Supplementary material

D Proofs of theorems under high-level assumptions

Assumptions R and H are assumed to hold throughout this section, including H5 with $l_0 = 0$. Whenever we require H5 to hold for some $l_0 \in \{1, 2\}$, this will be explicitly noted.

D.1 Preliminary results

Let $\beta_n := \beta_0 + n^{-1/2}\delta_n$ for a (possibly) random $\delta_n = o_p(n^{1/2})$. Define

$$\Delta_n^k(\beta) := n^{1/2}[\overline{\theta}_n^k(\beta, \lambda_n) - \overline{\theta}_n^k(\beta_0, \lambda_n)]$$

and recall that $G_n(\beta) := \partial_\beta \overline{\theta}_n^k(\beta, \lambda_n)$ and $G := [\partial_\beta \theta(\beta_0, 0)]^\mathsf{T}$. As per R5, we fix the order of jackknifing $k \in \{0, \ldots, k_0\}$ such that $n^{1/2}\lambda_n^{k+1} = o_p(1)$. Let $\mathcal{L}_n(\theta) := \mathcal{L}_n(y, x; \theta)$ and $\mathcal{L}(\theta) := \mathbb{E}\mathcal{L}_n(\theta)$. $\dot{\mathcal{L}}_n$ and $\ddot{\mathcal{L}}_n$ respectively denote the gradient and Hessian of $\mathcal{L}_n$, with $H := \mathbb{E}\ddot{\mathcal{L}}_n(\theta_0) = \mathcal{L}(\theta_0)$; $N(\theta, \epsilon)$ denotes an open ball of radius ϵ, centered at θ.

Proposition D.1.

(i) $\sup_{\beta \in \mathrm{B}} \|\overline{\theta}_n^k(\beta, \lambda_n) - \theta^k(\beta, \lambda_n)\| \xrightarrow{p} 0$;

(ii) $\theta^k(\beta_0, \lambda_n) - \theta(\beta_0, 0) = O_p(\lambda_n^{k+1})$;

(iii) $\Delta_n^k(\beta_n) = G\delta_n + o_p(1 + \|\delta_n\|)$.

Proposition D.2. *For* $V = (1 + \frac{1}{M})(\Sigma - \mathrm{R})$,

$$Z_n := n^{1/2}[\overline{\theta}_n^k(\beta_0, \lambda_n) - \theta^k(\beta_0, \lambda_n)] - n^{1/2}(\hat{\theta}_n - \theta_0) \rightsquigarrow N[0, H^{-1}VH^{-1}]. \tag{D.1}$$

Proposition D.3.

(i) $Q_{nk}^e(\beta, \lambda_n) \xrightarrow{p} Q_k^e(\beta, 0) =: Q^e(\beta)$ *uniformly on* B*;*

(ii) *for every* $\epsilon > 0$, $\inf_{\beta \in \mathrm{B} \setminus N(\beta_0, \epsilon)} Q^e(\beta) > Q(\beta_0)$*; and*

Proposition D.4. *If* H5 *holds for* $l_0 = 1$*, then*

(i) $G_n(\beta_n) \xrightarrow{p} G$*; and*

if H5 *holds for* $l_0 \in \{1, 2\}$ *then, uniformly on* B,

(ii) $\sup_{\beta \in \mathrm{B}} \|\partial_\beta^l \overline{\theta}_n^k(\beta, \lambda_n) - \partial_\beta^l \theta(\beta, 0)\| = o_p(1)$*; and*

(iii) $\partial_\beta^l Q_{nk}^e(\beta, \lambda_n) \xrightarrow{p} \partial_\beta^l Q_k^e(\beta, 0) = \partial_\beta^l Q^e(\beta)$

for $l \in \{1, \ldots, l_0\}$.

Define, for some $c_n = o_p(n^{-1/2})$, the sets of approximate and exact roots

$$R^e_{nk} \coloneqq \{\beta \in \mathrm{B} \mid \|\partial_\beta Q^e_{nk}(\beta, \lambda_n)\| \leq c_n\} \qquad R^e \coloneqq \{\beta \in \mathrm{B} \mid \partial_\beta Q^e(\beta, 0) = 0\}$$

of $\partial_\beta Q^e_{nk}(\beta, \lambda_n) = 0$ and $\partial_\beta Q^e(\beta, 0) = 0$ respectively; and let

$$S^e_{nk} \coloneqq \{\beta \in R^e_{nk} \mid \varrho_{\min}[\partial^2_\beta Q^e_{nk}(\beta, \lambda_n)] \geq -c_n\} \qquad S^e \coloneqq \{\beta \in R^e \mid \varrho_{\min}[\partial^2_\beta Q^e(\beta, 0)] \geq 0\},$$

denote those subsets on which the second-order conditions for a local minimum are also approximately satisfied.

Proposition D.5. *Let* B_0 *be a compact set with* $\beta_0 \in \operatorname{int} \mathrm{B}_0$*, and* $\{\tilde{\beta}_n\}$ *a random sequence in* B_0*. Suppose* H5 *holds with* $l_0 = 1$*. Then*

(i) *if* $R^e \cap \mathrm{B}_0 = \{\beta_0\}$*, and* $\tilde{\beta}_n \in R^e_{nk}$ *w.p.a.1, then* $n^{1/2}(\tilde{\beta}_n - \hat{\beta}^e_{nk}) = o_p(1)$*; and*

(ii) *if* H5 *holds with* $l_0 = 2$*, the preceding holds with* (S^e_{nk}, S^e) *in place of* (R^e_{nk}, R^e)*.*

For the next result, let $U : \Gamma \to \mathbb{R}$ be twice continuously differentiable with a unique global minimum at γ^*. For some ϵ, let $R_U \coloneqq \{\gamma \in \Gamma \mid \|\partial_\gamma U(\gamma)\| < \epsilon\}$, and $S_U \coloneqq \{\gamma \in R_U \mid \varrho_{\min}[\partial^2_\gamma U(\gamma)] \geq -\epsilon\}$. Applying a routine $r \in \{\mathrm{GN}, \mathrm{QN}, \mathrm{TR}\}$ to U yields the iterates $\{\gamma^{(s)}\}$; let

$$\overline{\gamma}(\gamma^{(0)}, r) \coloneqq \begin{cases} \gamma^{(s^*)} & \text{if } \gamma^{(s)} \in R_U \text{ for some } s \in \mathbb{N} \\ \gamma^{(0)} & \text{otherwise,} \end{cases}$$

where s^* denotes the smallest s for which $\gamma^{(s)} \in R_U$. When $r = \mathrm{TR}$, the definition of $\overline{\gamma}(\gamma^{(0)}, \mathrm{TR})$ is analogous, but with S_U in place of R_U. In the statement of the next result, $\Gamma_0 \coloneqq \{\gamma \in \Gamma \mid U(\gamma) \leq U(\gamma_1)\}$ for some $\gamma_1 \in \Gamma$, and is a compact set with $\gamma^* \in \operatorname{int} \Gamma_0$. For a continuously differentiable function $m : \Gamma \mapsto \mathbb{R}^{d_m}$, let $M(\gamma) \coloneqq [\partial_\gamma m(\gamma)]^\mathsf{T}$ denote its Jacobian.

Proposition D.6. *Let* $r \in \{\mathrm{QN}, \mathrm{TR}\}$*, and suppose that in addition to the preceding, either*

(i) $r = \mathrm{GN}$ *and* $U(\gamma) = \|m(\gamma)\|^2$*, with* $\inf_{\gamma \in \Gamma_0} \sigma_{\min}[M(\gamma)] > 0$*; or*

(ii) $r = \mathrm{QN}$ *and* U *is strictly convex on* Γ_0*;*

then $\overline{\gamma}(\gamma^{(0)}, r) \in R_U \cap \Gamma_0$ *for all* $\gamma^{(0)} \in \Gamma_0$*. Alternatively, if* $r = \mathrm{TR}$*, then* $\overline{\gamma}(\gamma^{(0)}, r) \in S_U \cap \Gamma_0$ *for all* $\gamma^{(0)} \in \Gamma_0$*.*

D.2 Proofs of Theorems 4.1–4.3

Throughout this section, $\beta_n \coloneqq \beta_0 + n^{-1/2}\delta_n$ for a (possibly) random $\delta_n = o_p(n^{1/2})$. Let $Q^{\mathrm{W}}_n(\beta) \coloneqq Q^{\mathrm{W}}_{nk}(\beta, \lambda_n)$, $Q^{\mathrm{LR}}_n(\beta) \coloneqq Q^{\mathrm{LR}}_{nk}(\beta, \lambda_n)$, and $\overline{\theta}_n(\beta) \coloneqq \overline{\theta}^k_n(\beta, \lambda_n)$.

