. The second-order term of the Taylor approximation (8) can be bounded as follows. For given x, θ , and ν

$$\sup_{\theta \in \Theta, \nu} \left| n^{-\zeta} u'_*(\theta, \nu) g_n(x, \theta) \right| \leq n^{-\zeta} \sup_{\theta \in \Theta} \|g_n(x, \theta)\| \leq n^{-\zeta + \xi} \leq n^{-\zeta/2}$$

since $\xi < \frac{1}{2\alpha} < \frac{\zeta}{2}$. Therefore,

$$\sup_{\theta \in \Theta, \nu} n^{-\zeta} \frac{u(\theta, \nu)' g_n(x, \theta) g_n(x, \theta)' u(\theta, \nu)}{(1 + n^{-\zeta} u'_*(\theta, \nu) g_n(x, \theta))^2} \leq \sup_{\theta \in \Theta, \nu} n^{-\zeta} \frac{\|g_n(x, \theta)\|^2 \|u(\theta, \nu)\|^2}{(1 - n^{-\zeta/2})^2} \leq n^{-\zeta + 2\xi} = o(1). \quad (9)$$

Now consider the expected value of $n^\zeta q_n(x, \theta, \nu)$. From (8), (9), and by the dominated convergence theorem, we have

$$\begin{aligned} n^\zeta \mathbb{E}[q_n(X_i, \theta, \nu)] &= u'(\theta, \nu)(\mathbb{E}[g_n(X_i, \theta)] - M'\nu) + o(1) \\ &= \begin{cases} o(1) & \text{if } \theta = \theta_0, \nu = \nu_{n,0} \\ \|\mathbb{E}[g(X_i, \theta)] - M'\nu\| + o(1) > 0 & \text{otherwise} \end{cases}. \end{aligned} \quad (10)$$

The $o(1)$ terms absorb the second-order term of the Taylor approximation and the discrepancy between $\mathbb{E}[g_n(X, \theta)]$ and $\mathbb{E}[g(X, \theta)]$, which vanishes as \mathcal{X}_n expands. From (10) and the monotone convergence theorem we can deduce that

$$\lim_{n \rightarrow \infty} n^\zeta \lim_{\delta \downarrow 0} \mathbb{E} \left[\inf_{\theta^* \in \mathbb{B}(\theta, \delta), \nu^* \in \mathbb{B}(\nu, \delta)} q_n(X_i, \theta^*, \nu^*) \right] \begin{cases} = 0 & \text{if } \theta = \theta_{n,0}, \nu = \nu_{n,0} \\ > 0 & \text{otherwise} \end{cases}.$$

Since Θ and \mathbb{V} are compact by assumption, the sets $\Theta \cap \mathbb{B}(\theta_{n,0}, \delta)^c$ and $\mathbb{V} \cap \mathbb{B}(\nu_{n,0}, \delta)^c$ are compact. We can cover both $\Theta \cap \mathbb{B}(\theta_{n,0}, \delta)^c$ and $\mathbb{V} \cap \mathbb{B}(\nu_{n,0}, \delta)^c$ with $\Theta_j = \mathbb{B}(\theta_j, \delta_j)$ and $N_j = \mathbb{B}(\nu_j, \delta_j)$'s, $j = 1, \dots, J$ taking each δ_j small enough such there exist η_j 's such that

$$n^\zeta \mathbb{E} \left[\inf_{\theta \in \Theta_j, \nu \in N_j} q_n(X_i, \theta, \nu) \right] \geq 2\eta_j, \quad n \geq n_j \quad (11)$$

for some positive numbers $\eta_j = \eta_j(\delta)$, $j = 1, \dots, J$. By the WLLN¹ and (11), for a given $\varepsilon > 0$, we can find \bar{n}_j 's such that $n \geq \bar{n}_j$ implies that

$$\begin{aligned} \frac{\varepsilon}{2J} &\geq P \left\{ \left| \frac{1}{n} \sum_{i=1}^n n^\zeta \inf_{\theta \in \Theta_j, \nu \in N_j} q_n(X_i, \theta, \nu) - \mathbb{E} \left[n^\zeta \inf_{\theta \in \Theta_j, \nu \in N_j} q_n(X_i, \theta, \nu) \right] \right| > \eta_j \right\} \\ &\geq P \left\{ \frac{1}{n} \sum_{i=1}^n \inf_{\theta \in \Theta_j, \nu \in N_j} q_n(X_i, \theta, \nu) < \mathbb{E} \left[\inf_{\theta \in \Theta_j, \nu \in N_j} q_n(X_i, \theta, \nu) \right] - n^{-\zeta} \eta_j \right\} \\ &\geq P \left\{ \frac{1}{n} \sum_{i=1}^n \inf_{\theta \in \Theta_j, \nu \in N_j} q_n(X_i, \theta, \nu) < n^{-\zeta} \eta_j \right\} \\ &\geq P \left\{ \inf_{\theta \in \Theta_j, \nu \in N_j} \frac{1}{n} \sum_{i=1}^n q_n(X_i, \theta, \nu) < n^{-\zeta} \eta_j \right\} \end{aligned}$$

for $j = 1, \dots, J$. Now let letting $\eta = \min \{\eta_1, \dots, \eta_J\}$ and $\bar{n} = \max_{j=1, \dots, J} \bar{n}_j$, we have for $n \geq \bar{n}$

$$\begin{aligned} &P \left\{ \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \frac{1}{n} \sum_{i=1}^n q_n(X_i, \theta, \nu) < n^{-\zeta} \eta \right\} \\ &\leq P \left\{ \min_{j=1, \dots, J} \left\{ \inf_{\theta \in \Theta_j, \nu \in N_j} \frac{1}{n} \sum_{i=1}^n q_n(X_i, \theta, \nu) \right\} < n^{-\zeta} \eta \right\} \\ &\leq \sum_{j=1}^J P \left\{ \inf_{\theta \in \Theta_j, \nu \in N_j} \frac{1}{n} \sum_{i=1}^n q_n(X_i, \theta, \nu) < n^{-\zeta} \eta_j \right\} \leq \frac{\varepsilon}{2}, \end{aligned}$$

¹Notice that

$$\mathbb{E} \left[\left(n^\zeta \inf_{\theta \in \Theta_j, \nu \in N_j} q_n(X_i, \theta, \nu) \right)^2 \right] \leq \mathbb{E} \left[\sup_{\theta \in \Theta} 2 \|g(X_i, \theta)\|^2 \right] + 2K + n^{-2\zeta + 4\varepsilon} < \infty. \quad (12)$$

as required part (ii)-(b).

Combining (ii)-(a) and (ii)-(b) we have

$$P \left\{ \min_{\theta \in \Theta_0^c, \nu \in N_0^c} \bar{G}_n^*(\theta, \nu) < n^{-\zeta} \eta \right\} \leq \frac{\varepsilon}{2},$$

as required for (ii).

