Supplement to "Estimation and inference with a (nearly) singular Jacobian"

(*Quantitative Economics*, Vol. 10, No. 3, July 2019, 1019–1068)

Sukjin Han
Department of Economics, University of Texas at Austin

Adam McCloskey
Department of Economics, University of Colorado at Boulder

This Online Appendix provides, in Section A, the proofs of the main results of Han and McCloskey (Forthcoming) and, in Section B, a proof of the claim that the quantity $\|\tilde{\eta}(\hat{\mu}_n)\|$ diverges in the context of the threshold crossing model made in Section 4 of Han and McCloskey (Forthcoming) as well as a set of primitive conditions, and corresponding verification of the high level assumptions made in Han and McCloskey (Forthcoming) for the threshold crossing model.

Online Appendix A: Proofs of main results

Proof of Theorem 3.1. When $\beta = 0$,

$$\frac{\partial Q_n(\theta)}{\partial \boldsymbol{\pi}'} = \frac{\partial \Psi_n}{\partial \boldsymbol{g}'} \frac{\partial \bar{\boldsymbol{g}}_n(\beta, h(\mu))}{\partial \boldsymbol{\pi}'} = \frac{\partial \Psi_n}{\partial \boldsymbol{g}'} \frac{\partial \bar{\boldsymbol{g}}_n(\beta, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}'} \frac{\partial h(\mu)}{\partial \boldsymbol{\pi}'} = 0_{1\times d_\pi}$$

for all $\theta = (0, \mu) \in \Theta \equiv \{(\beta, \mu) \in \mathbb{R}^{d_\theta} : (\beta, h(\mu)) \in \boldsymbol{\Theta}\}$ since $\partial h(\mu)/\partial \boldsymbol{\pi}'$ is in the null space of $\partial \bar{\boldsymbol{g}}_n(\beta, \boldsymbol{\mu})/\partial \boldsymbol{\mu}'$ by Assumption Jac. □

Proof of Theorem 3.2. First, note that

$$\frac{\partial g^{(1)}(0, \mu^{(1)})}{\partial \boldsymbol{\pi}_1^{(1)}} = \frac{\partial \boldsymbol{g}^*(0, h^{(1)}(\mu^{(1)}))}{\partial \boldsymbol{\pi}_1^{(1)}} = \left.\frac{\partial \boldsymbol{g}^*(0, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=h^{(1)}(\mu^{(1)})} \times \frac{\partial h^{(1)}(\mu^{(1)})}{\partial \boldsymbol{\pi}_1^{(1)}} = 0$$

by Steps 1 and 2. By way of induction, for $1 \le i-1 \le d_\pi - 1$, assume that the first $i-1$ columns of $\partial g^{(i-1)}(0, \mu^{(i-1)})/\partial \boldsymbol{\pi}^{(i-1)\prime}$ are equal to zero. Then by Step 8 of the algorithm,

$$\frac{\partial g^{(i)}(0, \mu^{(i)})}{\partial \boldsymbol{\pi}^{(i)\prime}}$$
$$= \frac{\partial g^{(i-1)}(0, h^{(i)}(\mu^{(i)}))}{\partial \boldsymbol{\pi}^{(i)\prime}} = \left.\frac{\partial g^{(i-1)}(0, \mu^{(i-1)})}{\partial \mu^{(i-1)\prime}}\right|_{\mu^{(i-1)}=h^{(i)}(\mu^{(i)})} \times \frac{\partial h^{(i)}(\mu^{(i)})}{\partial \boldsymbol{\pi}^{(i)\prime}}$$

Sukjin Han: sukjin.han@austin.utexas.edu
Adam McCloskey: adam.mccloskey@colorado.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE989

$$= \left[\frac{\partial g^{(i-1)}(0,\mu^{(i-1)})}{\partial \zeta^{(i-1)\prime}} : \frac{\partial g^{(i-1)}(0,\mu^{(i-1)})}{\partial (\pi_1^{(i-1)},\ldots,\pi_{i-1}^{(i-1)})} : \frac{\partial g^{(i-1)}(0,\mu^{(i-1)})}{\partial (\pi_i^{(i-1)},\ldots,\pi_{d_\pi}^{(i-1)})} \right] \Bigg|_{\mu^{(i-1)}=h^{(i)}(\mu^{(i)})}$$

$$\times \left[\frac{\partial h^{(i)}(\mu^{(i)})}{\partial (\pi_1^{(i)},\ldots,\pi_{i-1}^{(i)})} : \frac{\partial h^{(i)}(\mu^{(i)})}{\partial \pi_i^{(i)}} : \frac{\partial h^{(i)}(\mu^{(i)})}{\partial (\pi_{i+1}^{(i)},\ldots,\pi_{d_\pi}^{(i)})} \right]$$

$$= \left[\frac{\partial g^{(i-1)}(0,\mu^{(i-1)})}{\partial \zeta^{(i-1)\prime}} : 0_{d_g\times(i-1)} : \frac{\partial g^{(i-1)}(0,\mu^{(i-1)})}{\partial (\pi_i^{(i-1)},\ldots,\pi_{d_\pi}^{(i-1)})} \right] \Bigg|_{\mu^{(i-1)}=h^{(i)}(\mu^{(i)})}$$

$$\times \begin{bmatrix} \begin{matrix} 0_{(d_{\boldsymbol{\mu}}-d_\pi)\times(i-1)} \\ C^{(i)}(\mu^{(i)}) \\ 0_{(d_\pi-i+1)\times(i-1)} \end{matrix} : \frac{\partial h^{(i)}(\mu^{(i)})}{\partial \pi_i^{(i)}} : \frac{\partial h^{(i)}(\mu^{(i)})}{\partial (\pi_{i+1}^{(i)},\ldots,\pi_{d_\pi}^{(i)})} \end{bmatrix}$$

$$= \Bigg[0_{d_{g\times(i-1)}} : \frac{\partial g^{(i-1)}(0,\mu^{(i-1)})}{\partial \mu^{(i-1)\prime}} \frac{\partial h^{(i)}(\mu^{(i)})}{\partial \pi_i^{(i)}} :$$

$$\frac{\partial g^{(i-1)}(0,\mu^{(i-1)})}{\partial \mu^{(i-1)\prime}} \frac{\partial h^{(i)}(\mu^{(i)})}{\partial (\pi_{i+1}^{(i)},\ldots,\pi_{d_\pi}^{(i)})} \Bigg] \Bigg|_{\mu^{(i-1)}=h^{(i)}(\mu^{(i)})}$$

$$= \left[0_{d_{g\times(i-1)}} : 0_{d_g\times 1} : \frac{\partial g^{(i-1)}(0,\mu^{(i-1)})}{\partial \mu^{(i-1)\prime}} \frac{\partial h^{(i)}(\mu^{(i)})}{\partial (\pi_{i+1}^{(i)},\ldots,\pi_{d_\pi}^{(i)})} \right] \Bigg|_{\mu^{(i-1)}=h^{(i)}(\mu^{(i)})},$$

where the third equality results from the definition of $\mu^{(i)}$ in Step 6, the fourth equality follows from Step 7 and the final equality follows from Steps 5 and 6.

Hence, we have shown that for $1 \le i \le d_\pi$, the first i columns of $\partial g^{(i)}(0,\mu^{(i)})/\partial \pi^{(i)\prime}$ are equal to zero. In particular, $\partial g^{(d_\pi)}(0,\mu^{(d_\pi)})/\partial \pi^{(d_\pi)\prime} = 0_{d_g\times d_\pi}$. Also note that Step 8 defines θ as equal to $(\beta,\mu^{(d_\pi)})$ and

$$\begin{aligned} g^*(\theta) &= \boldsymbol{g}^*\big(\beta, h^{(1)}\circ\cdots\circ h^{(d_\pi)}(\mu^{(d_\pi)})\big) = g^{(1)}\big(\beta, h^{(2)}\circ\cdots\circ h^{(d_\pi)}(\mu^{(d_\pi)})\big) \\ &= g^2\big(\beta, h^{(3)}\circ\cdots\circ h^{(d_\pi)}(\mu^{(d_\pi)})\big) = \cdots = g^{(d_\pi)}\big(\beta,\mu^{(d_\pi)}\big), \end{aligned}$$

where the first equality follows from the definition of h in Step 8, the second equality follows from the definition of $g^{(1)}(\theta^{(1)})$ in Step 4 and the final two equalities follow from the definition of $g^{(i)}(\theta^{(i)})$ in Step 8. Thus for $\beta = 0$, using the definition of $h(\cdot)$ in Step 8, we have

$$\begin{aligned} \begin{bmatrix} \cdots & 0_{1\times d_\pi} \\ \vdots & \vdots \\ \cdots & 0_{1\times d_\pi} \end{bmatrix} &= \frac{\partial g^{(d_\pi)}(\theta^{(d_\pi)})}{\partial \mu^{(d_\pi)\prime}} = \frac{\partial \boldsymbol{g}^*\big(\beta, h^{(1)}\circ\cdots\circ h^{(d_\pi)}(\mu^{(d_\pi)})\big)}{\partial \mu^{(d_\pi)\prime}} \\ &= \frac{\partial \boldsymbol{g}^*(\beta, h(\mu))}{\partial \mu'} = \frac{\partial \boldsymbol{g}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\mu}'}\bigg|_{\boldsymbol{\theta}=(\beta,h(\mu))} \times \frac{\partial h(\mu)}{\partial \mu'} \end{aligned}$$

so that $h : \mathcal{M} \to \boldsymbol{\mathcal{M}}$ satisfies Procedure 3.1 if it is one-to-one. This latter property holds because each $\partial h^{(i)}(\mu^{(i)})/\partial \mu^{(i)\prime}$ for $i = 1,\ldots,d_\pi$ has full rank by Steps 3 and 7 and $h = h^{(1)}\circ\cdots\circ h^{(d_\pi)}$ by Step 8. □

Proof of Proposition 3.1. First, when $\beta = 0$, under Assumption ID, there exists at least one column in $J^*(\boldsymbol{\theta})$ that is linearly dependent on the other columns wp1, which implies that there exists a nonzero vector $m^{(1)}$ such that (3.3) holds. Thus, (3.4) is a well-defined system of ODEs with an initial condition that is determined by constants of integration. By the (global) Picard–Lindelöf theorem (Picard (1893), Lindelöf (1894)), since $m^{(1)}(\cdot)$ is Lipschitz continuous on compact $\mathcal{M}^{(1)}$, there exists a solution $h^{(1)}$ on $\mathcal{M}^{(1)}$ of (3.4). Since the choice of constants of integration for this solution does not affect (3.4), it is always possible to choose them to ensure full rank of $\partial h^{(1)}(\mu^{(1)})/\partial\mu^{(1)\prime}$. Now by way of induction, for $1 \le i-1 \le d_\pi - 1$, since $\partial h^{(i)}(\mu^{(i)})/\partial\mu^{(i)\prime}$ is full rank and $\operatorname{rank}(\partial g^{(i-1)}(\theta^{(i-1)})/\partial\mu^{(i-1)\prime}) = r$, it follows that

$$\operatorname{rank}\left(\frac{\partial g^{(i)}\big(\theta^{(i)}\big)}{\partial\mu^{(i)\prime}}\right) = \operatorname{rank}\left(\frac{\partial g^{(i-1)}\big(\theta^{(i-1)}\big)}{\partial\mu^{(i-1)\prime}}\frac{\partial h^{(i)}\big(\mu^{(i)}\big)}{\partial\mu^{(i)\prime}}\right) = r.$$

Thus, there exists a nonzero vector $m^{(i)}$ such that (3.5) holds. Given (3.6), since $m^{(i)}(\cdot)$ is Lipschitz continuous on compact $\mathcal{M}^{(i)}$, there exists a solution $h^{(i)}$ on $\mathcal{M}^{(i)}$. Similarly to before, since the choice of constants of integration for this solution does not affect (3.6), it is always possible to choose them to ensure (1) and (2) of Step 7 hold. Therefore, $h = h^{(1)} \circ \cdots \circ h^{(d_\pi)}$ exists on $\mathcal{M} = \mathcal{M}^{(d_\pi)}$. □

Proof of Lemma 4.1. Define $\bar{h}(\theta) \equiv (\beta, h(\mu))$. For any $\mu \in \mathcal{M}$, since $M(h(\mu))$ has full rank, $\partial h(\mu)/\partial\mu'$ has full rank by Step 2 of Procedure 3.1. Therefore,

$$\frac{\partial\bar{h}(\theta)}{\partial\theta'} = \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{\partial h(\mu)}{\partial\mu} \end{bmatrix}$$

has full rank for any $\theta \in \Theta$. Also, since $h : \mathcal{M} \to \boldsymbol{\mathcal{M}}$ is proper, $\bar{h} : \Theta \to \boldsymbol{\Theta}$ is also proper. Combining these results with Assumption H(ii), we can apply Hadamard's global inverse function theorem Hadamard (1906a,b) to $\bar{h} : \Theta \to \boldsymbol{\Theta}$, and conclude that $\bar{h}$ is a homeomorphism. □

Proof of Lemma 4.2. Suppose Assumption Reg3*(v) holds. Without loss of generality, we may permute the elements of μ^s so that

$$h_\pi^s(\mu) = \begin{pmatrix} 0_{(d_s-\tilde{d}_\pi^*)\times\tilde{d}_\pi^*} & 0_{(d_s-\tilde{d}_\pi^*)\times(d_\pi-\tilde{d}_\pi^*)} \\ D(\mu) & 0_{\tilde{d}_\pi^*\times(d_\pi-\tilde{d}_\pi^*)} \end{pmatrix},$$

where $D(\mu)$ is a diagonal full rank $\tilde{d}_\pi^* \times \tilde{d}_\pi^*$ matrix. By definition, the column space of $h_\pi^s(\mu)$ is equal to

$$\begin{aligned} &\{v : v = h_\pi^s(\mu)x \text{ for some } x \in \mathbb{R}^{d_\pi}\} \\ &= \{(0_{1\times(d_s-\tilde{d}_\pi^*)}, v_2')' : v_2 \in \mathbb{R}^{\tilde{d}_\pi^*} \text{ and for each} \\ &\quad i = 1, \dots, \tilde{d}_\pi^*,\ v_{2,i} = D_{ii}(\mu)x_i \text{ for some } x_i \in \mathbb{R}\} \\ &= \{(0_{1\times(d_s-\tilde{d}_\pi^*)}, x_2)' : x_2 \in \mathbb{R}^{\tilde{d}_\pi^*}\}, \end{aligned}$$

which clearly satisfies the condition in Assumption Reg3*(iii) since it does not depend upon μ. □

The proofs of Theorem 4.1, Corollary 4.1, and Proposition 5.1 make use of the following auxiliary lemmas. The following lemma applies some of the main results of AC12.