Proof of Theorem 4.1. We first consider the Wald estimator. We have

$$n[Q^{\mathrm{W}}_n(\beta_n) - Q^{\mathrm{W}}_n(\beta_0)] = 2n^{1/2}[\overline{\theta}^k_n(\beta_0) - \hat{\theta}_n]^\mathsf{T} W_n \Delta^k_n(\beta_n) + \Delta^k_n(\beta_n)^\mathsf{T} W_n \Delta^k_n(\beta_n).$$

For Z_n as defined in (D.1), we see that by Proposition D.1(ii) and R5

$$n^{1/2}[\overline{\theta}_n^k(\beta_0) - \hat{\theta}] = Z_n + n^{1/2}[\theta^k(\beta_0, \lambda_n) - \theta_0] = Z_n + o_p(1), \tag{D.2}$$

whence by Proposition D.1(iii),

$$n[Q_n^{\mathrm{W}}(\beta_n) - Q_n^{\mathrm{W}}(\beta_0)] = 2Z_n^{\mathsf{T}} W G \delta_n + \delta_n^{\mathsf{T}} G^{\mathsf{T}} W G \delta_n + o_p(1 + \|\delta_n\| + \|\delta_n\|^2). \tag{D.3}$$

Now consider the LR estimator. Twice continuous differentiability of the likelihood yields

$$\begin{aligned} n[Q_n^{\mathrm{LR}}(\beta) - Q_n^{\mathrm{LR}}(\beta_0)] &= -n[\mathcal{L}_n(\overline{\theta}_n^k(\beta_n)) - \mathcal{L}_n(\overline{\theta}_n^k(\beta_0))] \\ &= -n^{1/2} \dot{\mathcal{L}}_n(\overline{\theta}_n^k(\beta_0))^{\mathsf{T}} \Delta_n^k(\beta_n) - \frac{1}{2} \Delta_n^k(\beta_n)^{\mathsf{T}} \ddot{\mathcal{L}}_n(\overline{\theta}_n^k(\beta_0)) \Delta_n^k(\beta_n) \\ &\qquad + o_p(\|\Delta_n^k(\beta_n)\|^2) \end{aligned}$$

where by Proposition D.1(ii) and H3,

$$\begin{aligned} n^{1/2} \dot{\mathcal{L}}_n[\overline{\theta}_n^k(\beta_0)] &= n^{1/2} \dot{\mathcal{L}}_n(\theta_0) + \ddot{\mathcal{L}}_n(\theta_0) n^{1/2} [\overline{\theta}_n^k(\beta_0) - \theta_0] + o_p(1) \\ &= H[Z_n + n^{1/2}(\theta^k(\beta_0, \lambda_n) - \theta_0)] \\ &= HZ_n + o_p(1) \end{aligned} \tag{D.4}$$

for Z_n as in (D.1). Thus by Proposition D.1(iii),

$$n[Q_n^{\mathrm{LR}}(\beta_n) - Q_n^{\mathrm{LR}}(\beta_n)] = -Z_n^{\mathsf{T}} H G \delta_n - \frac{1}{2} \delta_n^{\mathsf{T}} G^{\mathsf{T}} H G \delta_n + o_p(1 + \|\delta_n\| + \|\delta_n\|^2). \tag{D.5}$$

Consistency of $\hat{\beta}_{nk}^e$ follows from parts (i) and (ii) of Proposition D.3 and Corollary 3.2.3 in van der Vaart and Wellner (1996). Thus by applying Theorem 3.2.16 in van der Vaart and Wellner (1996) – or more precisely, the arguments following their (3.2.17) – to (D.3) and (D.5), we have

$$n^{1/2}(\hat{\beta}_{nk}^e - \beta_0) = -(G^{\mathsf{T}} U_e G)^{-1} G^{\mathsf{T}} U_e Z_n + o_p(1) \tag{D.6}$$

for U_e as in (4.7); the result now follows by Proposition D.2. □

Proof of Theorem 4.2. We first note that, in consequence of H3 and Theorem 4.1, $\hat{\beta}_{nk}^e \xrightarrow{p} \beta_0$, $\hat{\theta}_n \xrightarrow{p} \theta_0$, and $\hat{\theta}_n^m := \hat{\theta}_n^m(\hat{\beta}_{nk}^e, \lambda_n) \xrightarrow{p} \theta_0$. Part (i) then follows from R2, H2, and Lemma 2.4 in Newey and McFadden (1994). Defining $\dot{\ell}_i^m(\theta_0) := \dot{\ell}_i^m(\beta_0, 0; \theta_0)$ for $m \in \{1, \ldots, M\}$ and

$$\varsigma_i^{\mathsf{T}} := \begin{bmatrix} \dot{\ell}_i^0(\theta_0)^{\mathsf{T}} & \dot{\ell}_i^1(\beta_0, 0; \theta_0)^{\mathsf{T}} & \cdots & \dot{\ell}_i^M(\beta_0, 0; \theta_0)^{\mathsf{T}} \end{bmatrix},$$

H2 and H3 further imply that

$$A^{\mathsf{T}} \left(\frac{1}{n} \sum_{i=1}^n s_{ni} s_{ni}^{\mathsf{T}} \right) A \xrightarrow{p} A^{\mathsf{T}} (\mathbb{E} \varsigma_i \varsigma_i^{\mathsf{T}}) A = A^{\mathsf{T}} \begin{bmatrix} \Sigma & \mathrm{R} & \cdots & \mathrm{R} \\ \mathrm{R} & \Sigma & \cdots & \mathrm{R} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{R} & \mathrm{R} & \cdots & \Sigma \end{bmatrix} A = V.$$

Part (iii) is an immediate consequence of Proposition D.4(i). □

Proof of Theorem 4.3. For each $r \in \{\mathrm{GN}, \mathrm{QN}, \mathrm{TR}\}$, suppose that there exists a $\mathrm{B}_0 \subseteq \mathrm{B}$ such that $U = Q_n^e(\beta) := Q_{nk}^e(\beta, \lambda_n)$ satisfies the corresponding part of Proposition D.6, w.p.a.1. Then

$$\mathbb{P}\{\overline{\beta}_{nk}^e(\beta^{(0)}, r) \in R_{nk}^e \cap \mathrm{B}_0,\ \forall \beta^{(0)} \in \mathrm{B}_0\} \xrightarrow{p} 1 \tag{D.7}$$

for $r \in \{\mathrm{GN}, \mathrm{QN}\}$, and also for $r = \mathrm{TR}$ with S_{nk}^e in place of R_{nk}^e. Further, $R^e \cap \mathrm{B}_0 = \{\beta_0\}$ under O-GN and O-QN, while $S^e \cap \mathrm{B}_0 = \{\beta_0\}$ under O-TR.

Now let $\tilde{\beta}_n^{(0)}$ be a random sequence in B_0. When $r \in \{\mathrm{GN}, \mathrm{QN}\}$, it follows from (D.7) that $\overline{\beta}_{nk}^e := \overline{\beta}_{nk}^e(\tilde{\beta}_n^{(0)}, r) \in R_{nk}^e \in \mathrm{B}_0$ w.p.a.1, and so by Proposition D.5(i), $n^{1/2}(\overline{\beta}_{nk}^e - \hat{\beta}_{nk}^e) = o_p(1)$. When $r = \mathrm{TR}$, the result follows analogously from Proposition D.5(ii).