Since $\hat{\theta}_n - \theta_{n,0} \xrightarrow{p} 0$ and $\hat{\nu}_n - \nu_{n,0} \xrightarrow{p} 0$ we can deduce from Lemmas 2 and 3 that $\hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n) \xrightarrow{p} 0$. ■

1.3 Technical Lemmas

Lemma 1 *Suppose that Assumption 1 is satisfied. Then,*

$$(i) \quad \sup_{\theta \in \Theta, \lambda \in \Lambda_n^\zeta, 1 \leq i \leq n} |\lambda' g(X_i, \theta)| \xrightarrow{p} 0,$$

$$(ii) \quad \Lambda_n^\zeta \subseteq \bigcap_{\theta \in \Theta} \hat{\Lambda}_n(\theta) \text{ w.p.a. } 1.$$

Proof of Lemma 1: See proof of Lemma A1 in Newey and Smith (2004). ■

Lemma 2 *Suppose that Assumption 1 is satisfied. Let $\bar{\theta} \in \Theta$ and $\bar{\nu} \geq 0$ be sequences such that $\bar{\theta} - \theta_{n,0} \xrightarrow{p} 0$, and $\bar{\nu} - \nu_{n,0} \xrightarrow{p} 0$. Moreover, $\frac{1}{\sqrt{n}} \sum_{i=1}^n g_1(X_i, \bar{\theta}) = O_p(1)$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^n (g_2(X_i, \bar{\theta}) - \bar{\nu}) = O_p(1)$. Then,*

$$(i) \quad \hat{\lambda}(\bar{\theta}, \bar{\nu}) \text{ exists w.p.a. } 1,$$

$$(ii) \quad \hat{\lambda}(\bar{\theta}, \bar{\nu}) = O_p(n^{-1/2}),$$

$$(iii) \quad G_n^*(\bar{\theta}, \bar{\nu}, \hat{\lambda}(\bar{\theta}, \bar{\nu})) \leq O_p\left(\frac{1}{n}\right).$$

Proof of Lemma 2: The proof is similar to that of Lemma A2 in Newey and Smith (2004).

Proof of (i): Define

$$\tilde{\lambda}(\bar{\theta}, \bar{\nu}) = \arg \max_{\lambda \in \Lambda_n^\zeta} G_n^*(\bar{\theta}, \bar{\nu}, \lambda)$$

Since Λ_n^ζ is compact and $\ln(1 + \lambda' g(X_i, \bar{\theta})) - \bar{\nu}' M \lambda$ is continuous and strictly concave in λ the optimal solution $\tilde{\lambda}(\bar{\theta}, \bar{\nu})$ exists and is unique. Statement (i) then follows from Lemma 1.

Proof of (ii) and (iii): Write $\bar{g}_i = g(X_i, \bar{\theta})$. For some constant C

$$\begin{aligned}
0 = G_n^*(\bar{\theta}, \bar{\nu}, 0) &\leq G_n^*(\bar{\theta}, \bar{\nu}, \tilde{\lambda}(\bar{\theta}, \bar{\nu})) \\
&= \frac{1}{n} \sum_{i=1}^n \ln \left(1 + \tilde{\lambda}'(\bar{\theta}, \bar{\nu}) \bar{g}_i \right) - \bar{\nu}' M \tilde{\lambda}(\bar{\theta}, \bar{\nu}) \\
&= \tilde{\lambda}'(\bar{\theta}, \bar{\nu}) \left(\frac{1}{n} \sum_{i=1}^n \bar{g}_i - M' \bar{\nu} \right) - \frac{1}{2} \tilde{\lambda}''(\bar{\theta}, \bar{\nu}) \left(\frac{1}{n} \sum_{i=1}^n \frac{\bar{g}_i \bar{g}_i'}{(1 + \lambda'_* \bar{g}_i)^2} \right) \tilde{\lambda}(\bar{\theta}, \bar{\nu}) \\
&\leq \tilde{\lambda}'(\bar{\theta}, \bar{\nu}) \left(\frac{1}{n} \sum_{i=1}^n \bar{g}_i - M' \bar{\nu} \right) - \frac{C}{4} \tilde{\lambda}'(\bar{\theta}, \bar{\nu}) \tilde{\lambda}(\bar{\theta}, \bar{\nu}),
\end{aligned}$$

where λ_* lies on the line joining $\tilde{\lambda}(\bar{\theta}, \bar{\nu})$ and 0. The last inequality holds because

$$\max_{1 \leq i \leq n} |\lambda'_* \bar{g}_i| = o_p(1)$$

according to Lemma 1 and $\frac{1}{n} \sum_{i=1}^n \bar{g}_i \bar{g}_i'$ converges in probability to J , a positive definite matrix, by (3) and Assumption 1(f). The remainder of the proof follows the proof of Lemma A2 in Newey and Smith (2004). ■

Lemma 3 *Suppose Assumption 1 is satisfied. Then,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left[g(X_i, \hat{\theta}) - M' \hat{\nu} \right] = O_p(1).$$

Proof of Lemma 3: The proof is similar to that of Lemma A.3 in Newey and Smith (2004).

Let $\hat{g}_i = g(X_i, \hat{\theta}) - M' \hat{\nu}$ and $\hat{g} = \frac{1}{n} \sum_{i=1}^n \left[g(X_i, \hat{\theta}) - M' \hat{\nu} \right]$. Define $\hat{u}(\hat{\theta}, \hat{\nu}) = n^{-\zeta} \frac{\hat{g}}{\|\hat{g}\|}$. (Recall the definition of $u(\theta, \nu)$ in the proof of consistency.) Approximation $G_n^*(\theta, \nu, \lambda)$ with respect to λ around $\lambda = 0$ at $(\theta, \nu, \lambda) = (\hat{\theta}, \hat{\nu}, \hat{u}(\hat{\theta}, \hat{\nu}))$. Then,

$$\begin{aligned}
&G_n^*(\hat{\theta}, \hat{\nu}, \hat{u}(\hat{\theta}, \hat{\nu})) \\
&= G_n^*(\hat{\theta}, \hat{\nu}, 0) + \frac{\partial G_n^*(\hat{\theta}, \hat{\nu}, 0)}{\partial \lambda'} \hat{u}(\hat{\theta}, \hat{\nu}) + \frac{1}{2} \hat{u}'(\hat{\theta}, \hat{\nu}) \frac{\partial^2 G_n^*(\hat{\theta}, \hat{\nu}, \tilde{\lambda})}{\partial \lambda \partial \lambda'} \hat{u}(\hat{\theta}, \hat{\nu}) \\
&= \hat{g}' \hat{u}(\hat{\theta}, \hat{\nu}) - \frac{1}{2} \hat{u}'(\hat{\theta}, \hat{\nu}) \left(\frac{1}{n} \sum_{i=1}^n \frac{\hat{g}_i \hat{g}_i'}{(1 + \tilde{\lambda}' \hat{g}_i)^2} \right) \hat{u}(\hat{\theta}, \hat{\nu}),
\end{aligned}$$

where $\tilde{\lambda}$ is located between 0 and $\hat{u}(\hat{\theta}, \hat{\nu})$.

Notice that $\max_{1 \leq i \leq n} |\hat{u}'(\hat{\theta}, \hat{\nu}) \hat{g}_i| \rightarrow_p 0$ and $\hat{u}(\hat{\theta}, \hat{\nu}) \in \hat{\Lambda}_n(\hat{\theta})$ by Lemma A.1 w.p.a.1. Also, under Assumption 1 $\left\| \frac{1}{n} \sum_{i=1}^n \hat{g}_i \hat{g}_i' \right\| \leq 2 \left(\frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} \|g(X_i, \theta)\|^2 + K \right) = O_p(1)$.