LEMMA A.1. (i) *Suppose Assumptions* ID, CF, Reg1 *and* Jac, *and Assumptions* B1–B3 *and* C1–C6 *of* AC12, *applied to the* θ *and* $Q_n(\theta)$ *of this paper, hold. Under parameter sequences* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ *with* $\|b\| < \infty$,

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_n - \beta_n) \\ \sqrt{n}(\hat{\zeta}_n - \zeta_n) \\ \hat{\pi}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \tau^{\beta}_{0,b}(\pi^*_{0,b}) \\ \tau^{\zeta}_{0,b}(\pi^*_{0,b}) \\ \pi^*_{0,b} \end{pmatrix}.$$

(ii) *Suppose Assumptions* ID, CF, Reg1 *and* Jac, *and Assumptions* B1–B3, C1–C5, C7–C8 *and* D1–D3 *of* AC12, *applied to the* θ *and* $Q_n(\theta)$ *of this paper, hold. Under parameter sequences* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_n - \beta_n \\ \hat{\zeta}_n - \zeta_n \\ \iota(\beta_n)(\hat{\pi}_n - \pi_n) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_\beta \\ Z_\zeta \\ Z_\pi \end{pmatrix}.$$

PROOF. Theorem 3.1 directly implies that Assumption A of AC12 holds when applied to the θ and $Q_n(\theta)$ of this paper. Then (i) and (ii) follow by direct application of Theorems 3.1(a) and 3.2(a) of AC12. □

The next lemma ensures we can write $\hat{\boldsymbol{\theta}}_n = (\hat{\beta}_n, h(\hat{\mu}_n))$.

LEMMA A.2. *Suppose Assumptions* ID, Jac *and* H *hold. Then* $\hat{\boldsymbol{\theta}}_n = (\hat{\beta}_n, h(\hat{\mu}_n))$ *for some* $\hat{\theta}_n = (\hat{\beta}_n, \hat{\mu}_n) \in \Theta$ *such that* $Q_n(\hat{\theta}_n) = \inf_{\theta \in \Theta} Q_n(\theta) + o(n^{-1})$.

PROOF. The reparameterization function $\bar{h} : \Theta \to \boldsymbol{\Theta}$ is bijective by Lemma 4.1, which implies $\boldsymbol{\Theta} = \bar{h}(\Theta)$ and $\Theta = h^{-1}(\boldsymbol{\Theta})$ so that

$$\begin{aligned} \boldsymbol{Q}_n(\hat{\boldsymbol{\theta}}_n) &= \inf_{\boldsymbol{\theta} \in \bar{h}(\Theta)} \boldsymbol{Q}_n(\boldsymbol{\theta}) + o(n^{-1}) = \inf_{\bar{h}^{-1}(\boldsymbol{\theta}) \in \Theta} \boldsymbol{Q}_n\big(\bar{h}(\bar{h}^{-1}(\boldsymbol{\theta}))\big) + o(n^{-1}) \\ &= \inf_{\bar{h}^{-1}(\boldsymbol{\theta}) \in \Theta} Q_n\big(\bar{h}^{-1}(\boldsymbol{\theta})\big) + o(n^{-1}) \\ &= \inf_{\theta \in \Theta} Q_n(\theta) + o(n^{-1}) = Q_n(\hat{\theta}_n) \end{aligned}$$

for some $\hat{\theta}_n \in \Theta$. □

PROOF OF THEOREM 4.1. (i) Using Lemma A.2, begin by decomposing $\hat{\boldsymbol{\mu}}_n^s - \boldsymbol{\mu}_n^s = h^s(\hat{\mu}_n) - h^s(\mu_n)$ as follows:

$$\begin{aligned} h^s(\hat{\mu}_n) - h^s(\mu_n) &= \big[h^s(\hat{\zeta}_n, \hat{\pi}_n) - h^s(\zeta_n, \hat{\pi}_n)\big] + \big[h^s(\zeta_n, \hat{\pi}_n) - h^s(\zeta_n, \pi_n)\big] \\ &= h^s_\zeta(\hat{\mu}_n)(\hat{\zeta}_n - \zeta_n) + \big[h^s(\zeta_n, \hat{\pi}_n) - h^s(\zeta_n, \pi_n)\big] + o_p\big(n^{-1/2}\big), \end{aligned}$$

where the second equality uses a mean value expansion (with respect to ζ) that holds by Lemma A.1(i) and Lemma 4.1(ii). Using this decomposition, we have

$$\begin{aligned} \begin{pmatrix} \sqrt{n}(\hat{\beta}_n - \beta_n) \\ \sqrt{n}\tilde{A}_1(\hat{\mu}_n)\big(\hat{\boldsymbol{\mu}}_n^s - \boldsymbol{\mu}_n^s\big) \\ \tilde{A}_2(\hat{\mu}_n)\big(\hat{\boldsymbol{\mu}}_n^s - \boldsymbol{\mu}_n^s\big) \end{pmatrix} &= \begin{pmatrix} \sqrt{n}(\hat{\beta}_n - \beta_n) \\ \sqrt{n}\tilde{A}_1(\hat{\mu}_n)h^s_\zeta(\hat{\mu}_n)(\hat{\zeta}_n - \zeta_n) \\ \tilde{A}_2(\hat{\mu}_n)\big[h^s(\zeta_n, \hat{\pi}_n) - h^s(\zeta_n, \pi_n)\big] \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 \\ \sqrt{n}\tilde{A}_1(\hat{\mu}_n)\big[h^s(\zeta_n, \hat{\pi}_n) - h^s(\zeta_n, \pi_n)\big] \\ \tilde{A}_2(\hat{\mu}_n)h^s_\zeta(\hat{\mu}_n)(\hat{\zeta}_n - \zeta_n) \end{pmatrix} + o_p(1) \\ &= \begin{pmatrix} \sqrt{n}(\hat{\beta}_n - \beta_n) \\ \tilde{A}_1(\hat{\mu}_n)h^s_\zeta(\hat{\mu}_n)\sqrt{n}(\hat{\zeta}_n - \zeta_n) + \tilde{\eta}^*_{0,b} \\ \tilde{A}_2(\hat{\mu}_n)\big[h^s(\zeta_n, \hat{\pi}_n) - h^s(\zeta_n, \pi_n)\big] \end{pmatrix} + o_p(1) \\ &\xrightarrow{d} \begin{pmatrix} \tau^\beta_{0,b}(\pi^*_{0,b}) \\ \tilde{A}_1\big(\zeta_0, \pi^*_{0,b}\big)h^s_\zeta\big(\zeta_0, \pi^*_{0,b}\big)\tau^\zeta_{0,b}\big(\pi^*_{0,b}\big) + \tilde{\eta}^*_{0,b} \\ \tilde{A}_2\big(\zeta_0, \pi^*_{0,b}\big)\big[h^s\big(\zeta_0, \pi^*_{0,b}\big) - \boldsymbol{\mu}_0^s\big] \end{pmatrix} \end{aligned}$$

under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$, where the second equality follows from Assumptions Reg2 and Reg3, Lemma A.1(i) and the CMT and the weak convergence follows from Assumption Reg2, Lemma A.1(i), the CMT and the fact that $h^s(\zeta_0, \pi_0) = \boldsymbol{\mu}_0^s$.

(ii) For the $\beta_0 = 0$ case, the same decomposition of $\hat{\boldsymbol{\mu}}_n^s - \boldsymbol{\mu}_n^s = h^s(\hat{\mu}_n) - h^s(\mu_n)$ as that used in the proof of part (i) and similar reasoning imply

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_n - \beta_n \\ \tilde{A}_1(\hat{\mu}_n)\big(\hat{\boldsymbol{\mu}}_n^s - \boldsymbol{\mu}_n^s\big) \\ \iota(\beta_n)\tilde{A}_2(\hat{\mu}_n)\big(\hat{\boldsymbol{\mu}}_n^s - \boldsymbol{\mu}_n^s\big) \end{pmatrix} = \begin{pmatrix} \sqrt{n}(\hat{\beta}_n - \beta_n) \\ \tilde{A}_1(\hat{\mu}_n)h^s_\zeta(\hat{\mu}_n)\sqrt{n}(\hat{\zeta}_n - \zeta_n) \\ \tilde{A}_2(\hat{\mu}_n)\sqrt{n}\iota(\beta_n)\big[h^s(\zeta_n, \hat{\pi}_n) - h^s(\zeta_n, \pi_n)\big] \end{pmatrix} + o_p(1).$$

A mean-value expansion, Lemma 4.1(ii) and the consistency of $\hat{\mu}_n$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ given by Lemma A.1(ii) provide that

$$\begin{aligned} &\tilde{A}_2(\hat{\mu}_n)\sqrt{n}\iota(\beta_n)\big[h^s(\zeta_n, \hat{\pi}_n) - h^s(\zeta_n, \pi_n)\big] \\ &\quad = \tilde{A}_2(\hat{\mu}_n)\sqrt{n}\iota(\beta_n)\big[\big(h^s_\pi(\zeta_n, \hat{\pi}_n) + o_p(1)\big)(\hat{\pi}_n - \pi_n)\big] \\ &\quad = \tilde{A}_2(\hat{\mu}_n)h^s_\pi(\zeta_n, \hat{\pi}_n)\sqrt{n}\iota(\beta_n)(\hat{\pi}_n - \pi_n) + o_p(1), \end{aligned}$$

where the second equality follows from Lemma 4.1(ii) and Lemma A.1(ii). Putting these results together, we have

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_n-\beta_n\\ \tilde{A}_1(\hat{\mu}_n)\big(\hat{\boldsymbol{\mu}}_n^s-\boldsymbol{\mu}_n^s\big)\\ \iota(\beta_n)\tilde{A}_2(\hat{\mu}_n)\big(\hat{\boldsymbol{\mu}}_n^s-\boldsymbol{\mu}_n^s\big)\end{pmatrix}\xrightarrow{d}\begin{pmatrix}Z_\beta\\ \tilde{A}_1(\mu_0)h_\zeta^s(\mu_0)Z_\zeta\\ \tilde{A}_2(\mu_0)h_\pi^s(\mu_0)Z_\pi\end{pmatrix}$$

by Assumption Reg2, Lemma A.1(ii) and the CMT. Finally, for the $\beta_0\neq 0$ case, note that a standard mean value expansion for $\hat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}_n=h(\hat{\mu}_n)-h(\mu_n)$, Lemma 4.1(ii), Lemma A.1(ii) and the CMT imply

$$\begin{aligned}\sqrt{n}\begin{pmatrix}\hat{\beta}_n-\beta_n\\ \hat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}_n^s\end{pmatrix}&=\begin{pmatrix}\sqrt{n}(\hat{\beta}_n-\beta_n)\\ \sqrt{n}h_\mu(\hat{\mu}_n)(\hat{\mu}_n-\mu_n)\end{pmatrix}+o_p(1)\\ &=\begin{pmatrix}\sqrt{n}(\hat{\beta}_n-\beta_n)\\ h_\zeta(\hat{\mu}_n)\sqrt{n}(\hat{\zeta}_n-\zeta_n)+h_\pi(\hat{\mu}_n)\sqrt{n}(\hat{\pi}_n-\pi_n)\end{pmatrix}+o_p(1)\\ &\xrightarrow{d}\begin{pmatrix}Z_\beta\\ h_\zeta(\mu_0)Z_\zeta+\iota(\beta_0)^{-1}h_\pi(\mu_0)Z_\pi\end{pmatrix}.\end{aligned}$$

□

Proof of Corollary 4.1. For case (i),

$$\sqrt{n}\big(\hat{\boldsymbol{\mu}}_n^1-\boldsymbol{\mu}_n^1\big)=\sqrt{n}\big[h^1(\hat{\zeta}_n)-h^1(\zeta_n)\big]=h_\zeta^1(\hat{\zeta}_n)\sqrt{n}(\hat{\zeta}_n-\zeta_n)+o_p(1)\xrightarrow{d}h_\zeta^1(\zeta_0)\tau_{0,b}^\zeta\big(\pi_{0,b}^*\big),$$

where the first equality follows from Lemma A.2, the second equality follows from the mean value theorem, Lemma 4.1(ii) and Lemma A.1(i) and the weak convergence follows from the CMT, Lemma 4.1(ii) and Lemma A.1(i). The results for $\hat{\beta}_n$, $\hat{\boldsymbol{\mu}}_n^2$ and the joint convergence of the three components follow directly from Lemmas A.2 and A.1(i), Lemma 4.1(ii) and the CMT.

For case (ii), note that

$$\begin{aligned}\sqrt{n}\iota(\beta_n)(\hat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}_n)&=\sqrt{n}\iota(\beta_n)\big[h(\hat{\zeta}_n,\hat{\pi}_n)-h(\zeta_n,\pi_n)\big]\\ &=\sqrt{n}\iota(\beta_n)\big[h(\hat{\zeta}_n,\hat{\pi}_n)-h(\zeta_n,\hat{\pi}_n)\big]+\sqrt{n}\iota(\beta_n)\big[h(\zeta_n,\hat{\pi}_n)-h(\zeta_n,\pi_n)\big]\\ &=\sqrt{n}\iota(\beta_n)\big[h_\zeta(\hat{\mu}_n)(\hat{\zeta}_n-\zeta_n)+o_p\big(n^{-1/2}\big)\big]\\ &\quad+\sqrt{n}\iota(\beta_n)\big[h_\pi(\zeta_n,\hat{\pi}_n)(\hat{\pi}_n-\pi_n)+o_p\big(n^{-1/2}\iota(\beta_n)^{-1}\big)\big]\\ &=h_\pi(\zeta_n,\hat{\pi}_n)\sqrt{n}\iota(\beta_n)(\hat{\pi}_n-\pi_n)+o_p(1)\xrightarrow{d}h_\pi(\mu_0)Z_\pi,\end{aligned}$$

where the first equality follows from Lemma A.2, the third equality follows from the mean value theorem, Lemma 4.1(ii) and Lemma A.1(ii), while the final equality and weak convergence result follow from the CMT, Lemma 4.1(ii) and Lemma A.1(ii). Nearly identical arguments to those used for case (i) provide that $\sqrt{n}(\hat{\boldsymbol{\mu}}_n^1-\boldsymbol{\mu}_n^1)\xrightarrow{d}h_\zeta^1(\zeta_0)Z_\zeta$. Joint convergence of the three components immediately follows from Lemma A.1(ii). □

PROOF OF PROPOSITION 6.1. The proof is nearly identical to the proof of Theorem 5.1(b)(iv) of AC12, using Proposition 5.1 in the place of Theorems 4.2 and 4.3 of AC12. □

PROOF OF PROPOSITION 6.2. The proof of this proposition verifies that the assumptions of Theorem Bonf-Adj of McCloskey (2017) hold, with some modifications. First, by Proposition 5.1 and Assumption FD, Assumption PS of McCloskey (2017) holds with $\gamma_1 = (\beta, \pi)$ and $\gamma_2 = (\zeta, \delta)$. For the definition of $\{\gamma_{n,h}\}$, $\gamma_{n,h,1} = (\beta_{n,h}, n^{-1/2}\pi_{n,h})$ and $\gamma_{n,h,2} = (\zeta_{n,h}, \delta_{n,h})$. Note that $h_{1,1} = b$, where $h_{1,1}$ denotes the first d_β elements of h_1. In the notation of McCloskey (2017), sequences $\{\gamma_{n,h}\}$ with $\|h_{1,1}\| < \infty$ ($\|h_{1,1}\| = \infty$) correspond to weak (semi-strong or strong) identification sequences $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$ ($\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$) in the notation of this paper.