It thus remains to verify that the requirements of Proposition D.6 hold w.p.a.1. When $r = \mathrm{GN}$, it follows from Proposition D.4(i), the continuity of $\sigma_{\min}(\cdot)$ and O-GN that

$$0 < \inf_{\beta \in \mathrm{B}_0} \sigma_{\min}[G(\beta)] = \inf_{\beta \in \mathrm{B}_0} \sigma_{\min}[G_n(\beta)] + o_p(1),$$

whence $\inf_{\beta \in \mathrm{B}_0} \sigma_{\min}[G_n(\beta)] > 0$ w.p.a.1. When $r = \mathrm{QN}$, it follows from Proposition D.4(iii) and O-QN that

$$0 < \inf_{\beta \in \mathrm{B}_0} \varrho_{\min}[\partial_\beta^2 Q^e(\beta)] = \inf_{\beta \in \mathrm{B}_0} \varrho_{\min}[\partial_\beta^2 Q_n^e(\beta)] + o_p(1)$$

whence Q_n^e is strictly convex on B_0 w.p.a.1. When $r = \mathrm{TR}$, there are no additional conditions to verify. □

D.3 Proofs of Propositions D.1–D.6

Proof of Proposition D.1. Part (i) follows by H5 and the continuous mapping theorem. Part (ii) is immediate from (3.10). For part (iii), we note that for $\beta_n = \beta_0 + n^{1/2}\delta_n$ with $\delta_n = o_p(n^{1/2})$ as above,

$$\begin{aligned}\Delta_n^k(\beta_n) = n^{1/2}[\overline{\theta}_n^k(\beta_n, \lambda_n) - \theta^k(\beta_n, \lambda_n)]&\\ - n^{1/2}[\overline{\theta}_n^k(\beta_0, \lambda_n) - \theta^k(\beta_0, \lambda_n)] &+ n^{1/2}[\theta^k(\beta_n, \lambda_n) - \theta^k(\beta_0, \lambda_n)].\end{aligned}$$

Since $\overline{\theta}_n^k$ is a linear combination of the $\hat{\theta}_n^m$'s, it is clear from H3 and H4 that the first two terms converge jointly in distribution to identical limits (since $\beta_n \xrightarrow{p} \beta_0$). For the final term, continuous differentiability of θ^k (R3 above) entails that

$$\begin{aligned}n^{1/2}[\theta^k(\beta_n, \lambda_n) - \theta^k(\beta_0, \lambda_n)] &= [\partial_\beta \theta^k(\beta_0, \lambda_n)]^\mathsf{T}(\beta_n - \beta_0) + o_p(\|\beta_n - \beta_0\|)\\ &= G\delta_n + o_p(1 + \|\delta_n\|).\end{aligned}$$

□

Proof of Proposition D.2. Note first that

$$n^{1/2}[\overline{\theta}_n^k(\beta_0,\lambda_n) - \theta^k(\beta_0,\lambda_n)] = \sum_{r=0}^{k} \gamma_{rk} \cdot n^{1/2}[\overline{\theta}_n(\beta_0,\delta^r\lambda_n) - \theta(\beta_0,\delta^r\lambda_n)]$$
$$= -\frac{1}{M}\sum_{m=1}^{M}\sum_{r=0}^{k}\gamma_{rk}H^{-1}\phi_n^m + o_p(1) \rightsquigarrow -\frac{1}{M}\sum_{m=1}^{M} H^{-1}\phi^m,$$

by (3.10), (3.11), H3, H4 and $\sum_{r=0}^{k}\gamma_{rk} = 1$. By H3 and H4, this holds jointly with

$$n^{1/2}(\hat{\theta}_n - \theta_0) \rightsquigarrow -H^{-1}\phi^0.$$

The limiting variance of Z_n is thus equal to

$$\operatorname{var}\left[-H^{-1}\phi^0 + \frac{1}{M}\sum_{m=1}^{M} H^{-1}\phi^m\right] = H^{-1}\operatorname{var}\left[-\phi^0 + \frac{1}{M}\sum_{m=1}^{M}\phi^m\right]H^{-1} = H^{-1}VH^{-1}$$

where the final equality follows from H4 and straightforward calculations. □

Proof of Proposition D.3. We first prove part (i). For the Wald estimator, this is immediate from Proposition D.1(i). For the LR estimator, it follows from Proposition D.1(i), H2 and the continuous mapping theorem (arguing as on pp. 144f. of Billingsley, 1968), that

$$Q_{nk}^{\mathrm{LR}}(\beta) = (\mathcal{L}_n \circ \overline{\theta}_n^k)(\beta,\lambda_n) \xrightarrow{p} (\mathcal{L} \circ \theta^k)(\beta,0) = Q^{\mathrm{LR}}(\beta),$$

uniformly on B.

For part (ii), we note that $\beta \mapsto \theta^k(\beta,0)$ is continuous by R3, while the continuity of $\mathcal{L}$ is implied by H2, since $\mathcal{L}_n$ is continuous. Thus Q^e is continuous for $e \in \{\mathrm{W},\mathrm{LR}\}$, and by R4 is uniquely minimized at β_0. Hence $\beta \mapsto Q^e(\beta)$ has a well-separated minimum, which by R1 is interior to B. □

Proof of Proposition D.4. Part (ii) is immediate from H5, (3.11) and the continuous mapping theorem; it further implies part (i). For part (iii), recall $\dot{Q}_n^e(\beta) = \partial_\beta Q_n^e(\beta)$, and $G_n(\beta) = [\partial_\beta \overline{\theta}_n^k(\beta)]^{\mathsf{T}}$. Then we have

$$\dot{Q}_n^{\mathrm{W}}(\beta) = G_n(\beta)^{\mathsf{T}} W_n[\overline{\theta}_n(\beta) - \hat{\theta}_n] \qquad \dot{Q}_n^{\mathrm{LR}}(\beta) = G_n(\beta)^{\mathsf{T}}\dot{\mathcal{L}}_n[\overline{\theta}_n^k(\beta)].$$

Part (i), and similar arguments as were used are used in the proof of part (i) of Proposition D.3, yield that $\dot{Q}_n^e(\beta) \xrightarrow{p} \partial_\beta Q^e(\beta,0) =: \dot{Q}^e(\beta)$ uniformly on B. The proof that the second derivatives converge uniformly is analogous. □

Proof of Proposition D.5. We first prove part (i). Let $\dot{Q}_n^e(\beta) := \partial_\beta Q_n^e(\beta)$ and $\dot{Q}^e(\beta) := \partial_\beta Q^e(\beta,0)$. By Proposition D.4(iii)

$$\dot{Q}^e(\tilde{\beta}_n) = \dot{Q}_n^e(\tilde{\beta}_n) + o_p(1) = o_p(1 + c_n) = o_p(1). \tag{D.8}$$

Since $\dot{Q}^e$ is continuous and B_0 compact, and $\beta_0 \in \operatorname{int} \mathrm{B}_0$ is the unique element of B_0 for which $\dot{Q}^e(\beta_0) = 0$, it follows that $\tilde{\beta}_n \xrightarrow{p} \beta_0$. Hence we may write $\tilde{\beta}_n = \beta_0 + n^{1/2}\tilde{\delta}_n$, with $\tilde{\delta}_n = o_p(n^{1/2})$.

For the Wald criterion, we have

$$o_p(1) = n^{1/2}\dot{Q}_n^{\mathrm{W}}(\tilde{\beta}_n)^{\mathsf{T}} = 2[n^{1/2}(\overline{\theta}_n^k(\tilde{\beta}_n) - \hat{\theta}_n)]^{\mathsf{T}} W G_n(\tilde{\beta}_n)$$

where, for Z_n as in (D.1),

$$n^{1/2}(\overline{\theta}_n^k(\tilde{\beta}_n) - \hat{\theta}_n) = n^{1/2}(\overline{\theta}_n^k(\beta_0) - \hat{\theta}_n) + \Delta_n^k(\tilde{\beta}_n) = Z_n + G\tilde{\delta}_n + o_p(1 + \|\tilde{\delta}_n\|)$$

by (D.2), R5, and parts (ii) and (iii) of Proposition D.1. Hence, using Proposition D.4(i),

$$o_p(1) = 2[\tilde{\delta}_n^{\mathsf{T}} G^{\mathsf{T}} W G + Z_n^{\mathsf{T}} W G] + o_p(1 + \|\tilde{\delta}_n\|). \tag{D.9}$$