Then, w.p.a.1, for some constant C ,

$$\begin{aligned}
& \hat{g}'\hat{u}(\hat{\theta}, \hat{\nu}) - \frac{1}{2}\hat{u}'(\hat{\theta}, \hat{\nu}) \left(\frac{1}{n} \sum_{i=1}^n \frac{\hat{g}_i \hat{g}'_i}{(1 + \ddot{\lambda}' \hat{g}_i)^2} \right) \hat{u}(\hat{\theta}, \hat{\nu}) \\
&= n^{-\zeta} \|\hat{g}\| - \frac{1}{2}\hat{u}'(\hat{\theta}, \hat{\nu}) \left(\frac{1}{n} \sum_{i=1}^n \frac{\hat{g}_i \hat{g}'_i}{(1 + \ddot{\lambda}' \hat{g}_i)^2} \right) \hat{u}(\hat{\theta}, \hat{\nu}) \\
&\geq n^{-\zeta} \|\hat{g}\| - \frac{1}{2} \max_{1 \leq i \leq n} \left(\frac{1}{(1 + \ddot{\lambda}' \hat{g}_i)^2} \right) \hat{u}'(\hat{\theta}, \hat{\nu}) \left(\frac{1}{n} \sum_{i=1}^n \hat{g}_i \hat{g}'_i \right) \hat{u}(\hat{\theta}, \hat{\nu}) \\
&\geq n^{-\zeta} \|\hat{g}\| - Cn^{-2\zeta}. \tag{13}
\end{aligned}$$

Then,

$$n^{-\zeta} \|\hat{g}\| - Cn^{-2\zeta} \leq G_n^*(\hat{\theta}, \hat{\nu}, \hat{u}(\hat{\theta}, \hat{\nu})) \leq G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}) \leq \sup_{\lambda \in \bar{\Lambda}_n(\theta_{n,0})} G_n^*(\theta_{n,0}, \nu_{n,0}, \lambda) \leq O_p\left(\frac{1}{n}\right), \tag{14}$$

where the first inequality is from (13), the second and third inequalities hold because $(\hat{\theta}, \hat{\nu}, \hat{\lambda})$ is a saddle point, and the last inequality is from Lemma A.2 with

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n [g(X_i, \theta_{n,0}) - M' \nu_{n,0}] = O_p(1)$$

by Assumption 1(g). Also, by $\zeta < \frac{1}{2}$, $\zeta - 1 < -\frac{1}{2} < -\zeta$. Solving (14) for $\|\hat{g}\|$ gives

$$\|\hat{g}\| \leq O_p(n^{-\zeta}). \tag{15}$$

For a given sequence $\varepsilon_n \rightarrow 0$, let $\bar{\lambda} = \varepsilon_n \hat{g}$. According to (15) $\bar{\lambda} = o_p(n^{-\zeta})$. Hence, $\bar{\lambda} \in \Lambda_n^\zeta$ w.p.a.1. Then, as in (14), we have

$$\bar{\lambda}' \hat{g} - C \|\bar{\lambda}\|^2 = \varepsilon_n \|\hat{g}\|^2 - C\varepsilon_n^2 \|\hat{g}\|^2 \leq \varepsilon_n \|\hat{g}\|^2 (1 - C\varepsilon_n) \leq O_p\left(\frac{1}{n}\right).$$

For large enough n the term $1 - C\varepsilon_n$ is bounded away from zero and it follows that $\varepsilon_n \|\hat{g}\|^2 = O_p\left(\frac{1}{n}\right)$. Since ε_n is an arbitrary sequence that tends to zero, we deduce that

$$\|\hat{g}\| = O_p\left(\frac{1}{\sqrt{n}}\right),$$

as required. ■

Lemma 4 Suppose that Assumption 1 is satisfied. Let $g_n(x, \theta) = I\{x \in \mathcal{X}_n\}g(x, \theta)$ where

$$\mathcal{X}_n = \left\{ x : \sup_{\theta \in \Theta} \|g(x, \theta)\| \leq n^\xi \right\},$$

where $\frac{1}{\alpha^2} < \xi < \frac{1}{2\alpha}$ and $\alpha > 2$ as in Assumption 1(i). Define

$$\begin{aligned} q_n(X_i, \theta, \nu) &= \ln [1 + n^{-\zeta} u'(\theta, \nu) g_n(X_i, \theta)] - n^{-\zeta} u'(\theta, \nu) M' \nu \\ \tilde{q}_n(X_i, \theta, \nu) &= \ln [1 + n^{-\zeta} u'(\theta, \nu) g(X_i, \theta)] - n^{-\zeta} u'(\theta, \nu) M' \nu \end{aligned}$$

and assume that $\|u(\theta, \nu)\| \leq 1$. Then,

$$\sup_{\theta \in \Theta, \nu \geq 0} \left| \frac{1}{n} \sum_{i=1}^n \left(q_n(X_i, \theta, \nu) - \tilde{q}_n(X_i, \theta, \nu) \right) \right| = o_p(n^{-\zeta}).$$

Proof of Lemma 4: By the mean value theorem,

$$\begin{aligned} & \sup_{\theta \in \Theta, \nu \geq 0} \left| \frac{1}{n} \sum_{i=1}^n \{q_n(X_i, \theta, \nu) - \tilde{q}_n(X_i, \theta, \nu)\} \right| \\ &= \sup_{\theta \in \Theta, \nu \geq 0} \left| \frac{1}{n} \sum_{i=1}^n \left(\frac{n^{-\zeta} u'(\theta, \nu) g(X_i, \theta)}{1 + n^{-\zeta} u'_*(\theta, \nu) g(X_i, \theta)} \right) I\{X_i \notin \mathcal{X}_n\} \right| \quad (16) \\ &\leq \max_{1 \leq i \leq n} \sup_{\theta \in \Theta, \nu \geq 0} \left| \frac{n^{-\zeta} u'(\theta, \nu) g(X_i, \theta)}{1 + n^{-\zeta} u'_*(\theta, \nu) g(X_i, \theta)} \right| \frac{1}{n} \sum_{i=1}^n I\left\{ \sup_{\theta \in \Theta} \|g(X_i, \theta)\| > n^\xi \right\} \\ &\leq \frac{1}{n^{\alpha\xi}} \left(\max_{1 \leq i \leq n} \sup_{\theta \in \Theta, \nu \geq 0} \left| \frac{n^{-\zeta} u'(\theta, \nu) g(X_i, \theta)}{1 + n^{-\zeta} u'_*(\theta, \nu) g(X_i, \theta)} \right| \right) \left(\frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} \|g(X_i, \theta)\|^\alpha \right) \end{aligned}$$

where $u_*(\theta, \nu)$ is located between 0 and $u(\theta, \nu)$. The second term on the right-hand side of (16) can be bounded as follows. According to (1)

$$n^{-\zeta} \max_{1 \leq i \leq n} \sup_{\theta \in \Theta} \|g(X_i, \theta)\| = n^{-\zeta+1/\alpha} O_p(1).$$