Second, for Assumption DS of McCloskey (2017), $T_n(\theta_n) = W_n(v_n)$ $\hat{h}_{n,1} = (\hat{b}_n, \hat{\pi}_n)$ and $\hat{h}_{n,2} = (\hat{\zeta}_n, \hat{\delta}_n)$. Proposition 5.1 provides the marginal weak convergence of $T_{\omega_n}(\theta_{\omega_n})$ for all sequences $\{\gamma_{\omega_n,h}\}$, where in the notation of McCloskey (2017), $W_h = \lambda(\pi^*_{0,b}; \gamma_0, b)$ when $\|h_{1,1}\| < \infty$ and W_h is distributed $\chi^2_{d_r}$ when $\|h_{1,1}\| = \infty$. Lemma A.1 and Assumption FD provide the marginal weak convergence of $\hat{h}_{\omega_n} = (\hat{h}_{\omega_n,1}, \hat{h}_{\omega_n,2})$ for all sequences $\{\gamma_{\omega_n,h}\}$, where in the notation of McCloskey (2017), $\tilde{h}_1 = (b + \tau^\beta_{0,b}(\pi^*_{0,b}), \pi^*_{0,b})$ when $\|h_{1,1}\| < \infty$, $\tilde{h}_1 = (b + Z_\beta, \pi_0)$ when $\|h_{1,1}\| = \infty$ and $h_2 = (\zeta_0, \delta_0)$. Joint convergence of $(T_{\omega_n}(\theta_{\omega_n}), \hat{h}_{\omega_n})$ follows from nearly identical arguments for joint convergence to those used in the proof of Theorem 5.1 of AC14.

Third, for Definition MLLD of McCloskey (2017), we are in what McCloskey (2017) refers to as "the usual case" for which $u = 1$, $\tilde{W}^{(1)}_h = \lambda(\pi^*_{0,b}; \gamma_0, b)$ and $\bar{H}^{(1),c} = \emptyset$ since $P(|\lambda(\pi^*_{0,b}; \gamma_0, b)| < \infty) = 1$ under the assumptions of Proposition 5.1. Since we are in the usual case, there is no need to define the auxiliary sequence of parameters $\{\zeta_n\}$ in that assumption (it can be any arbitrary sequence in $\mathbb{R}^r$ for arbitrary $r > 0$) and $P = \mathbb{R}^r_\infty$ for any $r > 0$. Since $W_h = \lambda(\pi^*_{0,b}; \gamma_0, b) = \tilde{W}^{(1)}_h$ when $\|h_{1,1}\| < \infty$ and $W_h = \tilde{W}^{(1)}_h$ is distributed $\chi^2_{d_r}$ when $\|h_{1,1}\| = \infty$, the only item left to verify is that $\lambda(\pi^*_{0,b}; \gamma_0, b)$ is completely characterized by $h^{(1)} = h = (b, \pi_0, \zeta_0, \delta_0)$. This holds by Assumption FD.

Fourth, for Assumption Cont-Adj of McCloskey (2017), $\bar{H}^{(1)} = H$. This assumption holds for any $\underline{\delta}^{(1)} > 0$ and $\bar{\delta}^{(1)} \leq \alpha$ since $\lambda(\pi^*_{0,b}; \gamma_0, b)$ is an absolutely continuous random variable with quantiles that are continuous in b and π_0 and $\lambda(\pi^*_{0,b}; \gamma_0, b) \overset{d}{\sim} \chi^2_{d_r}$ for any b such that $\|b\| = \infty$. Fifth, Assumption Sel holds trivially since we are in the "usual case."

Sixth, Assumption CS of McCloskey (2017) can be modified and applied to $\hat{I}^a_n(\cdot)$ and its limit counterpart $I^a_0(\cdot)$ so that: (i)

$$\sup_{(b,\pi)\in\mathcal{P}(\zeta_0,\delta_0)} d_H\big(\hat{I}^a_n(b,\pi), I^a_0(b,\pi)\big) \xrightarrow{p} 0$$

under any $\{\gamma_n\} \in \Gamma(\gamma_0)$, where $d_H(A, B)$ denotes the Hausdorff distance between the two sets A and B; (ii) $I^a_0(\cdot)$ is a continuous and compact-valued correspondence; (iii) $P_{\gamma_n}(\hat{I}^a_n(\hat{b}_n, \hat{\pi}_n) \subset \bar{H}^{(1)}_1(\hat{h}^c_{n,2})) = 1$ for all $n \geq 1$ and $\{\gamma_n\} \in \Gamma(\gamma_0)$ and $P(I^a_0(b+$

$\tau^{\beta}_{0,b}(\pi^*_{0,b}), \pi^*_{0,b}) \subset \bar{H}^{(1)}_1(h^c_2)) = 1$; and (iv) $I^a_0(b + \tau^{\beta}_{0,b}(\pi^*_{0,b}), \pi^*_{0,b})$ need not satisfy a coverage requirement (i.e., $P(h_1 \in I^a_0(b + \tau^{\beta}_{0,b}(\pi^*_{0,b}), \pi^*_{0,b}) \geq 1 - a)$. The proof of Theorem Bonf-Adj in McCloskey (2017) still goes through with this modification of Assumption CS. Condition (i) is satisfied by the consistency of $(\hat{\zeta}_n, \hat{\delta}_n)$ and the uniform consistency of $\hat{\Sigma}_n(\cdot)$ under any $\{\gamma_n\} \in \Gamma(\gamma_0)$. The former holds by Lemma A.1 and Assumption FD while the latter holds by Assumptions V1 and V2 of AC12. For condition (ii), $I^a_0(\cdot)$ is clearly continuous and compact-valued. Note that $\mathcal{P}(\hat{\zeta}_n, \hat{\delta}_n)$ and $\mathcal{P}(\zeta_0, \delta_0)$ are equal to $\bar{H}^{(1)}(\hat{h}^c_{n,2})$ and $\bar{H}^{(1)}(h^c_2)$ in the notation of McCloskey (2017) so that condition (iii) holds by construction.

Seventh, note that rather than using a quantile adjustment function ($a^{(j)}(\cdot)$ in the notation of McCloskey (2017)), we are fixing the quantile at level $1 - \alpha$ and adding a size-correction function $\varsigma(\cdot)$ to it. The proof of Theorem Bonf-Adj of McCloskey (2017) can be easily adjusted to this modification. Rather than requiring the quantile adjustment function to be continuous, the proof requires $\varsigma(\cdot)$ to be continuous. That is, Assumption a(i) of McCloskey (2017) may be replaced by the analogous assumption: $\varsigma(\cdot)$ is continuous. In practice, $\varsigma(\cdot)$ is only evaluated at the point $(\hat{\zeta}_n, \hat{\delta}_n, \hat{\bar{\Sigma}}_n)$, which is consistent with this assumption. Due to the replacement of quantile adjustment by additive size-correction, Assumption a(ii) of McCloskey (2017) should also be replaced by the analogous assumption: $\sup_{(b,\gamma_0)\in\mathbb{R}^{d_\beta}_\infty\times\Gamma:(b,\zeta_0,\pi_0,\delta_0)\in\hat{\mathcal{L}}_n\cap\mathcal{L}(v)} P(\lambda(\pi^*_{0,b};\gamma_0,b) \geq \sup_{\ell\in\mathcal{L}^a_0(b,\gamma_0)\cap\mathcal{L}(v)} c_{1-\alpha}(\ell) + \varsigma(\zeta_0, \delta_0, \bar{\Sigma}(b,\gamma_0))) \leq \alpha$. This assumption holds by the construction of $\varsigma(\hat{\zeta}_n, \hat{\delta}_n, \hat{\bar{\Sigma}}_n)$ and the (uniform) consistency of $(\hat{\zeta}_n, \hat{\delta}_n, \hat{\Sigma}_n(\cdot))$.

Finally, Assumption Inf-Adj of McCloskey (2017) holds vacuously since $\bar{H}^{(1),c} = \emptyset$ and Assumption LB-Adj of that paper is imposed by Assumption DF2. □

Online Appendix B: Assumption verifications for threshold crossing example

Before proceeding to verify the assumptions imposed for the threshold crossing model example, we provide the details for the claim that $\|\tilde{\eta}(\hat{\mu}_n)\|$ diverges for $\hat{\boldsymbol{\mu}}^s_n = (\hat{\boldsymbol{\mu}}_{n,3}, \hat{\boldsymbol{\mu}}_{n,4})$ made in the continuation of Example 2.3 in Section 4 of Han and McCloskey (Forthcoming) (HM18 henceforth).

Proof $\|\tilde{\eta}(\hat{\mu}_n)\|$ diverges in Example 2.3. Note that

$$
\begin{aligned}
\tilde{\eta}_n(\hat{\mu}_n) &= \sqrt{n}\mathcal{S}(\hat{\mu}_n)\left[h_3(\zeta_n, \hat{\pi}_n) - h_3(\zeta_n, \pi_n) + \frac{C_3\big(h_3(\hat{\mu}_n), \hat{\zeta}_{1,n}; \hat{\pi}_n\big)}{C_1\big(h_3(\hat{\mu}_n), \hat{\zeta}_{1,n}; \hat{\pi}_n\big)}(\hat{\pi}_n - \pi_n)\right] \\
&= \sqrt{n}\mathcal{S}(\hat{\mu}_n)\Bigg[\frac{\zeta_{3,n}(\zeta_{1,n}-1)(\zeta_{1,n}-\zeta_{3,n})}{(\zeta_{1,n} - \zeta_{3,n}\hat{\pi}_n + \zeta_{1,n}\zeta_{3,n}\hat{\pi}_n)(\zeta_{1,n} - \zeta_{3,n}\pi_n + \zeta_{1,n}\zeta_{3,n}\pi_n)} \\
&\quad + \frac{C_3\big(h_3(\hat{\mu}_n), \hat{\zeta}_{1,n}; \hat{\pi}_n\big)}{C_1\big(h_3(\hat{\mu}_n), \hat{\zeta}_{1,n}; \hat{\pi}_n\big)}\Bigg](\hat{\pi}_n - \pi_n) \\
&= \sqrt{n}\mathcal{S}(\hat{\mu}_n)\Bigg[\frac{\zeta_{3,n}(\zeta_{1,n}-1)(\zeta_{1,n}-\zeta_{3,n})}{(\zeta_{1,n} - \zeta_{3,n}\hat{\pi}_n + \zeta_{1,n}\zeta_{3,n}\hat{\pi}_n)(\zeta_{1,n} - \zeta_{3,n}\pi_n + \zeta_{1,n}\zeta_{3,n}\pi_n)}
\end{aligned}
$$

$$- \frac{\hat{\zeta}_{3,n}(\hat{\zeta}_{1,n}-1)(\hat{\zeta}_{1,n}-\hat{\zeta}_{3,n})}{(\hat{\zeta}_{1,n}-\hat{\zeta}_{3,n}\hat{\pi}_n+\hat{\zeta}_{1,n}\hat{\zeta}_{3,n}\hat{\pi}_n)^2}\bigg](\hat{\pi}_n-\pi_n)$$
$$= \sqrt{n}\mathcal{S}(\hat{\mu}_n)\left[\frac{\tilde{\eta}_n^N(\hat{\mu}_n)}{\tilde{\eta}_n^D(\hat{\mu}_n)}\right](\hat{\pi}_n-\pi_n),$$

where

$$\begin{aligned}
\tilde{\eta}_n^N(\hat{\mu}_n) &= \zeta_{3,n}(\zeta_{1,n}-1)(\zeta_{1,n}-\zeta_{3,n})(\hat{\zeta}_{1,n}-\hat{\zeta}_{3,n}\hat{\pi}_n+\hat{\zeta}_{1,n}\hat{\zeta}_{3,n}\hat{\pi}_n) \\
&\quad - \hat{\zeta}_{3,n}(\hat{\zeta}_{1,n}-1)(\hat{\zeta}_{1,n}-\hat{\zeta}_{3,n})(\zeta_{1,n}-\zeta_{3,n}\pi_n+\zeta_{1,n}\zeta_{3,n}\pi_n) \\
&= \zeta_{3,n}(\zeta_{1,n}-1)(\zeta_{1,n}-\zeta_{3,n}) \\
&\quad \times \big[(\hat{\zeta}_{1,n}-\hat{\zeta}_{3,n}\hat{\pi}_n+\hat{\zeta}_{1,n}\hat{\zeta}_{3,n}\hat{\pi}_n)-(\zeta_{1,n}-\zeta_{3,n}\pi_n+\zeta_{1,n}\zeta_{3,n}\pi_n)\big] \\
&\quad + \big[\zeta_{3,n}(\zeta_{1,n}-1)(\zeta_{1,n}-\zeta_{3,n})-\hat{\zeta}_{3,n}(\hat{\zeta}_{1,n}-1)(\hat{\zeta}_{1,n}-\hat{\zeta}_{3,n})\big] \\
&\quad \times (\zeta_{1,n}-\zeta_{3,n}\pi_n+\zeta_{1,n}\zeta_{3,n}\pi_n) \\
&= \zeta_{3,n}^2(\zeta_{1,n}-1)^2(\zeta_{1,n}-\zeta_{3,n})(\hat{\pi}_n-\pi_n)+O_p\big(n^{-1/2}\big) = O_p\big(n^{-1/2}\|\beta_n\|^{-1}\big)
\end{aligned}$$

with the final two equalities resulting from Lemma A.1 of HM18 and a mean value expansion of the term $\hat{\zeta}_{3,n}(\hat{\zeta}_{1,n}-1)(\hat{\zeta}_{1,n}-\hat{\zeta}_{3,n})$, and

$$\tilde{\eta}_n^D(\hat{\mu}_n) = \big(\zeta_{1,n}-\zeta_{3,n}\hat{\pi}_n+\zeta_{1,n}\zeta_{3,n}\hat{\pi}_n+O_p\big(n^{-1/2}\big)\big)^2(\zeta_{1,n}-\zeta_{3,n}\pi+\zeta_{1,n}\zeta_{3,n}\pi) = O_p(1)$$

by Lemma A.1 of HM18. Noting that both $\mathcal{S}(\hat{\mu}_n)$ and $\tilde{\eta}_n^D(\hat{\mu}_n)^{-1}$ are also $O_p(1)$ by Lemma A.1 of HM18, we may combine the expressions for $\tilde{\eta}_n(\hat{\mu}_n)$, $\mathcal{S}(\hat{\mu}_n)$, $\tilde{\eta}_n^N(\hat{\mu}_n)$ and $\tilde{\eta}_n^D(\hat{\mu}_n)$ to conclude that $\|\tilde{\eta}_n(\hat{\mu}_n)\| = \|O_p(n^{-1/2}\|\beta_n\|^{-1})\sqrt{n}(\hat{\pi}_n-\pi_n)\| = \|O_p(n^{-1/2}\|\beta_n\|^{-2})\| \to \infty$, according to Lemma A.1 of HM18. □

We now proceed to verify the imposed assumptions for the threshold crossing model example. Hereafter, Andrews and Cheng (2013a) and Han and Vytlacil (2017) are abbreviated as AC13 and HV17. The supplemental material for AC12, AC13, and AC14, Andrews and Cheng (2012, 2013b, 2014), are abbreviated as AC12supp, AC13supp, and AC14supp. The working paper version of AC13 is abbreviated as ACMLwp. And "with respect to" is abbreviated as "w.r.t."