Similarly, for the LR criterion,

$$o_p(1) = n^{1/2}\partial_\beta Q_n^{\mathrm{LR}}(\tilde{\beta}_n)^{\mathsf{T}} = n^{1/2}\dot{\mathcal{L}}_n[\overline{\theta}_n^k(\tilde{\beta}_n)]^{\mathsf{T}} G_n(\tilde{\beta}_n)$$

where by the twice continuous differentiability of the likelihood, Proposition D.1(iii) and (D.4),

$$\begin{aligned} n^{1/2}\dot{\mathcal{L}}_n[\overline{\theta}_n^k(\tilde{\beta}_n)] &= n^{1/2}\dot{\mathcal{L}}_n[\overline{\theta}_n^k(\beta_0)] + \ddot{\mathcal{L}}_n(\overline{\theta}_n^k(\beta_0))\Delta_n^k(\tilde{\beta}_n) + o_p(\|\Delta_n^k(\tilde{\beta}_n)\|) \\ &= HZ_n + HG\tilde{\delta}_n + o_p(1 + \|\tilde{\delta}_n\|). \end{aligned}$$

Thus by Proposition D.4(i),

$$o_p(1) = \tilde{\delta}_n^{\mathsf{T}} G^{\mathsf{T}} H G + Z_n^{\mathsf{T}} H G + o_p(1 + \|\tilde{\delta}_n\|). \tag{D.10}$$

Hence using (D.9) and (D.10), we see that for U_e as in (4.7),

$$n^{1/2}(\tilde{\beta}_{nk}^e - \beta_0) = -(G^{\mathsf{T}} U_e G)^{-1} G^{\mathsf{T}} U_e Z_n + o_p(1) = n^{1/2}(\hat{\beta}_{nk}^e - \beta_0) + o_p(1) \tag{D.11}$$

for $e \in \{\mathrm{W}, \mathrm{LR}\}$. The final equality follows from Theorem 4.1: or more precisely, from (D.6) in the proof of Theorem 4.1.

We now turn to part (ii). Let $\ddot{Q}_n^e(\beta) := \partial_\beta^2 Q_n^e(\beta)$, $\ddot{Q}^e(\beta) := \partial_\beta^2 Q^e(\beta, 0)$. By Proposition D.4(iii) and the continuity of the minimum eigenvalue,

$$\varrho_{\min}[\ddot{Q}^e(\tilde{\beta}_n)] = \varrho_{\min}[\ddot{Q}_n^e(\tilde{\beta}_n)] + o_p(1) \geq -c_n + o_p(1) \to 0.$$

Since (D.8) also holds, and $S^e \cap \mathrm{B}_0 = \{\beta_0\}$, it follows that $\tilde{\beta}_n \xrightarrow{p} \beta_0$. Since $\tilde{\beta}_n \in S_{nk}^e \subseteq R_{nk}^e$ w.p.a.1, (D.11) follows immediately from the arguments given in the proof of part (i). □

Proof of Proposition D.6. For $r = \mathrm{GN}$, the result follows by Theorem 10.1 in Nocedal and Wright (2006); for $r = \mathrm{QN}$, by their Theorem 6.5; for $r = \mathrm{TR}$, by Theorem 4.7 in Moré and Sorensen (1983). □

E Sufficiency of the low-level assumptions

We shall henceforth maintain both Assumptions L and R, and address the question of whether these are sufficient for Assumption H; that is, we shall prove Proposition 4.1.

Recall that, as per L9, the auxiliary model is the Gaussian SUR displayed in (B.1) above. For simplicity, we shall consider only the case where Σ_ξ is unrestricted, but our arguments extend straightforwardly to the case where Σ_ξ is block diagonal (as would typically be imposed when $T > 1$). Recall that θ collects the elements of α and Σ_ξ^{-1}. Fix an $m \in \{0, 1, \ldots M\}$, and define

$$\xi_{ri}(\alpha) \coloneqq y_r(z_i; \beta, \lambda) - \alpha_{xr}^{\mathsf{T}} \Pi_{xr} x(z_i) - \alpha_{yr}^{\mathsf{T}} \Pi_{yr} y(z_i; \beta, \lambda),$$

temporarily suppressing the dependence of y (and hence ξ_{ri}) on m. Collecting $\xi_i \coloneqq (\xi_{1i}, \ldots, \xi_{d_y i})^{\mathsf{T}}$, the average log-likelihood of the auxiliary model can be written as

$$\mathcal{L}_n(y, x; \theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(y_i, x_i; \theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\det\Sigma_\xi - \frac{1}{2}\operatorname{tr}\left[\Sigma_\xi^{-1}\frac{1}{n}\sum_{i=1}^{n}\xi_i(\alpha)\xi_i(\alpha)^{\mathsf{T}}\right].$$

Deduce that there are functions L and l, which are three times continuously differentiable in both arguments (at least on $\operatorname{int}\Theta$), such that

$$\mathcal{L}_n(y, x; \theta) = L(T_n; \theta) \qquad \ell(y_i, x_i; \theta) = l(t_i; \theta) \tag{E.1}$$

where

$$t_i^m(\beta, \lambda) = \begin{bmatrix} y(z_i^m; \beta, \lambda) \\ x(z_i^m) \end{bmatrix}$$

and $T_n^m \coloneqq \operatorname{vech}(\mathcal{T}_n^m)$, for

$$\mathcal{T}_n^m(\beta, \lambda) \coloneqq \frac{1}{n}\sum_{i=1}^{n} t_i^m(\beta, \lambda) t_i^m(\beta, \lambda)^{\mathsf{T}}. \tag{E.2}$$

Further, direct calculation gives

$$\partial_{\alpha_{xr}}\ell_i(\theta) = \sum_{s=1}^{d_y}\sigma^{rs}\xi_{si}(\alpha)\Pi_{xr}x(z_i) \qquad \partial_{\alpha_{yr}}\ell_i(\theta) = \sum_{s=1}^{d_y}\sigma^{rs}\xi_{si}(\alpha)\Pi_{yr}y(z_i; \beta, \lambda) \tag{E.3}$$

and

$$\partial_{\sigma^{rs}}\ell_i(\theta) = \frac{1}{2}\sigma_{rs} - \frac{1}{2}\xi_{ri}(\alpha)\xi_{si}(\alpha). \tag{E.4}$$

Since the elements of the score vector $\dot{\ell}_i(\theta) = \partial_\theta \ell_i(\theta)$ necessarily take one of the forms displayed in (E.3) or (E.4), we may conclude that, for any compact subset $A \subset \Theta$, there exists a C_A such that

$$\mathbb{E}\sup_{\theta\in A}\|\dot{\ell}_i(\theta)\|^2 \le C_A \mathbb{E}\|z_i\|^4 < \infty \tag{E.5}$$

with the second inequality following from L7.

Regarding the maximum likelihood estimator (MLE), we note that the concentrated average log-likelihood is given by

$$\mathcal{L}_n(y,x;\alpha) = -\frac{d_y}{2}(\log 2\pi + 1) - \frac{1}{2}\log\det\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i(\alpha)\xi_i(\alpha)^{\mathsf{T}}\right] = L_c(T_n;\alpha)$$

which is three times continuously differentiable in α and T_n, so long as $\mathcal{T}_n$ is non-singular. By the implicit function theorem, it follows that $\hat{\alpha}_n$ may be regarded as a smooth function of T_n. Noting the usual formula for the ML estimates of Σ_ξ, this holds also for the components of θ referring to Σ_ξ^{-1}, whence

$$\hat{\theta}_n^m(\beta,\lambda) = h[T_n^m(\beta,\lambda)] \tag{E.6}$$

for some h that is twice continuously differentiable on the set where $\mathcal{T}_n^m$ has full rank. Under L8, this occurs uniformly on $\mathrm{B}\times\Lambda$ w.p.a.1., and so to avoid tiresome circumlocution, we shall simply treat h as if it were everywhere twice continuously differentiable throughout the sequel. Letting $T(\beta,\lambda) \coloneqq \mathbb{E}T_n^0(\beta,\lambda)$, we note that the population binding function is given by

$$\theta(\beta,\lambda) = h[T(\beta,\lambda)]. \tag{E.7}$$

Define $\varphi_n^m(\beta,\lambda) \coloneqq n^{1/2}[T_n^m(\beta,\lambda) - T(\beta,\lambda)]$, and let $[\varphi^m(\beta,\lambda)]_{m=0}^M$ denote a vector-valued continuous Gaussian process on $\mathrm{B}\times\Lambda$ with covariance kernel

$$\operatorname{cov}(\varphi^{m_1}(\beta_1,\lambda_1),\varphi^{m_2}(\beta_2,\lambda_2)) = \operatorname{cov}(T_n^{m_1}(\beta_1,\lambda_1),T_n^{m_2}(\beta_2,\lambda_2)).$$

Note that L7, in particular the requirement that $\mathbb{E}\|z_i\|^4 < \infty$, ensures that this covariance exists and is finite.