Moreover, $\|u(\theta, \nu)\| \leq 1$. Therefore,

$$\begin{aligned} \max_{1 \leq i \leq n} \sup_{\theta \in \Theta, \nu \geq 0} \left| \frac{n^{-\zeta} u'(\theta, \nu) g(X_i, \theta)}{1 + n^{-\zeta} u'_*(\theta, \nu) g(X_i, \theta)} \right| &\leq \frac{2n^{-\zeta} \max_{1 \leq i \leq n} \sup_{\theta \in \Theta} \|g(X_i, \theta)\|}{1 - 2n^{-\zeta} \max_{1 \leq i \leq n} \sup_{\theta \in \Theta} \|g(X_i, \theta)\|} \\ &= \frac{n^{-\zeta+1/\alpha} O_p(1)}{1 - n^{-\zeta+1/\alpha} O_p(1)} = n^{-\zeta+1/\alpha} O_p(1). \end{aligned}$$

By Assumption 1(i) and the Markov inequality, the third term on the right-hand side of (16) is $O_p(1)$. Since $\frac{1}{\alpha^2} < \xi < \frac{1}{2\alpha}$, we are able to deduce that

$$n^\zeta \sup_{\theta \in \Theta, \nu \geq 0} \left| \frac{1}{n} \sum_{i=1}^n \left(q_n(X_i, \theta, \nu) - \tilde{q}_n(X_i, \theta, \nu) \right) \right| = n^{-\alpha\xi + \frac{1}{\alpha}} O_p(1) = o_p(1),$$

as required. ■

2 Limit Distributions

Let $\beta = [\theta', \nu', \lambda']'$, $\beta_{n,0} = [\theta'_{n,0}, \nu'_{n,0}, 0_{1 \times h}]'$, and abbreviate $G_n^*(\theta, \nu, \lambda)$ as $G_n^*(\beta)$. The objective function is expanded around $\beta_{n,0}$ as follows:

$$G_n^*(\beta) = G_{nq}^*(\beta) + \frac{1}{n} \mathcal{R}_n(\beta), \quad (17)$$

where

$$G_{nq}^*(\beta) = G_n^*(\beta_{n,0}) + G_n^{*(1)}(\beta_{n,0})(\beta - \beta_{n,0}) + \frac{1}{2}(\beta - \beta_{n,0})' G_n^{*(2)}(\beta_{n,0})(\beta - \beta_{n,0}).$$

We begin by deriving the coefficient matrices for the quadratic approximation of the objective function

$$G_{nq}^*(\beta) = G_n^*(\beta_{n,0}) + G_n^{*(1)}(\beta_{n,0})'(\beta - \beta_{n,0}) + \frac{1}{2}(\beta - \beta_{n,0})' G_n^{*(2)}(\beta_{n,0})(\beta - \beta_{n,0}). \quad (18)$$

A direct calculation shows that

$$G_n^{*(1)}(\beta) = \left[G_n^{*(1)}(\beta)'_{\theta}, G_n^{*(1)}(\beta)'_{\nu}, G_n^{*(1)}(\beta)'_{\lambda} \right]', \quad (19)$$

where

$$\begin{aligned} G_n^{*(1)}(\beta)_{\theta} &= \frac{1}{n} \sum_{i=1}^n \left(\frac{g^{(1)}(X_i, \theta) \lambda}{1 + \lambda' g(X_i, \theta)} \right), \\ G_n^{*(1)}(\beta)_{\nu} &= -M\lambda, \\ G_n^{*(1)}(\beta)_{\lambda} &= \frac{1}{n} \sum_{i=1}^n \left(\frac{g(X_i, \theta)}{1 + \lambda' g(X_i, \theta)} \right) - M'v. \end{aligned}$$

At $\beta_{n,0}$ the first derivatives simplify to

$$G_n^{*(1)}(\beta_{n,0}) = [0, 0, n^{-1/2} Z_n']. \quad (20)$$

We proceed by partitioning the matrix of second derivative as follows

$$G_n^{*(2)}(\beta) = \begin{pmatrix} G_n^{*(2)}(\beta)_{\theta\theta'} & G_n^{*(2)}(\beta)_{\theta\nu'} & G_n^{*(2)}(\beta)_{\theta\lambda'} \\ G_n^{*(2)}(\beta)_{\nu\theta'} & G_n^{*(2)}(\beta)_{\nu\nu'} & G_n^{*(2)}(\beta)_{\nu\lambda'} \\ G_n^{*(2)}(\beta)_{\lambda\theta'} & G_n^{*(2)}(\beta)_{\lambda\nu'} & G_n^{*(2)}(\beta)_{\lambda\lambda'} \end{pmatrix}, \quad (21)$$

where

$$\begin{aligned} G_n^{*(2)}(\beta)_{\theta\theta'} &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\sum_{j=1}^h \lambda_j g_j^{(2)}(X_i, \theta)}{1 + \lambda' g(X_i, \theta)} - \frac{g^{(1)}(X_i, \theta) \lambda \lambda' g^{(1)}(X_i, \theta)'}{(1 + \lambda' g(X_i, \theta))^2} \right), \\ G_n^{*(2)}(\beta)_{\theta\nu'} &= 0, \quad G_n^{*(2)}(\beta)_{\nu\nu'} = 0, \quad G_n^{*(2)}(\beta)_{\lambda\nu'} = -M', \\ G_n^{*(2)}(\beta)_{\lambda\theta'} &= \frac{1}{n} \sum_{i=1}^n \left(\frac{g^{(1)}(X_i, \theta)'}{1 + \lambda' g(X_i, \theta)} - \frac{g(X_i, \theta) \lambda' g^{(1)}(X_i, \theta)'}{(1 + \lambda' g(X_i, \theta))^2} \right), \\ G_n^{*(2)}(\beta)_{\lambda\lambda'} &= -\frac{1}{n} \sum_{i=1}^n \frac{g(X_i, \theta) g(X_i, \theta)'}{(1 + \lambda' g(X_i, \theta))^2}. \end{aligned}$$

At $\beta_{n,0}$ the second derivatives simplify to

$$G_n^{*(2)}(\beta_{n,0}) = \begin{bmatrix} 0 & 0 & Q_n \\ 0 & 0 & -M \\ Q_n' & -M' & -J_n \end{bmatrix}. \quad (22)$$

The objective function $G_{nq}^*(\beta)$ in terms of the transformed parameters is:

$$\begin{aligned} \mathcal{G}_{nq}^*(\phi, l) &= -\frac{1}{2}(l - J_n^{-1}[Z_n - R'_n(\phi - \phi_{n,0})])' J_n (l - J_n^{-1}[Z_n - R'_n(\phi - \phi_{n,0})]) \\ &\quad + \frac{1}{2}(Z_n - R'_n(\phi - \phi_{n,0}))' J_n^{-1} (Z_n - R'_n(\phi - \phi_{n,0})) \end{aligned} \quad (23)$$

The function $\mathcal{G}_{nq}^*(\phi, l)$ is maximized with respect to $l \in \mathbb{R}^h$ by

$$\tilde{l}_q(\phi) = J_n^{-1} (Z_n - R'_n(\phi - \phi_{n,0})) \quad (24)$$

and the concentrated objective function is:

$$\bar{\mathcal{G}}_{nq}^*(\phi) = \mathcal{G}_{nq}^*(\phi, \tilde{l}_q(\phi)) = \frac{1}{2}(Z_n - R'_n(\phi - \phi_{n,0}))' J_n^{-1} (Z_n - R'_n(\phi - \phi_{n,0})). \quad (25)$$