B.1 *Assumptions for threshold crossing models*

The assumptions in the main text of the current paper and the assumptions in AC12 on objects involving the transformed parameter θ are verified under assumptions introduced in this section. The assumptions in AC12 are verified by verifying those in AC13.

Assumption TC1. *$\{W_i = (Y_i, D_i, Z_i) : i \geq 1\}$ is an i.i.d. sequence.*

ASSUMPTION TC2. (i) $Z \perp (\varepsilon, \nu)$;

(ii) F_ε *and* F_ν *are known marginal distributions of* ε *and* ν*, respectively, that are strictly increasing and absolutely continuous with respect to the Lebesgue measure such that* $E[\varepsilon] = E[\nu] = 0$ *and* $\mathrm{Var}(\varepsilon) = \mathrm{Var}(\nu) = 1$;

(iii) $(\varepsilon, \nu)' \sim F_{\varepsilon\nu}(\varepsilon, \nu) = C(F_\varepsilon(\varepsilon), F_\nu(v); \pi)$ *where* $C : (0,1)^2 \to (0,1)$ *is a copula known up to a scalar parameter* $\pi \in \Pi$ *such that* $C(u_1, u_2, ; \pi)$ *is three-times differentiable in* $(u_1, u_2, \pi) \in (0,1)^2 \times \Pi$;

(iv) *The copula* $C(u_1, u_2, ; \pi)$ *satisfies*

$$C(u_1|u_2; \pi) \prec_S C\big(u_1|u_2; \pi'\big) \text{ for any } \pi < \pi', \tag{B.1}$$

where "$\prec_S$*" is a stochastic ordering defined in HV*17 (*Definition* 3.2);

(v) $(1, Z)$ *does not lie in a proper linear subspace of* $\mathbb{R}^2$ *wp*1;

(vi) $\boldsymbol{\Theta}^*$ *is compact and convex.*

Given the form of h in (3.8) of HM18 with $c_4(\zeta)$ set equal to zero, we write $\pi = \boldsymbol{\pi}_3$ in this assumption and below. The conditions in TC2 are sufficient for (global) identification of $\boldsymbol{\theta}$ when $\beta \neq 0$. The argument is similar to that in HV17, except that the condition for the parameter space TC2(vi) is stronger than that in HV17.

For the next assumption, define $\boldsymbol{\Theta}^*_\delta \equiv \{\boldsymbol{\theta} \in \boldsymbol{\Theta}^* : |\beta| < \delta\}$ for some $\delta > 0$.

ASSUMPTION TC3. (i) $\boldsymbol{\Theta} \equiv \boldsymbol{\Theta}_{-\pi} \times \Pi$*, and* $\boldsymbol{\Theta}_{-\pi}$ *and* Π *are compact and simply connected*;

(ii) $\mathrm{int}(\boldsymbol{\Theta}) \supset \boldsymbol{\Theta}^*$;

(iii) *For some* $\delta > 0$, $\boldsymbol{\Theta} \supset \{\beta \in \mathbb{R}^{d_\beta} : |\beta| < \delta\} \times \boldsymbol{\mathcal{Z}}^0 \times \Pi \supset \boldsymbol{\Theta}^*_\delta$ *for some nonempty open set* $\boldsymbol{\mathcal{Z}}^0 \subset \mathbb{R}^{d_{\boldsymbol{\mu}} - d_\pi}$ *and* Π.

(iv) $h^{-1}(\boldsymbol{\mathcal{Z}}^0 \times \Pi) = \mathcal{Z}^0 \times \Pi$ *for some nonempty open set* $\mathcal{Z}^0 \subset \mathbb{R}^{d_{\boldsymbol{\mu}} - d_\pi}$.

As is typical, Assumption TC3(i)–(ii) will be satisfied by a proper choice of the optimization parameter space. For concreteness, we define

$$\begin{aligned}\boldsymbol{\Theta}^* \equiv \big\{\boldsymbol{\theta} = (\beta, \boldsymbol{\zeta}, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3) \in [-0.98, 0.98] \times [0.01, 0.99] \times [0.01, 0.99] \\ \times [0.01, 0.99] \times [-0.99, 0.99] : 0.01 \leq \beta + \boldsymbol{\zeta} \leq 0.99\big\}\end{aligned} \tag{B.2}$$

and

$$\begin{aligned}\boldsymbol{\Theta} \equiv \big\{\boldsymbol{\theta} = (\beta, \boldsymbol{\zeta}, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3) \in [-0.98 - \boldsymbol{\epsilon}, 0.98 + \boldsymbol{\epsilon}] \times [0.01 - \boldsymbol{\epsilon}, 0.99 + \boldsymbol{\epsilon}] \\ \times [0.01 - \boldsymbol{\epsilon}, 0.99 + \boldsymbol{\epsilon}] \times [0.01 - \boldsymbol{\epsilon}, 0.99 + \boldsymbol{\epsilon}] \times [-0.99 - \boldsymbol{\epsilon}, 0.99 + \boldsymbol{\epsilon}] : \\ 0.01 - \boldsymbol{\epsilon} \leq \beta + \boldsymbol{\zeta} \leq 0.99 + \boldsymbol{\epsilon}\big\}\end{aligned} \tag{B.3}$$

for some $\boldsymbol{\epsilon} > 0$ so that TC3(i)–(ii) is clearly satisfied for small enough $\boldsymbol{\epsilon}$. Given the definition (B.2), TC4 below also holds if we define the parameter space $\boldsymbol{\Phi}^*(\boldsymbol{\theta})$ of $\phi \equiv \phi_1$

as

$$\mathbf{\Phi}^*(\boldsymbol{\theta}) = \mathbf{\Phi}^* \equiv [0.01, 0.99]. \tag{B.4}$$

TC3(iii) is satisfied by setting

$$\boldsymbol{\mathcal{Z}}^0 \equiv (0.01 - \delta, 0.99 + \delta)^3$$

for $\delta < \boldsymbol{\epsilon}/2$. For TC3(iv), let $\tilde{h}^{-1}(\boldsymbol{\zeta}, \boldsymbol{\pi}) = (h_1^{-1}(\boldsymbol{\zeta}, \boldsymbol{\pi}), h_2^{-1}(\boldsymbol{\zeta}, \boldsymbol{\pi}), h_3^{-1}(\boldsymbol{\zeta}, \boldsymbol{\pi}))$, the first three elements of (3.12) of HM18. Note that $h_4(\zeta, \pi) = \pi$ (i.e., $\boldsymbol{\pi}_3 = \pi$) and for any given $\pi \in \Pi$, $\tilde{h}^{-1}(\boldsymbol{\mathcal{Z}}^0, \pi)$ does not depend on π. Thus, we may set $\mathcal{Z}^0 = \tilde{h}^{-1}(\boldsymbol{\mathcal{Z}}^0, \pi)$ for any $\pi \in \Pi$, noting that $\mathcal{Z}^0$ must be a nonempty open set by the continuity of the first three elements of $h(\cdot)$. The latter follows from TC2(iii) and (3.8) of HM18 after setting $c_1(\zeta) = \zeta_1$, $c_2(\zeta) = \zeta_2$ and $c_3(\zeta) = \zeta_3$.

Assumption TC4. (i) *$\mathbf{\Gamma}$ is compact and* $\mathbf{\Gamma} = \{\boldsymbol{\gamma} = (\boldsymbol{\theta}, \phi) : \boldsymbol{\theta} \in \boldsymbol{\Theta}^*, \phi \in \mathbf{\Phi}^*(\boldsymbol{\theta})\}$;

(ii) $\forall \delta > 0, \exists \boldsymbol{\gamma} = (\beta, \boldsymbol{\mu}, \phi) \in \mathbf{\Gamma}$ *with* $0 < |\beta| < \delta$;

(iii) $\forall \boldsymbol{\gamma} = (\beta, \boldsymbol{\mu}, \phi) \in \mathbf{\Gamma}$ *with* $0 < |\beta| < \delta$ *for some* $\delta > 0$, $\boldsymbol{\gamma}_a = (a\beta, \boldsymbol{\mu}, \phi) \in \mathbf{\Gamma}\ \forall a \in [0, 1]$.

Assumption TC4(ii) guarantees that the true parameter space includes a region where weak identification occurs and TC4(iii) ensures that $\mathbf{\Gamma}$ is consistent with the existence of $K(\theta; \gamma)$, defined later.

Assumption TC5. (i) *$C(u_1, u_2, ; \pi)$ is bounded away from zero over* $(0, 1)^2 \times \Pi$;

(ii) $0 < \phi_1 \equiv \Pr_\gamma[Z = 1] < 1\ \forall \gamma \in \Gamma$.

Lemma B.1. *TC5 and TC2(iii) imply the following: for* $(y, d, z) \in \{0, 1\}^3$, $\forall \boldsymbol{\gamma} = (\boldsymbol{\theta}, \phi) \in \mathbf{\Gamma}$, *and* $\forall \gamma = (\theta, \phi) \in \Gamma$,

(i) *the first-, second-, and third-order derivatives of $\boldsymbol{p}_{yd,z}(\boldsymbol{\theta})$ are bounded over $\boldsymbol{\Theta}$*;

(ii) *$\boldsymbol{p}_{yd,z}(\boldsymbol{\theta})$ is bounded away from zero over $\boldsymbol{\Theta}$ and* $0 < \phi_1 < 1$;

(iii) *$\bar{h}(\theta)$ is three-times differentiable on Θ*;

(iv) *$p_{yd,z}(\theta) \equiv \boldsymbol{p}_{yd,z}(\bar{h}(\theta))$ is three-times differentiable on Θ and the first-, second-, and third-order derivatives of $p_{yd,z}(\theta)$ are bounded over Θ*;

(v) *$p_{yd,z}(\theta)$ is bounded away from zero over Θ.*

Proof of Lemma B.1. (i) holds by TC2(iii), the fact that the domain $\boldsymbol{\Theta}$ is compact by TC3(i), and the definitions of $\boldsymbol{p}_{yd,z}(\boldsymbol{\theta})$. (ii) immediately holds by TC5. For (iii), given (3.9) of HM18, TC2(iii), and TC3(i) imply that $h(\mu)$ is three-times differentiable in μ, and hence $\bar{h}(\theta) = (\beta, h(\mu))$ is three-times differentiable in θ. Next, (iv) holds by (i), (iii), and the chain rule, and (v) trivially holds by (ii). □

B.2 *Verification of assumptions in the main text*

Assumptions CF, ID, Jac, and Reg3 are verified in the main text. Assumption Reg1 is satisfied with $\bar{\mathbf{g}}_n(\boldsymbol{\theta}) = \hat{\xi}_n - \mathfrak{g}(\boldsymbol{\theta})$, where each element $\boldsymbol{p}_{yd,z}(\boldsymbol{\theta})$ of the vector $\mathfrak{g}(\boldsymbol{\theta})$ is continuously differentiable by TC2(iii). For Assumption H, H(i) holds since its sufficient conditions that Θ is bounded and h is continuous hold by S2(v), verified below, and by Proposition 3.1 of HM18, respectively. H(ii) is also trivially satisfied by TC3(i). For Reg2, $\text{rank}(h^s_\pi(\mu)) = 1$ if $h^s(\pi)$ contains $h_2(\pi)$, $h_3(\pi)$ or $h_4(\pi)$ and $\text{rank}(h^s_\pi(\mu)) = 0$ otherwise, as can be seen from the form of h in (3.8) of HM18 upon setting $c_1(\zeta) = \zeta_1$, $c_2(\zeta) = \zeta_2$, $c_3(\zeta) = \zeta_3$, and $c_4(\zeta) = 0$.

B.3 *Verification of assumptions in Andrews and Cheng (2013)*

In this section, given our transformed parameter θ and associated transformed objects, we verify the regularity conditions for the asymptotic theory of the ML estimator $\hat{\theta}_n$ in AC13. Specifically, we show that Assumptions TC1–TC5 are sufficient for Assumptions S1–S4, B1, B2, C6, C7, V1, and V2 of AC13. Then, under Assumptions B1 and B2, Assumptions S1–S3 of AC13 imply Assumptions A, B3, C1–C4, C8, and D1–D3 of AC12; see Lemma 9.1 in ACMLwp. Maintaining the same labels of AC13, below we rewrite the assumptions of AC13 before verifying them. Note that in our stylized threshold crossing model, β is scalar. Therefore, we do not consider Assumptions S3* and V1* of AC13 which apply to the vector β case.

Assumption S1. *$\forall\gamma_0 \in \Gamma$, $\{W_i : i \geq 1\}$ is an i.i.d. sequence and the constant q (that appears in Assumption S3 below) equals $2 + \delta$ for some $\delta > 0$.*

Assumption S2. (i) *For some function $\rho(w, \theta) \in \mathbb{R}$, $Q_n(\theta) = n^{-1}\sum_{i=1}^n \rho(W_i, \theta)$, where $\rho(w, \theta)$ is twice continuously differentiable in θ on an open set containing $\Theta^* \ \forall w \in \mathcal{W}$.*

(ii) *$\rho(w, \theta)$ does not depend on π when $\beta = 0 \ \forall w \in \mathcal{W}$.*

(iii) *$\forall\gamma_0 \in \Gamma$ with $\beta_0 = 0$, $E_{\gamma_0}\rho(W_i, \psi, \pi)$ is uniquely minimized by $\psi_0 \ \forall \pi \in \Pi$.*

(iv) *$\forall\gamma_0 \in \Gamma$ with $\beta_0 \neq 0$, $E_{\gamma_0}\rho(W_i, \theta)$ is uniquely minimized by θ_0.*

(v) *$\Psi(\pi)$ is compact $\forall \pi \in \Pi$, and Π and Θ are compact.*

(vi) *$\forall\epsilon > 0$, $\exists\delta > 0$ such that $d_H(\Psi(\pi_1), \Psi(\pi_2)) < \epsilon \ \forall \pi_1, \pi_2 \in \Pi$ with $|\pi_1 - \pi_2| < \delta$, where $d_H(\cdot, \cdot)$ is the Hausdorff metric.*

Verification of S2(i). By TC2(iii), $\boldsymbol{p}_{yd,z}(\boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$. Then, since $p_{yd,z}(\theta) \equiv \boldsymbol{p}_{yd,z}(\bar{h}(\theta))$ is twice continuously differentiable by Lemma B.1, so is $\rho(w, \theta) = -\sum_{y,d,z=0,1} \mathbf{1}_{ydz}(w) \log p_{yd,z}(\theta)$. □

Verification of S2(ii). It is easy to see from (2.5)–(2.6) of HM18 that, when $\beta = 0$, $\boldsymbol{p}_{yd,0}(\boldsymbol{\theta}) = \boldsymbol{p}_{yd,1}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$ and (y, d), which implies that $\boldsymbol{p}_{yd,0}(\bar{h}(\theta)) = \boldsymbol{p}_{yd,1}(\bar{h}(\theta))$ for all θ. Therefore,

$$\begin{aligned}
p_{11,1}(\theta) &= p_{11,0}(\theta) = \zeta_3, \\
p_{10,1}(\theta) &= p_{10,0}(\theta) = \zeta_2, \\
p_{01,1}(\theta) &= p_{01,0}(\theta) = \zeta_1 - \zeta_3,
\end{aligned} \tag{B.5}$$

where the second equality in each equation is from (7.1)–(7.2) of HM18. Therefore, $p_{yd,z}(\theta)$ does not depend on π when $\beta = 0$, and hence $\rho(w, \theta) = -\sum_{y,d,z=0,1} \mathbf{1}_{ydz}(w) \times \log p_{yd,z}(\theta)$ does not depend on π. □