Lemma E.1.

(i) $\varphi_n^m(\beta,\lambda) \rightsquigarrow \varphi^m(\beta,\lambda)$ *in* $b^\infty(\mathrm{B}\times\Lambda)$*, jointly for* $m\in\{0,\dots,M\}$*; and*

(ii) *if (4.3) holds for* $l' = l \in \{1,2\}$*, then*

$$\sup_{\beta\in\mathrm{B}}\|\partial_\beta^l T_n^m(\beta,\lambda_n) - \partial_\beta^l T(\beta,0)\| = o_p(1) \tag{E.8}$$

By an application of the delta method, we thus have

Corollary E.1. *For* $\dot{h}(\beta,\lambda) \coloneqq \partial_\beta h[T(\beta,\lambda)]$,

$$\psi_n^m(\beta,\lambda) \coloneqq n^{1/2}[\hat{\theta}_n^m(\beta,\lambda) - \theta(\beta,\lambda)] \rightsquigarrow \dot{h}(\beta,\lambda)\varphi^m(\beta,\lambda) =: \psi^m(\beta,\lambda) \tag{E.9}$$

in $b^\infty(\mathrm{B}\times\Lambda)$*, jointly for* $m\in\{0,\dots,M\}$.

The proof of Lemma E.1 appears in Appendix E.1.

Proof of Proposition 4.1. H1 follows from the twice continuous differentiability of L in (E.1). The first part of H2 is an immediate consequence of Lemma E.1(i) and the smoothness of L; the second part is implied by (E.5) and Lemma 2.4 in Newey and McFadden (1994).

By Corollary E.1, we have for any $\beta_n = \beta_0 + o_p(1)$ and $\lambda_n = o_p(1)$ that

$$n^{1/2}[\hat{\theta}_n^m(\beta_n, \lambda_n) - \theta(\beta_n, \lambda_n)] = n^{1/2}[\hat{\theta}_n^m(\beta_0, 0) - \theta(\beta_0, 0)] + o_p(1)$$
$$= -H^{-1}\frac{1}{n^{1/2}}\sum_{i=1}^{n} \dot{\ell}_i^m(\beta_0, 0; \theta_0) + o_p(1)$$

where for $m \in \{0, 1, \dots, M\}$; the final equality follows from the consistency of $\hat{\theta}_n^m(\beta_0, 0)$ (as implied by Corollary E.1) and the arguments used to prove Theorem 3.1 in Newey and McFadden (1994). By definition, $\phi_n^m := n^{-1/2}\sum_{i=1}^n \dot{\ell}_i^m(\beta_0, 0; \theta_0)$, and thus H3 holds. H4 follows by the central limit theorem, in view of L1 and (E.5). Finally, H5 follows from (E.6), (E.7), Lemma E.1(ii) and the chain rule. □

E.1 Proof of Lemma E.1

For the purposes of the proofs undertaken in this section, we may suppose without loss of generality that $\tilde{D} = I_{d_y}$ in L3, $\gamma(\beta) = \beta$ in L4, and $\|K\|_\infty \le 1$. Recalling (B.3) above, we have

$$y_r(\beta, \lambda) = \omega_r(\beta) \cdot \prod_{s \in \mathcal{S}_r} K_\lambda[\nu_s(\beta)] =: \omega_r(\beta) \cdot \mathbb{K}(\mathcal{S}_r; \beta, \lambda). \tag{E.10}$$

Let $\dot{K}$ and $\ddot{K}$ respectively denote the first and second derivatives of K. For future reference, we here note that

$$\partial_\beta y_r(\beta, \lambda) = z_{wr} \cdot \mathbb{K}(\mathcal{S}_r; \beta, \lambda) + \lambda^{-1} w_r(\beta) \sum_{s \in \mathcal{S}_r} z_{vs} \cdot \mathbb{K}_s(\mathcal{S}_r; \beta, \lambda) \tag{E.11}$$
$$=: D_{r1}(\beta, \lambda) + \lambda^{-1} D_{r2}(\beta, \lambda)$$

where $z_{vr} := \Pi_{vr}^\mathsf{T} z$, $z_{wr} := \Pi_{wr}^\mathsf{T} z$ and $\mathbb{K}_s(\mathcal{S}; \beta, \lambda) := \dot{K}_\lambda[v_s(\beta)] \cdot \mathbb{K}(\mathcal{S}\backslash\{s\}; \beta, \lambda)$; and

$$\partial_\beta^2 y_r(\beta, \lambda) = \lambda^{-1} \sum_{s \in \mathcal{S}_r} [z_{wr} z_{vs}^\mathsf{T} + z_{vs} z_{wr}^\mathsf{T}] \cdot \mathbb{K}_s(\mathcal{S}_r; \beta, \lambda) \tag{E.12}$$
$$+ \lambda^{-2} w_r(\beta) \sum_{s \in \mathcal{S}_r} \sum_{t \in \mathcal{S}_r} z_{vs} z_{vt}^\mathsf{T} \cdot \mathbb{K}_{st}(\mathcal{S}_r; \beta, \lambda)$$
$$=: \lambda^{-1} H_{r1}(\beta, \lambda) + \lambda^{-2} H_{r2}(\beta, \lambda)$$

for

$$\mathbb{K}_{st}(\mathcal{S}; \beta, \lambda) := \begin{cases} \ddot{K}_\lambda[v_s(\beta)] \cdot \mathbb{K}(\mathcal{S}\backslash\{s\}; \beta, \lambda) & \text{if } s = t, \\ \dot{K}_\lambda[v_s(\beta)] \cdot \dot{K}_\lambda[v_t(\beta)] \cdot \mathbb{K}(\mathcal{S}\backslash\{s, t\}; \beta, \lambda) & \text{if } s \neq t. \end{cases}$$

E.1.1 Proof of part (ii)

In view of (E.2), the scalar elements of $T_n(\beta, \lambda)$ that depend on (β, λ) take either of the following forms:

$$\tau_{n1}(\beta, \lambda) := \mathbb{E}_n[y_r(\beta, \lambda) y_s(\beta, \lambda)] \qquad \tau_{n2}(\beta, \lambda) := \mathbb{E}_n[y_r(\beta, \lambda) x_t] \tag{E.13}$$

for some $r, s \in \{1, \dots, d_y\}$, or $t \in \{1, \dots, d_x\}$, where $\mathbb{E}_n f(\beta, \lambda) := \frac{1}{n}\sum_{i=1}^n f(z_i; \beta, \lambda)$. (Throughout the following, all statements involving r, s and t should be interpreted as holding for all possible values of these indices.) For $k \in \{1, 2\}$ and $l \in \{0, 1, 2\}$, define $\tau_k(\beta, \lambda) := \mathbb{E}\tau_{nk}(\beta, \lambda)$ – a typical scalar element of $T(\beta, \lambda)$ – and $\tau_k^{[l]}(\beta, \lambda) := \mathbb{E}\partial_\beta^l \tau_{nk}(\beta, \lambda)$. Thus part (ii) of Lemma E.1 will follow once we have shown that

$$\partial_\beta^l \tau_{nk}(\beta, \lambda_n) = \tau_k^{[l]}(\beta, \lambda_n) + o_p(1) = \partial_\beta^l \tau_k(\beta, 0) + o_p(1) \tag{E.14}$$

uniformly in $\beta \in \mathrm{B}$. The second equality in (E.14) is implied by

Lemma E.2. $\tau_k^{[l]}(\beta, \lambda_n) \xrightarrow{p} \partial_\beta^l \tau_k(\beta, 0)$, *uniformly on* B, *for* $k \in \{1, 2\}$ *and* $l \in \{0, 1, 2\}$.