2.1 Assumptions

Assumption 2 (a) The true parameter θ_0 exists in an interior of Θ ;

(b) $g(X_i, \theta)$ is twice continuously differentiable with respect to θ ;

(c) the minimum eigenvalue of $(\mathbb{E}_n[g_1^{(1)}(X_i, \theta)])(\mathbb{E}_n[g_1^{(1)}(X_i, \theta)])'$ is bounded below by a constant $K > 0$;

(d) $\mathbb{E}_n[\sup_{\theta \in \Theta} \|g^{(1)}(X_i, \theta)\|^2] \leq K < \infty$, $\mathbb{E}_n[\sup_{\theta \in \Theta} \|g_j^{(2)}(X_i, \theta)\|] \leq K < \infty$ for $j = 1, \dots, h$;

(e) for any θ and θ^* , $\|g_j^{(2)}(X_i, \theta) - g_j^{(2)}(X_i, \theta^*)\| \leq L_j(X_i) l_j(\|\theta - \theta^*\|)$, for some measurable function L_j of X_i such that $\sup_n \mathbb{E}_n(L_j(X_i)) < \infty$, and $l_j(y) \downarrow 0$ as $y \downarrow 0$.

Assumption 3 (a) For each θ , $Q_n(\theta) \xrightarrow{p} Q(\theta)$ and $J_n(\theta) \xrightarrow{p} J(\theta)$. (b) For each θ , $\frac{1}{n} \sum_{i=1}^n g_j^{(2)}(X_i, \theta) \xrightarrow{p} (\lim_{n \rightarrow \infty} \mathbb{E}_n[g_j^{(2)}(X_i, \theta)])$. (c) $Z_n \implies Z$, where $Z \sim \mathcal{N}(0, J - M' \nu_0 \nu_0' M)$.

2.2 Negligible Remainder

Lemma 5 Suppose Assumptions 1 to 2 are satisfied, then for all $\gamma_n \rightarrow 0$

$$\sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \frac{|\mathcal{R}_n(\beta)|}{(1 + \|\sqrt{n}(\beta - \beta_{n,0})\|^2)} = o_p(1), \quad (26)$$

where $\mathcal{R}_n(\beta)$ is the remainder term in (17).

Proof of Lemma 5: By Lemma 1(a) of Andrews (1999), it is sufficient to prove

$$\sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| G_n^{*(2)}(\beta) - G_n^{*(2)}(\beta_{n,0}) \right\| = o_p(1),$$

for every sequence $\gamma_n \rightarrow 0$. $G_n^{*(2)}$ is defined in (21). To verify this sufficient condition we will subsequently show that

- (i) $\sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| G_n^{*(2)}(\beta)_{\theta\theta'} - G_n^{*(2)}(\beta_{n,0})_{\theta\theta'} \right\| = o_p(1),$
- (ii) $\sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| G_n^{*(2)}(\beta)_{\lambda\theta'} - G_n^{*(2)}(\beta_{n,0})_{\lambda\theta'} \right\| = o_p(1),$
- (iii) $\sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| G_n^{*(2)}(\beta)_{\lambda\lambda'} - G_n^{*(2)}(\beta_{n,0})_{\lambda\lambda'} \right\| = o_p(1).$

We begin by showing that

$$\sup_{\beta \in \mathcal{B}_n} \left| \frac{1}{1 + \lambda'g(X_i, \theta)} \right| = O_p(1). \quad (27)$$

For any given $0 < \delta < \frac{1}{2}$, set $K = \frac{1}{1-\delta}$. Then, since $\sup_{1 \leq i \leq n, \beta \in \mathcal{B}_n} |\lambda'g(X_i, \theta)| \leq \delta$ implies $\sup_{1 \leq i \leq n, \beta \in \mathcal{B}_n} \left| \frac{1}{1 + \lambda'g(X_i, \theta)} \right| \leq K$,

$$P \left\{ \sup_{1 \leq i \leq n, \beta \in \mathcal{B}_n} \left| \frac{1}{1 + \lambda'g(X_i, \theta)} \right| > K \right\} \leq P \left\{ \sup_{1 \leq i \leq n, \beta \in \mathcal{B}_n} |\lambda'g(X_i, \theta)| > \delta \right\} \rightarrow 0,$$

which proves (27). The convergence result for the upper bound can be deduced from Lemma 1.

(i) Notice that

$$\begin{aligned} & \sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| \frac{1}{n} \sum_{i=1}^n \left(\frac{\lambda_j g_j^{(2)}(X_i, \theta)}{1 + \lambda'g(X_i, \theta)} \right) \right\| \\ & \leq \sup_{\lambda \in \Lambda_n^\zeta} |\lambda_j| \left(\sup_{\beta \in \mathcal{B}_n, 1 \leq i \leq n} \left| \frac{1}{1 + \lambda'g(X_i, \theta)} \right| \right) \left(\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \|g_j^{(2)}(X_i, \theta)\| \right) \\ & = O(n^{-\zeta}) O_p(1) O_p(1) = o_p(1), \end{aligned}$$

where the last inequality holds by the definition of Λ_n^ζ , (27) and (5). Moreover,

$$\begin{aligned} & \sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| \frac{1}{n} \sum_{i=1}^n \left(\frac{g^{(1)}(X_i, \theta)' \lambda \lambda' g^{(1)}(X_i, \theta)}{(1 + \lambda'g(X_i, \theta))^2} \right) \right\| \\ & \leq \sup_{\lambda \in \Lambda_n^\zeta} \|\lambda\|^2 \left(\sup_{\beta \in \mathcal{B}_n, 1 \leq i \leq n} \frac{1}{(1 + \lambda'g(X_i, \theta))^2} \right) \left(\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \|g^{(1)}(X_i, \theta)\| \right) \\ & = O(n^{-2\zeta}) O_p(1) O_p(1) = o_p(1). \end{aligned}$$

The last inequality holds by the definition of Λ_n^ζ , (27) and (5).

(ii) Apply the triangle inequality to

$$\begin{aligned}
& \sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| \frac{1}{n} \sum_{i=1}^n \left(\frac{g^{(1)}(X_i, \theta)}{1 + \lambda' g(X_i, \theta)} - g^{(1)}(X_i, \theta_0) \right) \right\| \\
& \leq \sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| \frac{1}{n} \sum_{i=1}^n \left(\frac{g^{(1)}(X_i, \theta)}{1 + \lambda' g(X_i, \theta)} - g^{(1)}(X_i, \theta) \right) \right\| \\
& \quad + \sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n \left(g^{(1)}(X_i, \theta) - \mathbb{E}_n \left[g^{(1)}(X_i, \theta) \right] \right) \right\| \\
& \quad + \sup_{\theta \in \Theta: \|\theta - \theta_0\| \leq \gamma_n} \left\| \mathbb{E}_n \left[g^{(1)}(X_i, \theta) \right] - \mathbb{E}_n \left[g^{(1)}(X_i, \theta_0) \right] \right\| \\
& \quad + \left\| \frac{1}{n} \sum_{i=1}^n \left(g^{(1)}(X_i, \theta_0) - \mathbb{E}_n \left[g^{(1)}(X_i, \theta_0) \right] \right) \right\| \\
& = I_d + o_p(1) + o_p(1) + o_p(1),
\end{aligned}$$