Verification of S2(iii). When $\beta_0 = 0$, for $\psi \neq \psi_0$ and for a given π,

$$\begin{aligned}
E_{\gamma_0}\rho(W_i, \psi, \pi) - E_{\gamma_0}\rho(W_i, \psi_0, \pi) &= -\sum_{y,d,z=0,1} p_{yd,z}(\psi_0, \pi_0)\phi_{z,0} \log \frac{p_{yd,z}(\psi, \pi)}{p_{yd,z}(\psi_0, \pi)} \\
&\geq -\log \sum_{y,d,z=0,1} p_{yd,z}(\psi_0, \pi_0)\phi_{z,0} \frac{p_{yd,z}(\psi, \pi)}{p_{yd,z}(\psi_0, \pi)} \\
&= -\log \sum_{y,d,z=0,1} p_{yd,z}(\psi, \pi)\phi_{z,0} \\
&= 0,
\end{aligned}$$

where the last equality holds since $\sum_{y,d} p_{yd,1}(\theta) = \sum_{y,d} p_{yd,0}(\theta) = 1$ and $\phi_{0,0} = 1 - \phi_{1,0}$, and the second-to-last equality holds since

$$p_{yd,z}(\psi_0, \pi_0) = p_{yd,z}(\psi_0, \pi) \equiv p_{yd}^0 \tag{B.6}$$

when $\beta_0 = 0$, as in (B.5). Notationally, $p_{11} = \zeta_3$, $p_{10} = \zeta_2$, and $p_{01} = \zeta_1 - \zeta_3$. The Jensen's inequality is strict if there exist $(y, d, z) \in \{0, 1\}^3$ such that

$$\frac{p_{yd,z}(\psi, \pi)}{p_{yd,z}(\psi_0, \pi)} \neq 1.$$

Under TC2, this condition can be readily shown to hold by a slight modification of the identification proof of Theorem 4.1 in HV17, which is omitted here for brevity. □

Verification of S2(iv). For $\theta \neq \theta_0$,

$$\begin{aligned}
Q_0(\theta) - Q_0(\theta_0) &= -\sum_{y,d,z=0,1} p_{yd,z}(\theta_0)\phi_{z,0} \log \frac{p_{yd,z}(\theta)}{p_{yd,z}(\theta_0)} \\
&> -\log \sum_{y,d,z=0,1} p_{yd,z}(\theta)\phi_{z,0} \\
&= 0,
\end{aligned}$$

where the Jensen's inequality is strict because there exist $(y,d,z)\in\{0,1\}^3$ such that

$$\frac{p_{yd,z}(\theta)}{p_{yd,z}(\theta_0)}\neq 1$$

by Theorem 4.1 in HV17 under TC2. □

VERIFICATION OF S2(v). By TC3(i), Π is compact and the parameter space is the same before and after the transformation. Also, $\Theta=\bar{h}^{-1}(\boldsymbol{\Theta})$ is compact since $\boldsymbol{\Theta}$ is compact and Assumption H(i) holds. For compactness of $\Psi(\pi)$, first note that, for a given $\pi\in\Pi$, $\bar{h}_{-\pi}(\cdot,\pi)$, which is $\bar{h}(\cdot,\pi)$ except the last element, is a homeomorphism. This is because $\boldsymbol{\Theta}_{-\pi}$ is simply connected, $\bar{h}_{-\pi}(\cdot,\pi)$ is continuous, and $\Psi(\pi)$ is bounded since Θ is bounded. Then

$$\boldsymbol{\Theta}_{-\pi}=\boldsymbol{\Theta}_{-\pi}(\pi)\equiv\bar{h}_{-\pi}\big(\Psi(\pi),\pi\big)$$

where the first equality is because the dependence parameter π does not restrict the space of the remaining elements of $\boldsymbol{\theta}$ (or by TC3(i)), and thus $\Psi(\pi)=\bar{h}_{-\pi}^{-1}(\boldsymbol{\Theta}_{-\pi},\pi)$. Therefore, $\Psi(\pi)$ is compact since $\boldsymbol{\Theta}_{-\pi}$ is compact and $\bar{h}_{-\pi}(\cdot,\pi)$ is proper. □

VERIFICATION OF S2(vi). The space of $\psi=(\beta,\zeta)$ is continuous in π since $\Psi(\pi)=\bar{h}_{-\pi}^{-1}(\boldsymbol{\Theta}_{-\pi},\pi)$, where $\bar{h}_{-\pi}^{-1}(\boldsymbol{\theta}_{-\pi},\pi)$ is continuous in π by (3.12) of HM18 and TC2(iii). □

Let $\rho_\theta(w,\theta)$ and $\rho_{\theta\theta}(w,\theta)$ denote the first- and second-order partial derivatives of $\rho(w,\theta)$ w.r.t. θ, respectively. Also, let $\rho_\psi(w,\theta)$ and $\rho_{\psi\psi}(w,\theta)$ denote the first- and second-order partial derivatives of $\rho(w,\theta)$ w.r.t. ψ, respectively. Recall

$$B(\beta)\equiv\begin{bmatrix} I_{d_\psi} & 0_{d_\psi\times 1}\\ 0_{1\times d_\psi} & \beta\end{bmatrix}\in\mathbb{R}^{d_\theta\times d_\theta}.$$

For $\beta\neq 0$, let

$$\begin{aligned} B^{-1}(\beta)\rho_\theta(w,\theta)&\equiv\rho_\theta^\dagger(w,\theta),\\ B^{-1}(\beta)\rho_{\theta\theta}(w,\theta)B^{-1}(\beta)&\equiv\rho_{\theta\theta}^\dagger(w,\theta)+r(w,\theta),\end{aligned}\tag{B.7}$$

where $\rho_{\theta\theta}^\dagger(w,\theta)$ is symmetric and $\rho_\theta^\dagger(w,\theta)$, $\rho_{\theta\theta}^\dagger(w,\theta)$, and $r(w,\theta)$ satisfy Assumption S3 below;[1] see below for actual expressions of these terms. Next, define

$$V^\dagger(\theta_1,\theta_2;\gamma_0)\equiv\mathrm{Cov}_{\gamma_0}\big(\rho_\theta^\dagger(W_i,\theta_1),\rho_\theta^\dagger(W_i,\theta_2)\big).$$

Let $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximum and minimum eigenvalues, respectively, of a square matrix A.

[1]The remainder term $r(w,\theta)$ and related conditions in S3 are slightly more general than conditions on $\beta^{-1}\varepsilon(w,\theta)$ and related conditions in AC13.

In this example of a threshold crossing model, define $D_\theta p^\dagger_{yd,z}(\theta) \equiv B^{-1}(\beta) \times D_\theta p_{yd,z}(\theta)$ so that

$$\rho_\theta(w,\theta) = -\sum_{y,d,z=0,1} \mathbf{1}_{ydz}(w) \frac{1}{p_{yd,z}(\theta)} D_\theta p_{yd,z}(\theta),$$

$$\rho_{\theta\theta}(w,\theta) = -\sum_{y,d,z=0,1} \mathbf{1}_{ydz}(w) \Bigg[-\frac{1}{p_{yd,z}(\theta)^2} D_\theta p_{yd,z}(\theta) D_\theta p_{yd,z}(\theta)' + \frac{1}{p_{yd,z}(\theta)} D_{\theta\theta} p_{yd,z}(\theta) \Bigg],$$

$$\rho^\dagger_\theta(w,\theta) = -\sum_{y,d,z=0,1} \mathbf{1}_{ydz}(w) \frac{1}{p_{yd,z}(\theta)} D_\theta p^\dagger_{yd,z}(\theta),$$

$$\rho^\dagger_{\theta\theta}(w,\theta) = \rho^\dagger_\theta(w,\theta)\rho^\dagger_\theta(w,\theta)' = \sum_{y,d,z=0,1} \mathbf{1}_{ydz}(w) \frac{1}{p_{yd,z}(\theta)^2} D_\theta p^\dagger_{yd,z}(\theta) D_\theta p^\dagger_{yd,z}(\theta)',$$

$$r(w,\theta) = -\sum_{y,d,z=0,1} \mathbf{1}_{ydz}(w) \frac{1}{p_{yd,z}(\theta)} B^{-1}(\beta) D_{\theta\theta} p_{yd,z}(\theta) B^{-1}(\beta).$$

Suppressing the argument (ζ_1, ζ_3, π) in h_3 and its derivatives, and suppressing the argument (ζ_1, ζ_2, π) in h_2 and its derivatives, note that from (7.1)–(7.2) of HM18,

$$D_\theta p_{11,0}(\theta) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \qquad D_\theta p_{10,0}(\theta) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

$$D_\theta p_{01,0}(\theta) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \qquad D_\theta p_{00,0}(\theta) = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 0 \end{bmatrix},$$

$$D_\theta p_{11,1}(\theta) = \begin{bmatrix} C_2(h_3, \zeta_1+\beta; \pi) \\ C_2(h_3, \zeta_1+\beta; \pi) + C_1(h_3, \zeta_1+\beta; \pi) h_{3,\zeta_1} \\ 0 \\ C_1(h_3, \zeta_1+\beta; \pi) h_{3,\zeta_3} \\ C_\pi(h_3, \zeta_1+\beta; \pi) + C_1(h_3, \zeta_1+\beta; \pi) h_{3,\pi} \end{bmatrix}$$
$$= \begin{bmatrix} C_2(h_3, \zeta_1+\beta; \pi) \\ C_2(h_3, \zeta_1+\beta; \pi) + C_1(h_3, \zeta_1+\beta; \pi) h_{3,\zeta_1} \\ 0 \\ C_1(h_3, \zeta_1+\beta; \pi) h_{3,\zeta_3} \\ \beta\{C_{\pi 2}(h_3, \zeta_1+\beta^\dagger; \pi) + C_{12}(h_3, \zeta_1+\beta^\dagger; \pi) h_{3,\pi}\} \end{bmatrix}, \tag{B.8}$$

where $0 \leq |\beta^{\dagger}| \leq \beta$. The last equality is derived using a mean value expansion and the fact that $C_{\pi}(h_3, \zeta_1; \pi) + C_1(h_3, \zeta_1; \pi) h_{3,\pi} = 0$, obtained by differentiating $C(h_3, \zeta_1; \pi) = \zeta_3$ w.r.t. π. Furthermore,

$$
\begin{aligned}
D_{\theta} p_{10,1}(\theta) &= \begin{bmatrix} -C_2(h_2, \zeta_1 + \beta; \pi) \\ h_{2,\zeta_1} - C_2(h_2, \zeta_1 + \beta; \pi) - C_1(h_2, \zeta_1 + \beta; \pi) h_{2,\zeta_1} \\ h_{2,\zeta_2} - C_1(h_2, \zeta_1 + \beta; \pi) h_{2,\zeta_2} \\ 0 \\ h_{2,\pi} - C_{\pi}(h_2, \zeta_1 + \beta; \pi) - C_1(h_2, \zeta_1 + \beta; \pi) h_{2,\pi} \end{bmatrix} \\
&= \begin{bmatrix} -C_2(h_2, \zeta_1 + \beta; \pi) \\ h_{2,\zeta_1} - C_2(h_2, \zeta_1 + \beta; \pi) - C_1(h_2, \zeta_1 + \beta; \pi) h_{2,\zeta_1} \\ h_{2,\zeta_2} - C_1(h_2, \zeta_1 + \beta; \pi) h_{2,\zeta_2} \\ 0 \\ -\beta \left\{ C_{\pi 2}\big(h_2, \zeta_1 + \beta^{\dagger\dagger}; \pi\big) + C_{12}\big(h_2, \zeta_1 + \beta^{\dagger\dagger}; \pi\big) h_{2,\pi} \right\} \end{bmatrix},
\end{aligned} \tag{B.9}
$$

where $0 \leq |\beta^{\dagger\dagger}| \leq \beta$ and the last equality is derived using a mean value expansion and the fact that $h_{2,\pi} - C_{\pi}(h_2, \zeta_1; \pi) - C_1(h_2, \zeta_1; \pi) h_{2,\pi} = 0$. Finally,

$$
D_{\theta} p_{01,1}(\theta) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} - D_{\theta} p_{11,1}(\theta), \qquad D_{\theta} p_{00,1}(\theta) = \begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} - D_{\theta} p_{10,1}(\theta).
$$

Also, note that for all (y, d),

$$
D_{\theta\theta} p_{yd,0}(\theta) = 0 \tag{B.10}
$$

and

$$
D_{\theta\theta} p_{01,1}(\theta) = -D_{\theta\theta} p_{11,1}(\theta), \qquad D_{\theta\theta} p_{00,1}(\theta) = -D_{\theta\theta} p_{10,1}(\theta). \tag{B.11}
$$

Now, for $z = 0$,

$$
D_{\theta} p^{\dagger}_{yd,z}(\theta) = D_{\theta} p_{yd,z}(\theta) \tag{B.12}
$$

and, for $z = 1$,

$$
D_{\theta} p^{\dagger}_{11,1}(\theta) = \begin{bmatrix} C_2(h_3, \zeta_1 + \beta; \pi) \\ C_2(h_3, \zeta_1 + \beta; \pi) + C_1(h_3, \zeta_1 + \beta; \pi) h_{3,\zeta_1} \\ 0 \\ C_1(h_3, \zeta_1 + \beta; \pi) h_{3,\zeta_3} \\ C_{\pi 2}\big(h_3, \zeta_1 + \beta^{\dagger}; \pi\big) + C_{12}\big(h_3, \zeta_1 + \beta^{\dagger}; \pi\big) h_{3,\pi} \end{bmatrix}, \tag{B.13}
$$

$$D_\theta p^\dagger_{10,1}(\theta) = \begin{bmatrix} -C_2(h_2, \zeta_1+\beta; \pi) \\ h_{2,\zeta_1} - C_2(h_2, \zeta_1+\beta; \pi) - C_1(h_2, \zeta_1+\beta; \pi) h_{2,\zeta_1} \\ h_{2,\zeta_2} - C_1(h_2, \zeta_1+\beta; \pi) h_{2,\zeta_2} \\ 0 \\ -C_{\pi 2}\big(h_2, \zeta_1+\beta^{\dagger\dagger}; \pi\big) - C_{12}\big(h_2, \zeta_1+\beta^{\dagger\dagger}; \pi\big) h_{2,\pi} \end{bmatrix}, \tag{B.14}$$

and expressions for the remaining two derivatives can be derived analogously.