The proof appears at the end of this section. We turn next to the first equality in (E.14). We require the following definitions. A function $F : \mathcal{Z} \mapsto \mathbb{R}$ is an *envelope* for the class $\mathcal{F}$ if $\sup_{f \in \mathcal{F}} |f(z)| \leq F(z)$. For a probability measure $\mathbb{Q}$ and a $p \in (1, \infty)$, let $\|f\|_{p,\mathbb{Q}} := (\mathbb{E}_\mathbb{Q}|f(z_i)|^p)^{1/p}$. $\mathcal{F}$ is *Euclidean* for the envelope F if

$$\sup_\mathbb{Q} N(\epsilon\|F\|_{1,\mathbb{Q}}, \mathcal{F}, L_{1,\mathbb{Q}}) \leq C_1 \epsilon^{-C_2}$$

for some C_1 and C_2 (depending on $\mathcal{F}$), where $N(\epsilon, \mathcal{F}, L_{1,\mathbb{Q}})$ denotes the minimum number of $L_{1,\mathbb{Q}}$-balls of diameter ϵ needed to cover $\mathcal{F}$. For a parametrized family of functions $g(\beta, \lambda) = g(z; \beta, \lambda) : \mathscr{Z} \mapsto \mathbb{R}^{d_1 \times d_2}$, let $\mathcal{F}(g) := \{g(\beta, \lambda) \mid (\beta, \lambda) \in \mathrm{B} \times \Lambda\}$. Since B is compact, we may suppose without loss of generality that $\mathrm{B} \subseteq \{\beta \in \mathbb{R}^{d_\beta} \mid \|\beta\| \leq 1\}$, whence recalling (B.2) and (B.4) above,

$$|w_r(z; \beta)| \leq W_r \leq \begin{cases} \|z\| & \text{if } r \in \{1, \dots d_w\} \\ 1 & \text{if } r \in \{d_w + 1, \dots d_y\}. \end{cases}$$

Thus by Lemma 22 in Nolan and Pollard (1987)

E1 for $\mathbb{L} \in \{\mathbb{K}, \mathbb{K}_s, \mathbb{K}_{st}\}$, $s, t \in \{1, \dots, d_y\}$ and $\mathcal{S} \subseteq \{1, \dots, d_v\}$, the class

$$\mathcal{F}(\mathbb{L}, \mathcal{S}) := \{\mathbb{L}(\mathcal{S}; \beta, \lambda) \mid (\beta, \lambda) \in \mathrm{B} \times \Lambda\}$$

is Euclidean with constant envelope; and

E2 for $r \in \{1, \dots, d_y\}$, $\mathcal{F}(w_r)$ is Euclidean for W_r.

It therefore follows by a slight adaptation of the proof of Theorem 9.15 in Kosorok (2008) that

E3 $\mathcal{F}(y_r)$ is Euclidean for W_r;

E4 $\mathcal{F}(y_r D_{s1})$ and $\mathcal{F}(y_r D_{s2})$ are Euclidean for $W_r W_s \|z\|$

E5 $\mathcal{F}(x_t D_{s1})$ and $\mathcal{F}(x_t D_{s2})$ are Euclidean for $W_s \|z\|^2$;

E6 $\mathcal{F}(D_{s1} D_{r1}^\mathsf{T})$, $\mathcal{F}(D_{s1} D_{r2}^\mathsf{T})$, $\mathcal{F}(D_{s2} D_{r1}^\mathsf{T})$ and $\mathcal{F}(D_{s2} D_{r2}^\mathsf{T})$ are Euclidean for $W_r W_s \|z\|^2$;

E7 $\mathcal{F}(y_s H_{r1})$ and $\mathcal{F}(y_s H_{r2})$ are Euclidean for $W_r W_s \|z\|^2$; and

E8 $\mathcal{F}(x_t H_{r1})$ and $\mathcal{F}(x_t H_{r2})$ are Euclidean for $W_s\|z\|^3$.

Let $\mu_n f := \frac{1}{n}\sum_{i=1}^n [f(z_i) - \mathbb{E}f(z_i)]$. Using the preceding facts, and the uniform law of large numbers given as Proposition E.1 below, we may prove

Lemma E.3. *The convergence*

$$\sup_{\beta\in\mathrm{B}} \mu_n|\partial_\beta^l[y_s(\beta,\lambda_n)y_r(\beta,\lambda_n)]| + \sup_{\beta\in\mathrm{B}} \mu_n|x_t\partial_\beta^l y_r(\beta,\lambda_n)| = o_p(1). \tag{E.15}$$

holds for $l = 0$, and also for $l \in \{1,2\}$ if (4.3) holds with $l' = l$.

The first equality in (E.8) now follows, and thus part (ii) of Lemma E.1 is proved.

Proof of Lemma E.2. Suppose $l = 2$; the proof when $l = 1$ is analogous (and is trivial when $l = 0$). Noting that

$$\partial_\beta^2(y_r y_s) = y_s\partial_\beta^2 y_r + (\partial_\beta y_r)(\partial_\beta y_s)^\mathsf{T} + (\partial_\beta y_s)(\partial_\beta y_r)^\mathsf{T} + y_r\partial_\beta^2 y_s, \tag{E.16}$$

it follows from (E.11), (E.12), E6 and E7 that for every $\lambda \in (0,1]$,

$$\|\partial_\beta^2(y_r y_s)\| \lesssim \lambda^{-2}W_r W_s(\|z\|^2 \vee 1),$$

which does not depend on β, and is integrable by L7. (Here $a \lesssim b$ denotes that $a \leq Cb$ for some constant C not depending on b.) Thus by the dominated derivatives theorem, the second equality in

$$\tau_1^{[2]}(\beta,\lambda) = \mathbb{E}\partial_\beta^2\tau_{n1}(\beta,\lambda) = \partial_\beta^2\mathbb{E}\tau_{n1}(\beta,\lambda) = \partial_\beta^2\tau_1(\beta,\lambda)$$

holds for every $\lambda \in (0,1]$; the other equalities follow from the definitions of $\tau_k^{[l]}$ and τ_k. Deduce that, so long as $\lambda_n > 0$ (as per the requirements of Proposition 4.1 above),

$$\tau_1^{[2]}(\beta,\lambda_n) = \partial_\beta^2\tau_1(\beta,\lambda_n) \xrightarrow{p} \partial_\beta^2\tau_1(\beta,0)$$

by the uniform continuity of $\partial_\beta^2\tau_1$ on $\mathrm{B}\times\Lambda$. A similar reasoning – but now using E8 – gives the same result for $\tau_2^{[2]}$. □

The proof of Lemma E.3 requires the following result. Let $\mathcal{G}_{\omega,x}$ denote the σ-field generated by $\eta_\omega(z_i)$ and $x(z_i)$, and let η_ν denote those elements of η that are not present in η_ω. Recall that $\eta_\nu \perp\!\!\!\perp \mathcal{G}_{\omega,x}$.