where the last equality holds by (5) and (4). Next,

$$\begin{aligned}
I_d & \leq \sup_{\beta \in \mathcal{B}_n} |\lambda' g(X_i, \theta)| \left(\sup_{\beta \in \mathcal{B}_n} \left| \frac{1}{1 + \lambda' g(X_i, \theta)} \right| \right) \left(\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \left\| g^{(1)}(X_i, \theta) \right\| \right) \\
& = o_p(1) O_p(1) O_p(1) O_p(1) = o_p(1)
\end{aligned}$$

by Lemma 1, (27), and (5). Moreover,

$$\begin{aligned}
& \sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| \frac{1}{n} \sum_{i=1}^n \frac{g(X_i, \theta)}{1 + \lambda' g(X_i, \theta)} \frac{\lambda' g^{(1)}(X_i, \theta)}{1 + \lambda' g(X_i, \theta)} \right\| \\
& \leq \sup_{\lambda \in \Lambda_n^c} \|\lambda\| \left(\sup_{\beta \in \mathcal{B}_n, 1 \leq i \leq n} \frac{1}{(1 + \lambda' g(X_i, \theta))^2} \right) \left(\frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} \|g(X_i, \theta)\|^2 \right)^{1/2} \\
& \quad \times \left(\frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} \left\| g^{(1)}(X_i, \theta) \right\|^2 \right)^{1/2} \\
& = O(n^{-\zeta}) O_p(1) O_p(1) = o_p(1).
\end{aligned}$$

(iii) Similar as before, we have

$$\begin{aligned}
& \sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| \frac{1}{n} \sum_{i=1}^n \left(\frac{g(X_i, \theta) g(X_i, \theta)'}{(1 + \lambda' g(X_i, \theta))^2} - g(X_i, \theta_0) g(X_i, \theta_0)' \right) \right\| \\
& \leq \sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| \frac{1}{n} \sum_{i=1}^n \left(\frac{g(X_i, \theta) g(X_i, \theta)'}{(1 + \lambda' g(X_i, \theta))^2} - g(X_i, \theta) g(X_i, \theta)' \right) \right\| \\
& \quad + \sup_{\Theta} \left\| \frac{1}{n} \sum_{i=1}^n (g(X_i, \theta) g(X_i, \theta)' - \mathbb{E}_n [g(X_i, \theta) g(X_i, \theta)']) \right\| \\
& \quad + \sup_{\Theta} \left\| \mathbb{E}_n [g(X_i, \theta) g(X_i, \theta)'] - \mathbb{E}_n [g(X_i, \theta_0) g(X_i, \theta_0)'] \right\| \\
& \quad + \sup_{\Theta} \left\| \frac{1}{n} \sum_{i=1}^n (g(X_i, \theta_0) g(X_i, \theta_0)' - \mathbb{E}_n [g(X_i, \theta_0) g(X_i, \theta_0)']) \right\| \\
& = \sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| \frac{1}{n} \sum_{i=1}^n \left(\frac{g(X_i, \theta) g(X_i, \theta)'}{(1 + \lambda' g(X_i, \theta))^2} - g(X_i, \theta) g(X_i, \theta)' \right) \right\| + o_p(1).
\end{aligned}$$

Next,

$$\begin{aligned}
& \sup_{\beta \in \mathcal{B}_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \left\| \frac{1}{n} \sum_{i=1}^n \left(\frac{g(X_i, \theta) g(X_i, \theta)'}{(1 + \lambda' g(X_i, \theta))^2} - g(X_i, \theta) g(X_i, \theta)' \right) \right\| \\
& \leq \sup_{\beta \in \mathcal{B}_n, 1 \leq i \leq n} |\lambda' g(X_i, \theta)| \left(\sup_{\beta \in \mathcal{B}_n, 1 \leq i \leq n} \frac{1}{|1 + \lambda' g(X_i, \theta)|} \right) \\
& \quad \times \left(\sup_{\beta \in \mathcal{B}_n, 1 \leq i \leq n} \frac{1}{|1 + \lambda' g(X_i, \theta)|} + 1 \right) \left(\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \|g(X_i, \theta)\|^2 \right) \\
& = o_p(1) O_p(1) O_p(1) O_p(1) = o_p(1). \quad \blacksquare
\end{aligned}$$

2.3 \sqrt{n} Consistency

Theorem 2 *Suppose Assumptions 1 – 3 are satisfied. Then, (i) $\sqrt{n}(\tilde{\beta}_{nq} - \beta_{n,0}) = O_p(1)$, (ii) $\sqrt{n}(\hat{\beta}_n - \beta_{n,0}) = O_p(1)$, (iii) $nG_n^*(\hat{\beta}_n) = nG_{nq}^*(\hat{\beta}_n) + o_p(1)$, (iv) $nG_{nq}^*(\hat{\beta}_n) = nG_{nq}^*(\tilde{\beta}_{nq}) + o_p(1)$, and (v) $nG_n^*(\hat{\beta}_n) = nG_{nq}^*(\tilde{\beta}_{nq}) + o_p(1)$.*

Theorem 2 establishes that $\hat{\beta}_n$ and $\tilde{\beta}_{nq}$ are \sqrt{n} -consistent. Moreover, the theorem states that the discrepancy between $G_n^*(\beta)$ evaluated at $\hat{\beta}_n$ and $G_{nq}^*(\beta)$ evaluated at $\tilde{\beta}_{nq}$ vanishes. Thus, the large-sample behavior of likelihood ratios can be approximated by the behavior of $G_{nq}^*(\tilde{\beta}_{nq})$. Let

$$\hat{b} = [\hat{\phi}', \hat{l}'(\hat{\phi})]' \quad \text{and} \quad \tilde{b}_q = [\tilde{\phi}'_q, \tilde{l}'_q(\tilde{\phi}_q)]'$$

be re-scaled versions of $\hat{\beta}_n$ and $\tilde{\beta}_{nq}$. To prove the theorem, we will introduce a third estimator

$$\hat{b}_q = [\hat{\phi}'_q, \hat{l}'_q(\hat{\phi}_q)]'$$

where

$$\hat{l}_q(\phi) = \operatorname{argmax}_{l \in L_n(\phi)} \mathcal{G}_{nq}^*(\phi, l), \quad \hat{\phi}_q = \operatorname{argmin}_{\phi \in \Phi_n} \mathcal{G}_{nq}^*(\phi, \hat{l}_q(\phi)).$$

\hat{b}_q is based on the quadratic approximation of the objective function, but the domains of ϕ and l are restricted. In slight abuse of notation we let $B_n = \Phi_n(u_0) \otimes L_n$.

Proof of Theorem 2: (i) Follows from Lemma 7.