Note that

$$\rho_\psi(w,\theta) = -\sum_{y,d,z=0,1} \mathbf{1}_{ydz}(w) \frac{1}{p_{yd,z}(\theta)} D_\psi p_{yd,z}(\theta),$$

$$\rho_{\psi\psi}(w,\theta) = -\sum_{y,d,z=0,1} \mathbf{1}_{ydz}(w) \Big[-\frac{1}{p_{yd,z}(\theta)^2} D_\psi p_{yd,z}(\theta) D_\psi p_{yd,z}(\theta)' + \frac{1}{p_{yd,z}(\theta)} D_{\psi\psi} p_{yd,z}(\theta) \Big],$$

where, with $\psi = (\beta, \zeta) = (\beta, \zeta_1, \zeta_2, \zeta_3)$,

$$D_\psi p_{11,0}(\theta) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \qquad D_\psi p_{10,0}(\theta) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

$$D_\psi p_{01,0}(\theta) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \qquad D_\psi p_{00,0}(\theta) = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \end{bmatrix},$$

$$D_\psi p_{11,1}(\theta) = \begin{bmatrix} C_2(h_3, \zeta_1+\beta; \pi) \\ C_2(h_3, \zeta_1+\beta; \pi) + C_1(h_3, \zeta_1+\beta; \pi) h_{3,\zeta_1} \\ 0 \\ C_1(h_3, \zeta_1+\beta; \pi) h_{3,\zeta_3} \end{bmatrix},$$

$$D_\psi p_{10,1}(\theta) = \begin{bmatrix} -C_2(h_2, \zeta_1+\beta; \pi) \\ h_{2,\zeta_1} - C_2(h_2, \zeta_1+\beta; \pi) - C_1(h_2, \zeta_1+\beta; \pi) h_{2,\zeta_1} \\ h_{2,\zeta_2} - C_1(h_2, \zeta_1+\beta; \pi) h_{2,\zeta_2} \\ 0 \end{bmatrix},$$

and

$$D_\psi p_{01,1}(\theta) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - D_\psi p_{11,1}(\theta), \qquad D_\psi p_{00,1}(\theta) = \begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \end{bmatrix} - D_\psi p_{10,1}(\theta).$$

Also, for all (y, d) and θ,

$$D_{\psi\psi} p_{yd,0}(\theta) = 0 \tag{B.15}$$

and

$$D_{\psi\psi} p_{01,1}(\theta) = -D_{\psi\psi} p_{11,1}(\theta), \qquad D_{\psi\psi} p_{00,1}(\theta) = -D_{\psi\psi} p_{10,1}(\theta). \tag{B.16}$$

Assumption S3. (i) (a) $E_{\gamma_0} r(W_i, \theta_0) = 0$; *and* (b) $\|E_{\gamma_0} r(W_i, \psi_0, \pi)\| \le C|\pi - \pi_0|\ \forall \gamma_0 \in \Gamma$ *with* $0 < |\beta_0| < \delta$ *for some* $\delta > 0$.

(ii) (a) *For all* $\delta > 0$ *and some function* $M_1(w) : \mathcal{W} \to \mathbb{R}_+$, $\|\rho_{\psi\psi}(w, \theta_1) - \rho_{\psi\psi}(w, \theta_2)\| + \|\rho^\dagger_{\theta\theta}(w, \theta_1) - \rho^\dagger_{\theta\theta}(w, \theta_2)\| \le M_1(w)\delta$, $\forall \theta_1, \theta_2 \in \Theta$ *with* $\|\theta_1 - \theta_2\| \le \delta$, $\forall w \in \mathcal{W}$; *and* (b) *for all* $\delta > 0$ *and some function* $M_2(w) : \mathcal{W} \to \mathbb{R}_+$, $\|\rho^\dagger_{\theta}(w, \theta_1) - \rho^\dagger_{\theta}(w, \theta_2)\| + \|r(w, \theta_1) - r(w, \theta_2)\| \le M_2(w)\delta$, $\forall \theta_1, \theta_2 \in \Theta$ *with* $\|\theta_1 - \theta_2\| \le \delta$, $\forall w \in \mathcal{W}$.

(iii) $E_{\gamma_0} \sup_{\theta \in \Theta} \{|\rho(W_i, \theta)|^{1+\delta} + \|\rho_{\psi\psi}(W_i, \theta)\|^{1+\delta} + \|\rho^\dagger_{\theta\theta}(W_i, \theta)\|^{1+\delta} + M_1(W_i) + \|\rho^\dagger_{\theta}(W_i, \theta)\|^q + \|r(W_i, \theta)\|^q + M_2(W_i)^q \le C$ *for some* $\delta > 0\ \forall \gamma_0 \in \Gamma$, *where* q *is as in Assumption* S1.

(iv) (a) $\lambda_{\min}(E_{\gamma_0} \rho_{\psi\psi}(W_i, \psi_0, \pi)) > 0\ \forall \pi \in \Pi$ *when* $\beta_0 = 0$; *and* (b) $E_{\gamma_0} \rho^\dagger_{\theta\theta}(W_i, \theta_0)$ *is positive definite* $\forall \gamma_0 \in \Gamma$.

(v) $V^\dagger(\theta_0, \theta_0; \gamma_0)$ *is positive definite* $\forall \gamma_0 \in \Gamma$.

Verification of S3(i)(a). Note that

$$E_{\gamma_0} r(W_i, \theta_0) = -\sum_{y,d,z=0,1} \phi_{z,0} B^{-1}(\beta_0) D_{\theta\theta} p_{yd,z}(\theta_0) B^{-1}(\beta_0) = 0$$

by (B.10) and (B.11) since $\beta_0 \ne 0$. □

Verification of S3(i)(b). Using (B.10) and (B.11),

$$\begin{aligned}
&E_{\gamma_0} r(W_i, \psi_0, \pi) \\
&\quad = \sum_{y,d=0,1} p_{yd,1}(\theta_0) \phi_{1,0} B^{-1}(\beta_0) \frac{D_{\theta\theta} p_{yd,1}(\psi_0, \pi)}{p_{yd,1}(\psi_0, \pi)} B^{-1}(\beta_0) \\
&\quad = \phi_{1,0} B^{-1}(\beta_0) \bigg[\frac{p_{11,1}(\theta_0)}{p_{11,1}(\psi_0, \pi)} D_{\theta\theta} p_{11,1}(\psi_0, \pi) + \frac{p_{01,1}(\theta_0)}{p_{01,1}(\psi_0, \pi)} D_{\theta\theta} p_{01,1}(\psi_0, \pi) \\
&\qquad + \frac{p_{10,1}(\theta_0)}{p_{10,1}(\psi_0, \pi)} D_{\theta\theta} p_{10,1}(\psi_0, \pi) + \frac{p_{00,1}(\theta_0)}{p_{00,1}(\psi_0, \pi)} D_{\theta\theta} p_{00,1}(\psi_0, \pi) \bigg] B^{-1}(\beta_0) \\
&\quad = \phi_{1,0} B^{-1}(\beta_0) \bigg[\left(\frac{p_{11,1}(\theta_0)}{p_{11,1}(\psi_0, \pi)} - \frac{p_{01,1}(\theta_0)}{p_{01,1}(\psi_0, \pi)} \right) D_{\theta\theta} p_{11,1}(\psi_0, \pi) \\
&\qquad + \left(\frac{p_{10,1}(\theta_0)}{p_{10,1}(\psi_0, \pi)} - \frac{p_{00,1}(\theta_0)}{p_{00,1}(\psi_0, \pi)} \right) D_{\theta\theta} p_{10,1}(\psi_0, \pi) \bigg] B^{-1}(\beta_0)
\end{aligned}$$

$$
\begin{aligned}
&= \phi_{1,0} B^{-1}(\beta_0)\bigg[\bigg(\frac{(\zeta_{10}+\beta_0)\big(p_{11,1}(\theta_0)-p_{11,1}(\psi_0,\pi)\big)}{p_{11,1}(\psi_0,\pi)\big(\zeta_{10}+\beta_0-p_{11,1}(\psi_0,\pi)\big)}\bigg) D_{\theta\theta} p_{11,1}(\psi_0,\pi) \\
&\quad + \bigg(\frac{(1-\zeta_{10}-\beta_0)\big(p_{10,1}(\theta_0)-p_{10,1}(\psi_0,\pi)\big)}{p_{10,1}(\psi_0,\pi)\big(1-\zeta_{10}-\beta_0-p_{10,1}(\psi_0,\pi)\big)}\bigg) D_{\theta\theta} p_{10,1}(\psi_0,\pi)\bigg] B^{-1}(\beta_0), \quad \text{(B.17)}
\end{aligned}
$$

where the last equality uses $p_{01,1}(\theta) = \zeta_1 + \beta - p_{11,1}(\theta)$ and $p_{00,1}(\theta) = 1 - \zeta_1 - \beta - p_{10,1}(\theta)$. Apply the mean value theorem to $p_{11,1}(\theta_0) - p_{11,1}(\psi_0, \pi)$ w.r.t. π:

$$
\begin{aligned}
p_{11,1}(\psi_0,\pi_0) - p_{11,1}(\psi_0,\pi) &= \frac{\partial p_{11,1}(\psi_0,\pi^\dagger)}{\partial \pi}(\pi_0 - \pi) \\
&= \frac{\partial^2 p_{11,1}(\beta^\dagger, \zeta_0, \pi^\dagger)}{\partial \pi \partial \beta}(\pi_0 - \pi)\beta_0, \quad \text{(B.18)}
\end{aligned}
$$

where $\pi^\dagger$ is between π_0 and π and $0 \le |\beta^\dagger| \le |\beta_0|$. The second equality holds by another mean value expansion of $\frac{\partial p_{11,1}(\psi_0,\pi^\dagger)}{\partial \pi}$ w.r.t. β_0 around $\beta_0 = 0$ and the fact that $\frac{\partial p_{11,1}(\beta,\zeta_0,\pi^\dagger)}{\partial \pi}|_{\beta=0} = 0$ since

$$
C_\pi\big(h_3(\pi), \zeta_1; \pi\big) + C_1\big(h_3(\pi), \zeta_1; \pi\big) h_{3,\pi}(\pi) = 0
$$

for all (ζ_1, ζ_3, π). Similarly, using mean value expansions,

$$
p_{10,1}(\psi_0,\pi_0) - p_{10,1}(\psi_0,\pi) = \frac{\partial^2 p_{10,1}(\beta^{\dagger\dagger}, \zeta_0, \pi^{\dagger\dagger})}{\partial \pi \partial \beta}(\pi_0 - \pi)\beta_0 \quad \text{(B.19)}
$$

for some $\pi^{\dagger\dagger}$ between π_0 and π and $0 \le |\beta^{\dagger\dagger}| \le |\beta_0|$. Therefore, combining (B.17)–(B.19),

$$
\begin{aligned}
\big\|E_{\gamma_0} r(W_i, \psi_0, \pi)\big\| &\le |c_1| \big\|B^{-1}(\beta_0)\beta_0 D_{\theta\theta} p_{11,1}(\psi_0,\pi) B^{-1}(\beta_0)\big\| |\pi_0 - \pi| \\
&\quad + |c_2| \big\|B^{-1}(\beta_0)\beta_0 D_{\theta\theta} p_{10,1}(\psi_0,\pi) B^{-1}(\beta_0)\big\| |\pi_0 - \pi|,
\end{aligned}
$$

where c_1 and c_2 are collections of all other terms, whose norms are bounded by (7.1)–(7.2) of HM18 and Lemma B.1. Also $\|B^{-1}(\beta_0)\beta_0\|$ is bounded for $0 < |\beta_0| < \delta$. Note that $\|D_{\theta\theta} p_{11,1}(\psi_0,\pi) B^{-1}(\beta_0)\|$ and $\|D_{\theta\theta} p_{10,1}(\psi_0,\pi) B^{-1}(\beta_0)\|$ can be shown to be bounded for $0 < |\beta_0| < \delta$ by differentiating (B.13) and (B.14) w.r.t. θ, respectively, and applying Lemma B.1. □

Verification of S3(ii)(a). Generically, for $A = aa'$ where $a = (a_1, \ldots, a_p) \in \mathbb{R}^{d_a}$ and $a_1, \ldots, a_p$ are vectors,

$$
\|A\| \le \sum_{j=1}^{p} \|a_j\|^2,
$$

and for $A^* = a^* a^{*\prime}$

$$
\begin{aligned}
\|A - A^*\| &\le \|a(a - a^*)'\| + \|(a - a^*)a^{*\prime}\| \le \big(\|a\| + \|a^*\|\big)\|a - a^*\| \\
&\le \sum_{j=1}^{p}\big(\|a_j\| + \|a_j^*\|\big) \sum_{j=1}^{p} \|a_j - a_j^*\|.
\end{aligned}
$$

Applying this result to the last inequality below,

$$
\begin{aligned}
&\left\|\rho_{\psi\psi}(w,\theta_1)-\rho_{\psi\psi}(w,\theta_2)\right\| \\
&\quad\le \sum_{y,d,z=0,1}\left\|\frac{D_\psi p_{yd,z}(\theta_1)D_\psi p_{yd,z}(\theta_1)'}{p_{yd,z}(\theta_1)^2}-\frac{D_\psi p_{yd,z}(\theta_2)D_\psi p_{yd,z}(\theta_2)'}{p_{yd,z}(\theta_2)^2}\right\| \\
&\qquad+\sum_{y,d,z=0,1}\left\|\frac{D_{\psi\psi} p_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}-\frac{D_{\psi\psi} p_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right\| \\
&\quad\le \sum_{y,d,z=0,1}\left(\left\|\frac{D_\psi p_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}\right\|+\left\|\frac{D_\psi p_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right\|\right)\left\|\frac{D_\psi p_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}-\frac{D_\psi p_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right\| \\
&\qquad+\sum_{y,d,z=0,1}\left\|\frac{D_{\psi\psi} p_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}-\frac{D_{\psi\psi} p_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right\| \\
&\quad\le \sum_{y,d,z=0,1}\sum_{j=1}^{d_\psi}\left(\left|\frac{D_{\psi_j} p_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}\right|+\left|\frac{D_{\psi_j} p_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right|\right)\sum_{j=1}^{d_\psi}\left|\frac{D_{\psi_j} p_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}-\frac{D_{\psi_j} p_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right| \\
&\qquad+\sum_{y,d,z=0,1}\sum_{j,k=1}^{d_\psi}\left|\frac{D_{\psi_j\psi_k} p_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}-\frac{D_{\psi_j\psi_k} p_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right|,
\end{aligned}
$$

where $|\mathbf{1}_{ydz}(w)|\le 1$ is used in the first inequality. Applying the mean value theorem to the differential terms,

$$
\begin{aligned}
\left|\frac{D_{\psi_j} p_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}-\frac{D_{\psi_j} p_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right| &\le \left\|D_\theta\left\{\frac{D_{\psi_j} p_{yd,z}(\theta^\dagger)}{p_{yd,z}(\theta^\dagger)}\right\}\right\|\|\theta_1-\theta_2\|, \\
\left|\frac{D_{\psi_j\psi_k} p_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}-\frac{D_{\psi_j\psi_k} p_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right| &\le \left\|D_\theta\left\{\frac{D_{\psi_j\psi_k} p_{yd,z}(\theta^{\dagger\dagger})}{p_{yd,z}(\theta^{\dagger\dagger})}\right\}\right\|\|\theta_1-\theta_2\|,
\end{aligned}
$$