Lemma E.4. *For every $p \in \{0,1,2\}$, $s,t \in \{1,\dots,d_v\}$, $\mathcal{S} \subseteq \{1,\dots,d_v\}$ and $\mathbb{L} \in \{\mathbb{K}_s,\mathbb{K}_{st}\}$*

$$\mathbb{E}[\|z_{\nu s}\|^p\|z_{\nu t}\|^p\mathbb{L}(\mathcal{S};\beta,\lambda)^2 \mid \mathcal{G}_{\omega,x}] \lesssim \lambda\mathbb{E}[\|z_{\nu s}\|^p\|z_{\nu t}\|^p \mid \mathcal{G}_{\omega,x}]. \tag{E.17}$$

Proof. Note that for any $\mathbb{L} \in \{\mathbb{K}_s,\mathbb{K}_{st}\}$,

$$\mathbb{L}(\mathcal{S};\beta,\lambda) \lesssim L_\lambda[\nu_s(\beta)]$$

where $L(x) = \max\{|\dot{K}(x)|, |\ddot{K}(x)|\}$. Let d denote the dimensionality of η_ν, and fix a $\beta \in \mathrm{B}$. By L5 and L6, there is a $k \in \{1, \dots d\}$, possibly depending on β, and an $\epsilon > 0$ which does not, such that

$$\nu_s(\beta) = \nu_s^*(\beta) + \beta_k^* \eta_{\nu k}$$

with $|\beta_k^*| \geq \epsilon$ and $\nu_s^*(\beta) \perp\!\!\!\perp \eta_{\nu k}$. Let $\mathcal{G}_{\omega,x}^* := \mathcal{G}_{\omega,x} \vee \sigma(\{\eta_{\nu l}\}_{l \neq k})$, so that $\nu_s^*(\beta)$ is $\mathcal{G}_{\omega,x}^*$-measurable, and let f_k denote the density of $\eta_{\nu k}$. Then for any $q \in \{0, \dots, 4\}$,

$$\begin{aligned}
\mathbb{E}\left[|\eta_{\nu k}|^q \mathbb{L}(\mathcal{S}; \beta, \lambda)^2 \mid \mathcal{G}_{\omega,x}^*\right] &\lesssim \mathbb{E}\left[|\eta_{\nu k}|^q L_\lambda^2(\nu_s^*(\beta) + \beta_k^* \eta_{\nu k}) \mid \mathcal{G}_{\omega,x}^*\right] \\
&= \int_{\mathbb{R}} |u|^q L_\lambda^2(\nu_s^*(\beta) + \beta_k^* u) f_k(u) \, \mathrm{d}u \\
&\lesssim (\beta_k^*)^{-1} \lambda \int_{\mathbb{R}} L^2(u) \, \mathrm{d}u \cdot \sup_{u \in \mathbb{R}} |u|^q f_k(u) \\
&\lesssim \epsilon^{-1} \lambda, \qquad \text{(E.18)}
\end{aligned}$$

since $\sup_{u \in \mathbb{R}} |u|^q f_k(u) < \infty$ under L5. Finally, we may partition $z_{\nu s} = (z_{\nu s}^{*\mathsf{T}}, \eta_{\nu k})^\mathsf{T}$ and $z_{\nu t} = (z_{\nu t}^{*\mathsf{T}}, \eta_{\nu k})^\mathsf{T}$, with the possibility that $z_{\nu s} = z_{\nu s}^*$ and $z_{\nu t} = z_{\nu t}^*$. Then by (E.18),

$$\mathbb{E}\left[\|z_{\nu s}\|^p \|z_{\nu t}\|^p \mathbb{L}(\mathcal{S}; \beta, \lambda)^2 \mid \mathcal{G}_{\omega,x}^*\right] \lesssim \lambda \|z_{\nu s}^*\|^p \|z_{\nu t}^*\|^p \leq \lambda \|z_{\nu s}\|^p \|z_{\nu t}\|^p.$$

The result now follows by the law of iterated expectations. □

Proof of Lemma E.3. We shall only provide the proof for first term on the left side of (E.15), when $l = 2$; the proof in all other cases are analogous, requiring appeal only to Proposition E.1 (or Theorem 2.4.3 in van der Vaart and Wellner, 1996, when $l = 0$) and the appropriate parts of E3–E8.

Recalling the decomposition of $\partial_\beta^2(y_r y_s)$ given in (E.16) above, we are led to consider

$$(\partial_\beta y_r)(\partial_\beta y_s)^\mathsf{T} = D_{s1} D_{r1}^\mathsf{T} + \lambda^{-1} D_{s2} D_{r1}^\mathsf{T} + \lambda^{-1} D_{s1} D_{r2}^\mathsf{T} + \lambda^{-2} D_{s2} D_{r2}^\mathsf{T} \qquad \text{(E.19)}$$

and

$$y_s \partial_\beta^2 y_r = \lambda^{-1} y_s H_{r1} + \lambda^{-2} y_s H_{r2}. \qquad \text{(E.20)}$$

Note that by Lemma E.4, and L7

$$\begin{aligned}
\mathbb{E}\|y_s H_{r2}\|^2 &\lesssim \mathbb{E}\left[|\omega_s(\beta)|^2 |\omega_r(\beta)|^2 \sum_{s \in \mathcal{S}_r} \sum_{t \in \mathcal{S}_r} \mathbb{E}\left[\|z_{vs}\|^2 \|z_{vt}\|^2 |\mathbb{K}_{st}(\mathcal{S}_r; \beta, \lambda)|^2 \mid \mathcal{G}_{\omega,x}\right]\right] \\
&\lesssim \lambda \mathbb{E}\left[W_s^2 W_r^2 \sum_{s \in \mathcal{S}_r} \sum_{t \in \mathcal{S}_r} \mathbb{E}\|z_{vs}\|^2 \|z_{vt}\|^2\right] \\
&\lesssim \lambda
\end{aligned}$$

and analogously for each of H_{r1}, $D_{s1} D_{r1}^\mathsf{T}$, $D_{s2} D_{r1}^\mathsf{T}$, $D_{s1} D_{r2}^\mathsf{T}$ and $D_{s2} D_{r2}^\mathsf{T}$. By E6 and E7, the classes formed from these parametrized functions are Euclidean, with envelopes that are p_0-integrable under L7 ($p_0 \geq 2$).

Application of Proposition E.1 to each of the terms in E6 and E7, with λ playing the role of

δ^{-1} there, thus yields the result. Negligibility of the final terms in (E.19) and (E.20) entail the most stringent conditions on the rate at which λ_n may shrink to zero, due to the multiplication of these by λ^{-2}. □

E.1.2 Proof of part (i)

The typical scalar elements of T_n are as displayed in (E.13) above, i.e. they are averages of random functions of the form $\zeta_1(\beta,\lambda) \coloneqq y_r(\beta,\lambda)y_s(\beta,\lambda)$ or $\zeta_2(\beta,\lambda) \coloneqq x_t y_r(\beta,\lambda)$, for $r,s \in \{1,\ldots,d_y\}$ and $t \in \{1,\ldots,d_x\}$. It follows from E3 that $\mathcal{F}(\zeta_1)$ and $\mathcal{F}(\zeta_2)$ are Euclidean, with envelopes $F_1 \coloneqq W_r W_s$ and $F_2 \coloneqq \|z\| W_r$ respectively. Since both envelopes are square integrable under L7, we have

$$\sup_{\mathbb{Q}} N(\epsilon\|F_k\|_{2,\mathbb{Q}}, \mathcal{F}(\zeta_k), L_{2,\mathbb{Q}}) \leq C_1' \epsilon^{-C_2'}$$

for $k \in \{1,2\}$. Hence (E.9) follows by Theorem 2.5.2 in van der Vaart and Wellner (1996).

E.2 A uniform-in-bandwidth law of large numbers

This section provides a uniform law of large numbers (ULLN) for certain classes of parametrized functions, broad enough to cover products involving $K_\lambda[\nu_s(\beta)]$, and such generalizations as appear in Lemma E.4 above. Our ULLN holds *uniformly* in the inverse 'bandwith' parameter $\delta = \lambda^{-1}$; in this respect, it is related to some of the results proved in Einmahl and Mason (2005). However, while their arguments could be adapted to our problem, these would lead to stronger conditions on the bandwidth: in particular, p would have to be replaced by $2p$ in Proposition E.1 below. (On the other hand, their results yield explicit rates of uniform convergence, which are not of concern here.)

Consider the (pointwise measurable) function class

$$\mathcal{F}_\Delta \coloneqq \{z \mapsto f_{(\gamma,\delta)}(z) \mid (\gamma,\delta) \in \Gamma \times \Delta\},$$

and put $\mathcal{F} \coloneqq \mathcal{F}_{[1,\infty)}$. The functions $f_{(\gamma,\delta)} : \mathscr{Z} \to \mathbb{R}^d$ satisfy:

E1 $\sup_{\gamma\in\Gamma} \mathbb{E}\|f_{(\gamma,\delta)}(z_0)\|^2 \lesssim \delta^{-1}$ for every $\delta > 0$.