(ii) According to Lemma 2, $\hat{\lambda}(\hat{\theta}, \hat{\nu}) = O_p(n^{-1/2})$. It remains to show that

$$\hat{\phi} = \left[\sqrt{n}(\hat{\theta} - \theta_{n,0})', u'_{n,0} + \sqrt{n}(\hat{\nu} - \nu_{n,0})' \right]'$$

is stochastically bounded. The saddlepoint property implies that

$$0 = \mathcal{G}_n^*(\hat{\phi}, 0) \leq \mathcal{G}_n^*(\hat{\phi}, \hat{l}(\hat{\phi})) \leq \mathcal{G}_n^*(0, \hat{l}(0)). \quad (28)$$

Then using the quadratic approximation (17), the bound for the remainder term given in Lemma 5 and the definition of \hat{l} and $\hat{\phi}$ we obtain

$$\begin{aligned} \mathcal{G}_n^*(\hat{\phi}, \hat{l}(\hat{\phi})) &= \mathcal{G}_{nq}^*(\hat{\phi}, \hat{l}(\hat{\phi})) + (1 + \|\hat{\phi} - \phi_{n,0}\|^2 + \|\hat{l}(\hat{\phi})\|^2) o_p(1) \\ &= \frac{1}{2} (Z_n - R'_n(\hat{\phi} - \phi_{n,0}))' J_n^{-1} (Z_n - R'_n(\hat{\phi} - \phi_{n,0})) \\ &\quad - \frac{1}{2} (\hat{l}(\hat{\phi}) - J_n^{-1} [Z_n - R'_n(\hat{\phi} - \phi_{n,0})])' J_n (\hat{l}(\hat{\phi}) - J_n^{-1} [Z_n - R'_n(\hat{\phi} - \phi_{n,0})]) \\ &\quad + (1 + \|\hat{\phi} - \phi_{n,0}\|^2 + \|\hat{l}(\hat{\phi})\|^2) o_p(1) \\ &= \frac{1}{2} (Z_n - R'_n(\hat{\phi} - \phi_{n,0}))' J_n^{-1} (Z_n - R'_n(\hat{\phi} - \phi_{n,0})) + (1 + \|\hat{\phi} - \phi_{n,0}\|^2 + \|\hat{l}(\hat{\phi})\|^2) o_p(1), \end{aligned} \quad (29)$$

where $\phi_{n,0} = [0, u'_{n,0}]'$. The last equality is a consequence of Lemma 8. Similarly, we can deduce from Lemmas 2, 5, and Assumptions 2 and 3 that

$$\mathcal{G}_n^*(0, \hat{l}(0)) = -\frac{1}{2} \hat{l}(0)' J_n \hat{l}(0) + Z_n' \hat{l}(0) + (1 + \|\hat{l}(0)\|^2) o_p(1) = O_p(1). \quad (30)$$

Hence, from (28), (29), and (30) we obtain the inequality

$$0 \leq \frac{1}{2} (Z_n + o_p(1) - R'_n(\hat{\phi} - \phi_{n,0}))' J_n^{-1} (Z_n + o_p(1) - R'_n(\hat{\phi} - \phi_{n,0})) \leq O_p(1). \quad (31)$$

Notice that $Z_n + o_p(1) = O_p(1)$. According to Assumption 1, R_n is full rank and J_n is positive definite w.p.a. 1. Therefore, (31) implies that $\hat{\phi} - \phi_{n,0}$ is stochastically bounded.

(iii) We deduce from Lemma 5 and Part (ii) that

$$\begin{aligned} nG_{nq}^*(\hat{\beta}_n) &= \mathcal{G}_{nq}^*(\sqrt{n}(\hat{\beta}_n - \beta_{n,0})) + (1 + \|\sqrt{n}(\hat{\beta}_n - \beta_{n,0})\|^2) o_p(1) \\ &= nG_{nq}^*(\hat{\beta}_n) + O_p(1) o_p(1). \end{aligned}$$

(iv) We proceed by establishing $o_p(1)$ bounds for $nG_{nq}^*(\hat{\beta}_n) - nG_{nq}^*(\tilde{\beta}_{nq})$.

We begin with the upper bound. Using (iii) we can rewrite the differential as

$$\begin{aligned} nG_{nq}^*(\hat{\beta}_n) - nG_{nq}^*(\tilde{\beta}_{nq}) &= \mathcal{G}_n^*(\hat{\phi}, \hat{l}(\hat{\phi})) + o_p(1) - \mathcal{G}_{nq}^*(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) \\ &\leq \mathcal{G}_n^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) - \mathcal{G}_{nq}^*(\tilde{\phi}_q, \tilde{l}(\tilde{\phi}_q)) + o_p(1). \end{aligned} \quad (32)$$

Replacing $\hat{\phi}$ by $\hat{\phi}_q$ raises \mathcal{G}_n^* , whereas substituting \tilde{l}_q with \hat{l} lowers \mathcal{G}_{nq}^* . Using Lemma 5 the first term on the right-hand side of (32) can be rewritten as

$$\begin{aligned}\mathcal{G}_n^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) &= \mathcal{G}_{nq}^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) + o_p(1) \left(1 + \|\hat{\phi}_q - \phi_{n,0}\|^2 + \|\hat{l}(\hat{\phi}_q)\|^2\right) \\ &= \mathcal{G}_{nq}^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) + o_p(1).\end{aligned}\quad (33)$$

The second equality in (33) is a consequence of Lemmas 2 and 7. According to Lemma 8

$$\hat{l}(\bar{\phi}) = (J_n + o_p(1))^{-1} [Z_n - (R'_n + o_p(1))(\bar{\phi} - \phi_{n,0})]$$

for $\bar{\phi} = O_p(1)$. Hence,

$$\hat{l}(\tilde{\phi}_q) - \hat{l}(\hat{\phi}_q) = -(J_n + o_p(1))^{-1} [(R'_n + o_p(1))(\tilde{\phi}_q - \hat{\phi}_q)] = o_p(1)$$

by Lemma 7. Since $\mathcal{G}_{nq}^*(\phi, l)$ is continuous in its arguments we can now express the second term on the right-hand side of (32) as

$$\mathcal{G}_{nq}^*(\tilde{\phi}_q, \hat{l}(\tilde{\phi}_q)) = \mathcal{G}_{nq}^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) + o_p(1) \quad (34)$$

Plugging (33) and (34) into (32) we obtain the upper bound

$$n\mathcal{G}_{nq}^*(\hat{\beta}_n) - n\mathcal{G}_{nq}^*(\tilde{\beta}_{nq}) \leq o_p(1).$$

Using similar arguments, we can establish a lower bound as follows:

$$\begin{aligned}n\mathcal{G}_{nq}^*(\hat{\beta}_n) - n\mathcal{G}_{nq}^*(\tilde{\beta}_{nq}) &= \mathcal{G}_n^*(\hat{\phi}, \hat{l}(\hat{\phi})) - \mathcal{G}_{nq}^*(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) + o_p(1) \\ &\geq \mathcal{G}_n^*(\hat{\phi}, \hat{l}_q(\hat{\phi})) - \mathcal{G}_{nq}^*(\hat{\phi}, \tilde{l}_q(\hat{\phi})) + o_p(1) \\ &= \mathcal{G}_n^*(\hat{\phi}, \hat{l}_q(\hat{\phi})) - \mathcal{G}_{nq}^*(\hat{\phi}, \hat{l}_q(\hat{\phi})) + o_p(1) \\ &= o_p(1)\end{aligned}$$

which proves (iv). ■

(v) Follows from parts (iii) and (iv).