where $\theta^\dagger$ and $\theta^{\dagger\dagger}$ lie between θ_1 and θ_2 (element-wise). By Lemma B.1, $\sup_\theta|\frac{D_{\psi_j} p_{yd,z}(\theta)}{p_{yd,z}(\theta)}|<c_1$, $\sup_\theta|D_{\theta_k}\{\frac{D_{\psi_j} p_{yd,z}(\theta)}{p_{yd,z}(\theta)}\}|<c_2$ and $\sup_\theta|D_{\theta_l}\{\frac{D_{\psi_j\psi_k} p_{yd,z}(\theta)}{p_{yd,z}(\theta)}\}|<c_3$ for some positive constants c_1, c_2 and c_3 and, therefore, combining the inequalities,

$$
\begin{aligned}
\left\|\rho_{\psi\psi}(w,\theta_1)-\rho_{\psi\psi}(w,\theta_2)\right\| \le{}& \sum_{y,d,z=0,1}\sum_{j=1}^{d_\psi}2c_1\sum_{j=1}^{d_\psi}\sum_{k=1}^{d_\theta}c_2\|\theta_1-\theta_2\| \\
&+\sum_{y,d,z=0,1}\sum_{j,k=1}^{d_\psi}\sum_{l=1}^{d_\theta}c_3\|\theta_1-\theta_2\|. \qquad \text{(B.20)}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\|\rho^{\dagger}_{\theta\theta}(w,\theta_1)-\rho^{\dagger}_{\theta\theta}(w,\theta_2)\| \\
&\quad\le \sum_{y,d,z=0,1}\left\|\frac{D_\theta p^{\dagger}_{yd,z}(\theta_1)D_\theta p^{\dagger}_{yd,z}(\theta_1)'}{p_{yd,z}(\theta_1)^2}-\frac{D_\theta p^{\dagger}_{yd,z}(\theta_2)D_\theta p^{\dagger}_{yd,z}(\theta_2)'}{p_{yd,z}(\theta_2)^2}\right\| \\
&\quad\le \sum_{y,d,z=0,1}\left(\left\|\frac{D_\theta p^{\dagger}_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}\right\|+\left\|\frac{D_\theta p^{\dagger}_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right\|\right)\left\|\frac{D_\theta p^{\dagger}_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}-\frac{D_\theta p^{\dagger}_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right\| \\
&\quad\le \sum_{y,d,z=0,1}\sum_{j=1}^{d_\theta}\left(\left|\frac{D_{\theta_j} p^{\dagger}_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}\right|+\left|\frac{D_{\theta_j} p^{\dagger}_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right|\right)\sum_{j=1}^{d_\theta}\left|\frac{D_{\theta_j} p^{\dagger}_{yd,z}(\theta_1)}{p_{yd,z}(\theta_1)}-\frac{D_{\theta_j} p^{\dagger}_{yd,z}(\theta_2)}{p_{yd,z}(\theta_2)}\right|
\end{aligned}
$$

and by Lemma B.1, $\sup_\theta |\frac{D_{\theta_j} p^{\dagger}_{yd,z}(\theta)}{p_{yd,z}(\theta)}| < c_4$ and $\sup_\theta |D_{\theta_k}\{\frac{D_{\theta_j} p^{\dagger}_{yd,z}(\theta)}{p_{yd,z}(\theta)}\}| < c_5$ for some positive constants c_4 and c_5 and, therefore, by applying the mean value theorem as above,

$$
\|\rho^{\dagger}_{\theta\theta}(w,\theta_1)-\rho^{\dagger}_{\theta\theta}(w,\theta_2)\| \le \sum_{y,d,z=0,1}\sum_{j=1}^{d_\theta} 2c_4 \sum_{j,k=1}^{d_\theta} c_5\|\theta_1-\theta_2\|. \tag{B.21}
$$

By combining (B.20) and (B.21), we have the desired result. □

Verification of S3(ii)(b). For bounding $\|r(w,\theta_1)-r(w,\theta_2)\|$, the proof is very similar to the one above with $\|\rho^{\dagger}_{\theta\theta}(w,\theta_1)-\rho^{\dagger}_{\theta\theta}(w,\theta_2)\|$. Bounding $\|\rho^{\dagger}_{\theta}(w,\theta_1)-\rho^{\dagger}_{\theta}(w,\theta_2)\|$ can also be done analogously. □

Verification of S3(iii). First, $M_1(w)$ is finite and does not depend on w, as can be seen from the verification of S3(ii)(a). Now, since $|\mathbf{1}_{ydz}(w)| \le 1$

$$
\begin{aligned}
E_{\gamma_0}\sup_{\theta\in\Theta}\left|\rho(W_i,\theta)\right|^{1+\delta} &\le E_{\gamma_0}\left(\sum_{y,d,z=0,1}\sup_{\theta\in\Theta}\left|\mathbf{1}_{ydz}(w)\cdot\log p_{yd,z}(\theta)\right|\right)^{1+\delta} \\
&\le \left(\sum_{y,d,z=0,1}\sup_{\theta\in\Theta}\left|\log p_{yd,z}(\theta)\right|\right)^{1+\delta},
\end{aligned}
$$

which is bounded since $p_{yd,z}(\theta)$ is bounded away from zero for any $\theta\in\Theta$ and $(y,d,z)\in\{0,1\}$ by Lemma B.1. Next,

$$
\begin{aligned}
&E_{\gamma_0}\sup_{\theta\in\Theta}\left\|\rho_{\psi\psi}(W_i,\theta)\right\|^{1+\delta} \\
&\quad\le E_{\gamma_0}\left(\sum_{y,d,z=0,1}\sup_{\theta\in\Theta}\left\|\mathbf{1}_{ydz}(w)\left[-\frac{1}{p_{yd,z}(\theta)^2}D_\psi p_{yd,z}(\theta)D_\psi p_{yd,z}(\theta)'\right.\right.\right. \\
&\qquad\left.\left.\left.+\frac{1}{p_{yd,z}(\theta)}D_{\psi\psi}p_{yd,z}(\theta)\right]\right\|\right)^{1+\delta}
\end{aligned}
$$

$$\leq \Bigg(\sum_{y,d,z=0,1} C \sup_{\theta\in\Theta} \big\{ \big\| D_\psi p_{yd,z}(\theta) D_\psi p_{yd,z}(\theta)' \big\| + \big\| D_{\psi\psi} p_{yd,z}(\theta) \big\| \big\} \Bigg)^{1+\delta}$$

by Lemma B.1, where $\|D_\psi p_{yd,z}(\theta) D_\psi p_{yd,z}(\theta)'\| \leq \sum_{j=1}^{d_\psi} \|D_{\psi_j} p_{yd,z}(\theta)\|^2$, which is bounded by Lemma B.1, and similarly for $\|D_{\psi\psi} p_{yd,z}(\theta)\|$. Similar arguments to those used in the verification of S3(i)(b) and S3(ii)(a) provide the desired result for the remaining four terms in the assumption. □

Verification of S3(iv)(a). Note that, when $\beta_0 = 0$,

$$\begin{aligned} &E_{\gamma_0} \rho_{\psi\psi}(W_i, \psi_0, \pi) \\ &\quad = \sum_{y,d,z=0,1} p_{yd,z}(\theta_0) \phi_{z,0} \bigg[\frac{D_\psi p_{yd,z}(\psi_0, \pi) D_\psi p_{yd,z}(\psi_0, \pi)'}{p_{yd,z}(\psi_0, \pi)^2} - \frac{D_{\psi\psi} p_{yd,z}(\psi_0, \pi)}{p_{yd,z}(\psi_0, \pi)} \bigg] \\ &\quad = \sum_{y,d,z=0,1} \phi_{z,0} \bigg[\frac{D_\psi p_{yd,z}(\psi_0, \pi) D_\psi p_{yd,z}(\psi_0, \pi)'}{p^0_{yd}} - D_{\psi\psi} p_{yd,z}(\psi_0, \pi) \bigg] \\ &\quad = \sum_{y,d,z=0,1} \phi_{z,0} \frac{D_\psi p_{yd,z}(\psi_0, \pi) D_\psi p_{yd,z}(\psi_0, \pi)'}{p^0_{yd}}, \end{aligned}$$

where the second equality is by (B.6), and the third equality is by (B.15) and (B.16). Let $M_{yd,z} \equiv D_\psi p_{yd,z}(\psi_0, \pi) D_\psi p_{yd,z}(\psi_0, \pi)'$ and $\tilde{M}_{yd,z} \equiv M_{yd,z}/p^0_{yd}$ so that

$$E_{\gamma_0} \rho_{\psi\psi}(W_i, \psi_0, \pi) = \phi_{1,0} \sum_{y,d=0,1} \tilde{M}_{yd,1} + \phi_{0,0} \sum_{y,d=0,1} \tilde{M}_{yd,0}. \tag{B.22}$$

Let $h_3(\pi) \equiv h_3(\zeta_{10}, \zeta_{30}; \pi)$ and $h_2(\pi) \equiv h_2(\zeta_{10}, \zeta_{20}; \pi)$. Note that when $\beta_0 = 0$, the $D_\psi p_{yd,z}(\psi_0, \pi)$ terms can be expressed as

$$D_\psi p_{11,0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad D_\psi p_{10,0} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad D_\psi p_{01,0} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \quad D_\psi p_{00,0} = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \end{bmatrix},$$

$$D_\psi p_{11,1} = \begin{bmatrix} C_2\big(h_3(\pi), \zeta_1; \pi\big) \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad D_\psi p_{10,1} = \begin{bmatrix} -C_2\big(h_2(\pi), \zeta_1; \pi\big) \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

and

$$D_\psi p_{01,1} = \begin{bmatrix} 1 - C_2\big(h_3(\pi), \zeta_1; \pi\big) \\ 1 \\ 0 \\ -1 \end{bmatrix}, \quad D_\psi p_{00,1} = \begin{bmatrix} -1 + C_2\big(h_2(\pi), \zeta_1; \pi\big) \\ -1 \\ -1 \\ 0 \end{bmatrix},$$

where, in $D_\psi p_{11,1}$ and $D_\psi p_{10,1}$,

$$C_2(h_3,\zeta_1;\pi)+C_1(h_3,\zeta_1;\pi)h_{3,\zeta_1}=0, \tag{B.23}$$

$$C_1(h_3,\zeta_1;\pi)h_{3,\zeta_3}=1, \tag{B.24}$$

$$h_{2,\zeta_1}-C_2(h_2,\zeta_1;\pi)-C_1(h_2,\zeta_1;\pi)h_{2,\zeta_1}=0, \tag{B.25}$$

$$h_{2,\zeta_2}-C_1(h_2,\zeta_1;\pi)h_{2,\zeta_2}=1, \tag{B.26}$$

by differentiating the objects in (7.1)–(7.2) of HM18 w.r.t. ζ_1, ζ_2 and ζ_3 and (B.5). Let $c\equiv C_2(h_3(\pi),\zeta_{10};\pi)$ and $\tilde{c}\equiv C_2(h_2(\pi),\zeta_{10};\pi)$ for notational simplicity. Then

$$M_{11,1}=\begin{bmatrix} c^2 & 0 & 0 & c\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ c & 0 & 0 & 1\end{bmatrix},\qquad M_{10,1}=\begin{bmatrix} \tilde{c}^2 & 0 & -\tilde{c} & 0\\ 0 & 0 & 0 & 0\\ -\tilde{c} & 0 & 1 & 0\\ 0 & 0 & 0 & 0\end{bmatrix},$$

$$M_{01,1}=\begin{bmatrix} (1-c)^2 & 1-c & 0 & c-1\\ 1-c & 1 & 0 & -1\\ 0 & 0 & 0 & 0\\ c-1 & -1 & 0 & 1\end{bmatrix},\qquad M_{00,1}=\begin{bmatrix} (1-\tilde{c})^2 & 1-\tilde{c} & 1-\tilde{c} & 0\\ 1-\tilde{c} & 1 & 1 & 0\\ 1-\tilde{c} & 1 & 1 & 0\\ 0 & 0 & 0 & 0\end{bmatrix},$$

$$M_{11,0}=\begin{bmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix},\qquad M_{10,0}=\begin{bmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0\end{bmatrix},$$

$$M_{01,0}=\begin{bmatrix} 0 & 0 & 0 & 0\\ 0 & 1 & 0 & -1\\ 0 & 0 & 0 & 0\\ 0 & -1 & 0 & 1\end{bmatrix},\qquad M_{00,0}=\begin{bmatrix} 0 & 0 & 0 & 0\\ 0 & 1 & 1 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}.$$

By Weyl (1912),

$$\lambda_{\min}(A+B)\geq\lambda_{\min}(A)+\lambda_{\min}(B) \tag{B.27}$$

for symmetric matrices A and B. Thus, for (B.22),

$$\lambda_{\min}\big(E_{\gamma_0}\rho_{\psi\psi}(W_i,\psi_0,\pi)\big)\geq\lambda_{\min}\bigg(\phi_{1,0}\sum_{y,d=0,1}\tilde{M}_{yd,1}\bigg)+\lambda_{\min}\bigg(\phi_{0,0}\sum_{y,d=0,1}\tilde{M}_{yd,0}\bigg).$$

The second term on the right-hand side satisfies $\lambda_{\min}(\phi_{0,0}\sum_{y,d=0,1}\tilde{M}_{yd,0})\geq\phi_{0,0}\sum_{y,d=0,1}\lambda_{\min}(\tilde{M}_{yd,0})=0$ by (B.27), the above expressions for the $M_{yd,0}$'s and since $\lambda_{\min}(\tilde{M}_{yd,0})=\lambda_{\min}(M_{yd,0})=0$ because $p^0_{yd}>0$ for all (y,d) by Lemma B.1(v). The first term on the right-hand side satisfies $\lambda_{\min}(\phi_{1,0}\sum_{y,d=0,1}\tilde{M}_{yd,1})\geq\phi_{1,0}\lambda_{\min}(\{\tilde{M}_{11,1}+\tilde{M}_{01,1}+\tilde{M}_{00,1}\})$ by (B.27) and since $\lambda_{\min}(\tilde{M}_{10,1})=\lambda_{\min}(M_{10,1})=0$. Now we prove $\lambda_{\min}(\tilde{M}_{11,1}+\tilde{M}_{01,1}+\tilde{M}_{00,1})>0$, which then implies that $\lambda_{\min}(E_{\gamma_0}\rho_{\psi\psi}(W_i,\psi_0,\pi))>0$ as desired since $\phi_{1,0}>0$ by TC5(ii). Under TC5(i) and by Lemma B.1(v), let $a\equiv p^0_{11}/p^0_{01}$ and

$b \equiv p^0_{11}/p^0_{00}$ for simplicity. Then $\tilde{M}_{11,1} + \tilde{M}_{01,1} + \tilde{M}_{00,1} = (M_{11,1} + aM_{01,1} + bM_{00,1})/p^0_{11}$ and

$$\begin{aligned} M &\equiv M_{11,1} + aM_{01,1} + bM_{00,1} \\ &= \begin{bmatrix} a(1-c)^2 + b(1-\tilde{c})^2 + c^2 & a(1-c) + b(1-\tilde{c}) & b(1-\tilde{c}) & -a(1-c) + c \\ a(1-c) + b(1-\tilde{c}) & a+b & b & -a \\ b(1-\tilde{c}) & b & b & 0 \\ -a(1-c) + c & -a & 0 & a+1 \end{bmatrix}. \end{aligned}$$

Then one can easily show the following: For the kth leading principal minor $|M_k|$ and determinant $|M|$ of M,

$$\begin{aligned} |M_1| &= a(1-c)^2 + b(1-\tilde{c})^2 + c^2 > 0, \\ |M_2| &= ab\big[(1-c) + (1-\tilde{c})\big]^2 + (a+b)c^2 > 0, \\ |M_3| &= ab\tilde{c}^2 > 0, \\ |M| &= ab\big[a(2c-1)^2 + b(\tilde{c}-1)^2\big] > 0 \end{aligned}$$

and, therefore, M is positive definite and so is M/p^0_{11}, that is, $\lambda_{\min}(\tilde{M}_{11,1} + \tilde{M}_{01,1} + \tilde{M}_{00,1}) > 0$. □

Verification of S3(iv)(b). We divide this proof into two cases: (i) $\beta_0 \neq 0$ and (ii) $\beta_0 = 0$.