Let $F : \mathscr{Z} \to \mathbb{R}$ denote an envelope for $\mathcal{F}$, in the sense that

$$\sup_{(\gamma,\delta)\in\Gamma\times[1,\infty)} \|f_{(\gamma,\delta)}(z)\| \leq F(z)$$

for all $z \in \mathscr{Z}$. We will suppose that F may be chosen such that, additionally,

E2 $\mathbb{E}|F(z_0)|^p < \infty$; and

E3 $\sup_{\mathbb{Q}} N(\epsilon\|F\|_{1,\mathbb{Q}}, \mathcal{F}, L_{1,\mathbb{Q}}) \leq C\epsilon^{-d}$ for some $d \in (0,\infty)$.

Let $\{\overline{\delta}_n\}$ denote a real sequence with $\overline{\delta}_n \geq 1$, and $\Delta_n \coloneqq [1,\overline{\delta}_n]$.

Proposition E.1. *Under* E1–E3, *if* $n^{1-1/p}/\overline{\delta}_n^{2m-1} \log(\overline{\delta}_n \vee n) \to \infty$ *for some* $m \geq 1$, *then*

$$\sup_{(\gamma,\delta)\in\Gamma\times\Delta_n} \delta^m \|\mu_n f_{(\gamma,\delta)}\| = o_p(1). \tag{E.21}$$

Remark E.1. Suppose δ_n is an $\mathcal{F}$-measurable sequence for which $n^{1-1/p}/\delta_n^{2m-1}\log(\delta_n \vee n) \xrightarrow{p} \infty$. Then for every $\epsilon > 0$, there exists a deterministic sequence $\{\overline{\delta}_n\}$ satisfying the requirements of Proposition E.1, and for which $\limsup_{n\to\infty} \mathbb{P}\{\delta_n \leq \overline{\delta}_n\} > 1 - \epsilon$. Deduce that

$$\sup_{\gamma\in\Gamma} \delta_n^m \|\mu_n f_{(\gamma,\delta_n)}\| = o_p(1).$$

The proof requires the following

Lemma E.5. *Suppose $\mathcal{F}$ is a (pointwise measurable) class with envelope F, satisfying*

(i) $\|F\|_\infty \leq \tau$;

(ii) $\sup_{f\in\mathcal{F}}\|f\|_{2,\mathbb{P}} \leq \sigma$; *and*

(iii) $\sup_{\mathbb{Q}} N(\epsilon\|F\|_{1,\mathbb{Q}}, \mathcal{F}, L_{1,\mathbb{Q}}) \leq C\epsilon^{-d}$.

Let $\theta := \tau^{-1/2}\sigma$, $m \in \mathbb{N}$ and $x > 0$. Then there exist $C_1, C_2 \in (0,\infty)$, not depending on τ, σ or x, such that

$$\mathbb{P}\left\{\sigma^{-2}\sup_{f\in\mathcal{F}}|\mu_n f| > x\right\} \leq C_1 \exp[-C_2 n\theta^2(1+x^2) + d\log(\theta^{-2}x^{-1})] \tag{E.22}$$

for all $n \geq \frac{1}{8}x^{-2}\theta^{-2}$.

Proof of Proposition E.1. We first note that, by E2,

$$\max_{i\leq n}|F(z_i)| = o_p(n^{-1/p})$$

and so, letting $f^n_{(\gamma,\delta)}(z) := f_{(\gamma,\delta)}(z)\mathbf{1}\{F(z) \leq n^{1/p}\}$, we have

$$\mathbb{P}\left\{\sup_{(\gamma,\delta)\in\Gamma\times\Delta_n} \delta^m|\mu_n[f_{(\gamma,\delta)} - f^n_{(\gamma,\delta)}]| = 0\right\} \leq \mathbb{P}\left\{\max_{i\leq n}|F(z_i)| > n^{1/p}\right\} = o(1).$$

It thus suffices to show that (E.21) holds when $f_{(\gamma,\delta)}$ is replaced by $f^n_{(\gamma,\delta)}$. Since E1 and E3 continue to hold after this replacement, it suffices to prove (E.21) when E2 is replaced by the condition that $\|F\|_\infty \leq n^{1/p}$, which shall be maintained throughout the sequel. (The dependence of f and F upon n will be suppressed for notational convenience.)

Letting $\delta_k := \mathrm{e}^k$, define $\Delta_{nk} := [\delta_k, \delta_{k+1} \wedge \overline{\delta}_n]$ for $k \in \{0,\ldots,K_n\}$, where $K_n := \log\overline{\delta}_n$; observe that $\Delta_n = \bigcup_{k=0}^{K_n}\Delta_{nk}$. Set

$$\mathcal{F}_{nk} := \{z \mapsto f_{(\gamma,\delta)}(z) \mid (\gamma,\delta) \in \Gamma\times\Delta_{nk}\}$$

and note that $\|F\|_\infty \leq n^{1/p}$ and $\sup_{f\in\mathcal{F}_{nk}}\|f\|_{2,\mathbb{P}} \leq \delta_k^{-1/2}$. Under E3, we may apply apply Lemma E.5 to each $\mathcal{F}_{nk}$, with $(\tau,\sigma) = (n^{1/p}, \delta_k^{-1/2})$ and $x = \delta_k^{1-m}\epsilon$, for some $\epsilon > 0$. There thus exist $C_1, C_2 \in (0,\infty)$ depending on ϵ such that

$$\mathbb{P}\left\{\sup_{(\gamma,\delta)\in\Gamma\times\Delta_n} \delta^m |\mu_n f_{(\gamma,\delta)}| > \epsilon\right\} \leq \sum_{k=0}^{K_n} \mathbb{P}\left\{\delta_k^m \sup_{(\gamma,\delta)\in\Gamma\times\Delta_{nk}} |\mu_n f_{(\gamma,\delta)}| > \mathrm{e}^{-1}\epsilon\right\}$$
$$\leq C_1 \sum_{k=0}^{K_n} \exp[-C_2 n\theta_{nk}^2 \delta_k^{2(1-m)} + d\log(\theta_{nk}^{-2}\delta_k^{m-1})] \tag{E.23}$$

where $\theta_{nk} \coloneqq n^{-1/2p}\delta_k^{-1/2}$, provided

$$n \geq \tfrac{1}{8}\delta_k^{2(m-1)}\theta_{nk}^{-2}\epsilon^{-2}, \ \forall k \in \{0,\ldots,K_n\} \impliedby n^{1-1/p}/\overline{\delta}_n^{2m-1} \geq \tfrac{1}{8}\epsilon^{-2}, \tag{E.24}$$

which holds for all n sufficiently large. In obtaining (E.24) we have used $\delta_k \leq \overline{\delta}_n$ and $\theta_{nk} \geq n^{-1/2p}\overline{\delta}_n^{-1/2}$, and these further imply that (E.23) may be bounded by

$$C_1(\log\overline{\delta}_n)\exp[-C_2 n^{1-1/p}\overline{\delta}_n^{-2m-1}(1+\epsilon^2) + d\log(\overline{\delta}_n^m n^{1/p})] \to 0$$

as $n \to \infty$. Thus (E.21) holds. □

Proof of Lemma E.5. Suppose (iii) holds. Define $\mathcal{G} \coloneqq \{\tau^{-1}f \mid f \in \mathcal{F}\}$, and $G \coloneqq \tau^{-1}F$. Then

$$\sup_{g\in\mathcal{G}}\|g\|_{2,\mathbb{P}} \leq \tau^{-1}\sup_{f\in\mathcal{F}}\|f\|_{2,\mathbb{P}} \leq \tau^{-1/2}\sigma \eqqcolon \theta;$$

$\|g\|_\infty \leq 1$ for all $g \in \mathcal{G}$; and since $\|G_n\|_{1,\mathbb{Q}} \leq 1$, $N(\epsilon,\mathcal{G},L_{1,\mathbb{Q}}) \leq C\epsilon^{-d}$. Hence, by arguments given in the proof of Theorem II.37 in Pollard (1984), there exist $C_1, C_2 > 0$, depending on x, such that

$$\mathbb{P}\left\{\sigma^{-2}\sup_{f\in\mathcal{F}}|\mu_n f| > x\right\} = \mathbb{P}\left\{\sup_{g\in\mathcal{G}}|\mu_n g| > \theta^2 x\right\} \leq C_1\exp[-C_2 n\theta^2(1+x^2) + d\log(\theta^{-2}x^{-1})]$$

for all $n \geq \frac{1}{8}x^{-2}\theta^{-2}$. □