2.3.1 Technical Lemmas

Lemma 6 *Suppose Assumptions 1 to 3 are satisfied. Then, \tilde{b}_q exists uniquely w.p.a. 1.*

Proof of Lemma 6: The subsequent statements are true w.p.a. 1. Notice that $\bar{\mathcal{G}}_{nq}^*(\phi)$, defined in (25), is strictly convex function of ϕ because $R'_n = [-Q'_n, M']$ is a full rank matrix under Assumption 2(c) and J_n^{-1} is positive definite. Hence, $R_n J_n^{-1} R'_n$ is a positive definite matrix. Moreover, the domain Φ is convex. Therefore, $\tilde{\phi}_q$ is unique. Finally, from (24) we deduce that \tilde{l}_q exists uniquely. ■

Lemma 7 *Suppose Assumptions 1 to 3 are satisfied. Then*

$$(i) \tilde{b}_q = O_p(1),$$

$$(ii) \hat{b}_q = \tilde{b}_q + o_p(1).$$

Proof of Lemma 7:

Proof of (i): We will show that $\tilde{\phi}_q = O_p(1)$. For notational simplicity, denote

$$A_{1n} = R_n J_n^{-1} R_n', \quad A_{2n} = A_{1n}^{-1} R_n J_n^{-1} Z_n, \quad \text{and} \quad A_{3n} = Z_n' J_n^{-1} Z_n - A_{2n}' A_{1n} A_{2n},$$

and write the concentrated quadratic objective function (25) as

$$\bar{\mathcal{G}}_{nq}^*(\phi) = \frac{1}{2} (\phi - \phi_{n,0} + A_{2n})' A_{1n} (\phi - \phi_{n,0} + A_{2n}) + \frac{1}{2} A_{3n}.$$

Observe that J_n , R_n , and Z_n converge weakly according to Assumptions 2 and 3. Moreover based on Assumption 1, A_{1n} is positive definite w.p.a. 1. Let

$$\bar{\phi}_q = \operatorname{argmin}_{\phi \in \mathbb{R}^{m+h_2}} \bar{\mathcal{G}}_{nq}^*(\phi) = \phi_{n,0} - A_{2n} = O_p(1).$$

Notice that $\tilde{\phi}_q$ is the projection of $\bar{\phi}_q$ onto the set $\Phi(u_0)$ with respect to the inner product $\langle x, y \rangle = x' A_{1n} y$. Then,

$$\|\tilde{\phi}_q\| \leq \lambda_{\min}^{-1}(A_{1n}) \langle \tilde{\phi}_q, \bar{\phi}_q \rangle^{1/2} \leq \lambda_{\min}^{-1}(A_{1n}) \langle \bar{\phi}_q, \bar{\phi}_q \rangle^{1/2} = O_p(1)$$

where $\lambda_{\min}(A_{1n})$ denotes the smallest eigenvalue of A_{1n} and is strictly positive w.p.a. 1. Finally, from (24) we can deduce that $\tilde{l}_q(\tilde{\phi}_q) = O_p(1)$.

Proof of (ii): According to Lemma 6 the saddlepoint problem $\min_{\phi \in \Phi(u_0)} \max_{l \in \mathbb{R}^h} \mathcal{G}_{nq}^*(\phi, l)$ has a unique solution \tilde{b}_q on the domain $B = \Phi(u_0) \otimes \mathbb{R}^h$. Since $B_n \subset B$ for any $\epsilon > 0$

$$\begin{aligned} P \left\{ \|\hat{b}_q - \tilde{b}_q\| > \epsilon \right\} &\leq P \left\{ \tilde{b}_q \in B \setminus B_n \right\} \\ &\leq P \left\{ \tilde{b}_q \in B \setminus (\Phi_n(u_0) \otimes \sqrt{n} \Lambda_n^\zeta) \right\} + o(1), \end{aligned}$$

where the $o(1)$ term in the last line holds by Lemma 1(ii). The set $\sqrt{n} \Lambda_n^\zeta$ consists of the elements in Λ_n^ζ multiplied by \sqrt{n} and expands to \mathbb{R}^h because $\zeta < 1/2$. Since θ_0 is in the interior of Θ , the first m ordinates of $\Phi_n(u_0)$ expand to \mathbb{R}^m . Ordinate $m+j$ expands to \mathbb{R} if $u_{0,j} = \infty$ and to \mathbb{R}^+ otherwise. Since $\tilde{b}_q = O_p(1)$, we deduce $P\{\tilde{b}_q \in B \setminus (\Phi_n(u_0) \otimes \sqrt{n} \Lambda_n^\zeta)\} = o(1)$. Therefore $\hat{b}_q = \tilde{b}_q + o_p(1)$, as required. ■

Lemma 8 *Suppose that Assumptions 1 to 3 are satisfied. Let $\bar{\theta} \in \Theta$ and $\bar{\nu} \geq 0$ be sequences such that $\bar{\theta} - \theta_{n,0} \xrightarrow{p} 0$ and $\bar{\nu} - \nu_{n,0} \xrightarrow{p} 0$. Let $\hat{l}(\bar{\phi}) = \sqrt{n} \hat{\lambda}(\bar{\theta}, \bar{\nu})$, and $\bar{\phi} = [\bar{s}', \bar{u}']$, where $\bar{s} = \sqrt{n}(\bar{\theta} - \theta_{n,0})$ and $\bar{u} = u_{n,0} + \sqrt{n}(\bar{\nu} - \nu_{n,0})$. Then*

$$0 = Z_n - (R_n' + o_p(1))(\bar{\phi} - \phi_{n,0}) - (J_n + o_p(1))\hat{l}(\bar{\phi}).$$

Proof of Lemma 8: In view of Lemmas 1(ii) and 2, we deduce that $\hat{\lambda}(\bar{\theta}, \bar{v})$ is in the interior of $\hat{\Lambda}(\bar{\theta})$ w.p.a. 1. Hence, $\hat{\lambda}$ satisfies the first-order conditions associated with $\max_{\lambda \in \hat{\Lambda}(\bar{\theta})} G_n^*(\bar{\theta}, \bar{v}, \lambda)$:

$$0 = \frac{1}{n} \sum_{i=1}^n \frac{g(X_i, \bar{\theta})}{1 + \hat{\lambda}'g(X_i, \bar{\theta})} - M' \bar{v}.$$

We now apply the mean-value theorem and multiply by \sqrt{n} :

$$0 = \sqrt{n} G_n^{*(1)}(\beta_{n,0})_{\lambda} + G_n^{*(2)}(\beta_*)_{\lambda \theta' \bar{s}} - M'(\bar{u} - u_{n,0}) + G_n^{*(2)}(\beta_*)_{\lambda \lambda'} \hat{l},$$

where β_* lies on the line joining $\beta_{n,0}$ and $\bar{\beta} = [\bar{\theta}', \bar{v}', \hat{\lambda}(\bar{\theta}, \bar{v})']'$. The matrices $G_n^{*(1)}(\beta)$ and $G_n^{*(2)}(\beta)$ and their partitions are defined in (19) and (21). Using the same arguments as in the proof of Lemma 5 and the definitions of J_n , Q_n , R_n , and Z_n we obtain the desired result. ■