Case (i): Note that by S3(i)(a),

$$E_{\gamma_0} B^{-1}(\beta_0)\rho_{\theta\theta}(W_i, \theta_0)B^{-1}(\beta_0) = E_{\gamma_0}\rho^\dagger_{\theta\theta}(W_i, \theta_0).$$

First, note that $E_{\gamma_0}\rho_{\theta\theta}(w, \theta_0)$ is positive definite by the information matrix equality with the fact that the information matrix is nonsingular by the identification result of HV17. Thus, for a nonzero vector $a \in \mathbb{R}^{d_\theta}$, $a' E_{\gamma_0}\rho_{\theta\theta}(w, \theta_0)a > 0$, which implies that, for a nonzero vector $\tilde{a} \in \mathbb{R}^{d_\theta}$, $\tilde{a}' E_{\gamma_0}\rho^\dagger_{\theta\theta}(w, \theta_0)\tilde{a} = \tilde{a}' B^{-1}(\beta_0) E_{\gamma_0}\rho_{\theta\theta}(w, \theta_0) B^{-1}(\beta_0)\tilde{a} > 0$. Therefore, $E_{\gamma_0}\rho^\dagger_{\theta\theta}(w, \theta_0)$ is positive definite.

Case (ii): First, note that by (B.12)–(B.14) and (B.23)–(B.26), we can express $D_\theta p^\dagger_{yd,z}(\psi_0, \pi)$'s as follows when $\beta_0 = 0$,

$$\begin{aligned} &D_\theta p^\dagger_{11,0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \qquad D_\theta p^\dagger_{10,0} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \\ &D_\theta p^\dagger_{01,0} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \qquad D_\theta p^\dagger_{00,0} = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \end{aligned} \tag{B.28}$$

$$D_\theta p^\dagger_{11,1} = \begin{bmatrix} C_2\big(h_3(\pi), \zeta_{10}; \pi\big) \\ 0 \\ 0 \\ 1 \\ C_{\pi 2}\big(h_3(\pi), \zeta_{10}; \pi\big) + C_{12}\big(h_3(\pi), \zeta_{10}; \pi\big) h_{3,\pi}(\zeta_{10}, \zeta_{30}, \pi) \end{bmatrix}, \tag{B.29}$$

$$D_\theta p^\dagger_{10,1} = \begin{bmatrix} -C_2\big(h_2(\pi), \zeta_{10}; \pi\big) \\ 0 \\ 1 \\ 0 \\ -C_{\pi 2}\big(h_2(\pi), \zeta_{10}; \pi\big) - C_{12}\big(h_2(\pi), \zeta_{10}; \pi\big) h_{2,\pi}(\zeta_{10}, \zeta_{20}, \pi) \end{bmatrix}, \tag{B.30}$$

and

$$D_\theta p^\dagger_{01,1} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} - D_\theta p^\dagger_{11,1}, \qquad D_\theta p^\dagger_{00,1} = \begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} - D_\theta p^\dagger_{10,1}. \tag{B.31}$$

The remaining arguments are similar to those used to verify S3(iv)(a): Let $M^\dagger_{yd,z} \equiv D_\theta p^\dagger_{yd,z}(\theta_0) \times D_\theta p^\dagger_{yd,z}(\theta_0)'$ and $\tilde{M}^\dagger_{yd,z} \equiv M^\dagger_{yd,z}/p^0_{yd}$. Then

$$E_{\gamma_0} \rho^\dagger_{\theta\theta}(W_i, \theta_0) = E_{\gamma_0} \rho^\dagger_\theta(W_i, \theta_0) \rho^\dagger_\theta(W_i, \theta_0)' = \phi_{1,0} \sum_{y,d=0,1} \tilde{M}^\dagger_{yd,1} + \phi_{0,0} \sum_{y,d=0,1} \tilde{M}^\dagger_{yd,0}. \tag{B.32}$$

For notational simplicity, let $c \equiv C_2(h_3(\pi_0), \zeta_{10}; \pi_0)$ and $\tilde{c} \equiv C_2(h_2(\pi_0), \zeta_{10}; \pi_0)$. Also let $d \equiv C_{\pi 2}(h_3(\pi_0), \zeta_{10}; \pi_0) + C_{12}(h_3(\pi_0), \zeta_{10}; \pi_0) h_{3,\pi}(\zeta_{10}, \zeta_{30}, \pi_0)$ and $\tilde{d} \equiv C_{\pi 2}(h_2(\pi_0), \zeta_{10}; \pi_0) + C_{12}(h_2(\pi_0), \zeta_{10}; \pi_0) h_{2,\pi}(\zeta_{10}, \zeta_{20}, \pi_0)$. Therefore,

$$M^\dagger_{11,1} = \begin{bmatrix} c^2 & 0 & 0 & c & cd \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ c & 0 & 0 & 1 & d \\ cd & 0 & 0 & d & d^2 \end{bmatrix}, \qquad M^\dagger_{01,1} = \begin{bmatrix} (1-c)^2 & 1-c & 0 & c-1 & (c-1)d \\ 1-c & 1 & 0 & -1 & -d \\ 0 & 0 & 0 & 0 & 0 \\ c-1 & -1 & 0 & 1 & d \\ (c-1)d & -d & 0 & d & d^2 \end{bmatrix},$$

$$M^\dagger_{10,1} = \begin{bmatrix} \tilde{c}^2 & 0 & -\tilde{c} & 0 & \tilde{c}\tilde{d} \\ 0 & 0 & 0 & 0 & 0 \\ -\tilde{c} & 0 & 1 & 0 & -\tilde{d} \\ 0 & 0 & 0 & 0 & 0 \\ \tilde{c}\tilde{d} & 0 & -\tilde{d} & 0 & \tilde{d}^2 \end{bmatrix}, \qquad M^\dagger_{00,1} = \begin{bmatrix} (1-\tilde{c})^2 & 1-\tilde{c} & 1-\tilde{c} & 0 & (\tilde{c}-1)\tilde{d} \\ 1-\tilde{c} & 1 & 1 & 0 & -\tilde{d} \\ 1-\tilde{c} & 1 & 1 & 0 & -\tilde{d} \\ 0 & 0 & 0 & 0 & 0 \\ (\tilde{c}-1)\tilde{d} & -\tilde{d} & -\tilde{d} & 0 & \tilde{d}^2 \end{bmatrix},$$

$$M^\dagger_{11,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad M^\dagger_{01,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$M^{\dagger}_{10,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad M^{\dagger}_{00,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

By Lemma B.1, in analogy to the verification of S3(iv)(a), since $\sum_{y,d=0,1}\lambda_{\min}(\tilde{M}^{\dagger}_{yd,0}) = \lambda_{\min}(\tilde{M}^{\dagger}_{00,1}) = 0$, we consider the rest of the sum in (B.32) and apply (B.27). Let $a \equiv p^0_{11}/p^0_{01}$ and $b \equiv p^0_{11}/p^0_{10}$. Then $\tilde{M}^{\dagger}_{11,1} + \tilde{M}^{\dagger}_{01,1} + \tilde{M}^{\dagger}_{10,1} = (M^{\dagger}_{11,1} + aM^{\dagger}_{01,1} + bM^{\dagger}_{10,1})/p^0_{11}$ and

$$\begin{aligned} M^{\dagger} &\equiv M^{\dagger}_{11,1} + aM^{\dagger}_{01,1} + bM^{\dagger}_{10,1} \\ &= \begin{bmatrix} a(1-c)^2 + b\tilde{c}^2 + c^2 & a(1-c) & -b\tilde{c} & -a(1-c)+c & a(c-1)d + b\tilde{c}\tilde{d} + cd \\ a(1-c) & a & 0 & -a & -ad \\ -b\tilde{c} & 0 & b & 0 & -b\tilde{d} \\ -a(1-c)+c & -a & 0 & a+1 & (a+1)d \\ a(c-1)d + b\tilde{c}\tilde{d} + cd & -ad & -b\tilde{d} & (a+1)d & (a+1)d^2 + b\tilde{d}^2 \end{bmatrix}. \end{aligned}$$

For the kth leading principal minor $|M^{\dagger}_k|$ of $M^{\dagger}$,

$$\begin{aligned} |M^{\dagger}_1| &= a(1-c)^2 + b\tilde{c}^2 + c^2 > 0, \\ |M^{\dagger}_2| &= ab\tilde{c}^2 + ac^2 > 0, \\ |M^{\dagger}_3| &= abc^2 > 0, \\ |M^{\dagger}_4| &= a^2b(1-c)^2 + abc^2 > 0, \\ |M^{\dagger}_5| &= |M^{\dagger}| \\ &= ab\big\{a^2\big(1+(1-c)^2\big)d^2 + b^2\tilde{c}^2\tilde{d}^2 + c^2\big(d^2 + b\tilde{d}^2\big) + a\big(\big((1-c)^2 + c^2\big)d^2 + bc^2\tilde{d}^2\big)\big\} > 0. \end{aligned}$$

Therefore, $\tilde{M}^{\dagger}_{01,1} + \tilde{M}^{\dagger}_{10,1} + \tilde{M}^{\dagger}_{11,1}$ is positive definite and by (B.27), we can easily show that $\lambda_{\min}(E_{\gamma_0}\rho^{\dagger}_{\theta\theta}(W_i, \theta_0)) > 0$. □

Verification of S3(v). Recall

$$V^{\dagger}(\theta_1, \theta_2; \gamma_0) \equiv \mathrm{Cov}_{\gamma_0}\big(\rho^{\dagger}_{\theta}(W_i, \theta_1), \rho^{\dagger}_{\theta}(W_i, \theta_2)\big).$$

But

$$\begin{aligned} \mathrm{Cov}_{\gamma_0}\big(\rho^{\dagger}_{\theta}(W_i, \theta_0), \rho^{\dagger}_{\theta}(W_i, \theta_0)\big) &= E_{\gamma_0}\rho^{\dagger}_{\theta}(W_i, \theta_0)\rho^{\dagger}_{\theta}(W_i, \theta_0)' \\ &= E_{\gamma_0}\rho^{\dagger}_{\theta\theta}(W_i, \theta_0), \end{aligned} \tag{B.33}$$

where the first equality is by $E_{\gamma_0}\rho^{\dagger}_{\theta}(W_i, \theta_0) = B^{-1}(\beta_0)E_{\gamma_0}\rho_{\theta}(w, \theta_0) = 0$ and the second equality is by the definition of $\rho^{\dagger}_{\theta}(W_i, \theta)$ and $\rho^{\dagger}_{\theta\theta}(W_i, \theta)$. Since $E_{\gamma_0}\rho^{\dagger}_{\theta\theta}(W_i, \theta_0)$ is positive definite from S3(iv)(b), we have the desired result. □

Define the $d_\psi \times d_\beta$ matrix-valued function

$$K(\theta;\gamma_0) \equiv \frac{\partial}{\partial \beta_0'} E_{\gamma_0} \rho_\psi(W_i,\theta) \tag{B.34}$$

with domain $\Theta_\delta \times \Gamma_0$, where $\Theta_\delta \equiv \{\theta \in \Theta : |\beta| < \delta\}$ and

$$\Gamma_0 \equiv \left\{\gamma_a = (a\beta, \zeta, \pi, \phi) \in \Gamma : \gamma = (\beta, \zeta, \pi, \phi) \in \Gamma \text{ with } |\beta| < \delta \text{ and } a \in [0,1]\right\}$$

for some $\delta > 0$.

Assumption S4. (i) *$K(\theta;\gamma_0)$ exists $\forall(\theta,\gamma_0) \in \Theta_\delta \times \Gamma_0$.*

(ii) *$K(\theta;\gamma^*)$ is continuous in (θ,γ^*) at $(\theta,\gamma^*) = ((\psi_0,\pi),\gamma_0)$ uniformly over $\pi \in \Pi$ $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$, where ψ_0 is a subvector of γ_0.*

Verification of S4(i). Note that

$$\begin{aligned} K(\theta;\gamma_0) &\equiv \frac{\partial}{\partial \beta_0} E_{\gamma_0} \rho_\psi(W_i,\theta) \\ &= -\frac{\partial}{\partial \beta_0} \sum_{y,d,z=0,1} \frac{p_{yd,z}(\theta_0)\phi_{z,0}}{p_{yd,z}(\theta)} D_\psi p_{yd,z}(\theta) \\ &= -\sum_{y,d,z=0,1} \frac{\partial p_{yd,z}(\theta_0)}{\partial \beta_0} \frac{\phi_{z,0}}{p_{yd,z}(\theta)} D_\psi p_{yd,z}(\theta), \end{aligned}$$

where $\frac{\partial p_{yd,z}(\theta_0)}{\partial \beta_0}$ is the first element of $D_{\psi_0} p_{yd,z}(\theta_0)$ for all (y,d,z), whose expressions are above. □

Verification of S4(ii). For

$$K(\pi;\gamma_0) \equiv K(\psi_0,\pi;\gamma_0) = -\sum_{y,d,z=0,1} \frac{\partial p_{yd,z}(\theta_0)}{\partial \beta_0} \frac{\phi_{z,0}}{p_{yd,z}(\psi_0,\pi)} D_\psi p_{yd,z}(\psi_0,\pi),$$

let $a_{yd,z}(\pi,\theta_0,\phi_{1,0}) \equiv \frac{\partial p_{yd,z}(\theta_0)}{\partial \beta_0} \frac{\phi_{z,0}}{p_{yd,z}(\psi_0,\pi)} D_\psi p_{yd,z}(\psi_0,\pi)$ since $\phi_{0,0} = 1-\phi_{1,0}$. Note that $a_{yd,z}(\pi,\theta_0,\phi_{1,0})$ is continuous in its arguments by Lemma B.1(iv). We can show that $a_{yd,z}(\pi,\theta_0,\phi_{1,0})$ is continuous uniformly in $\pi \in \Pi$ by applying the uniform convergence result in Lemma 9.2 of ACMLwp to $a_{yd,z}(\pi,\theta_n,\phi_{1,n}) - a_{yd,z}(\pi,\theta_0,\phi_{1,0})$, using (i) the pointwise convergence (i.e., pointwise continuity) above, (ii) $a_{yd,z}(\pi,\theta_0,\phi_{1,0})$'s differentiability in π with derivatives bounded over $\pi \in \Pi$ by Lemma B.1 and (iii) the compactness of Π (B1(iii) below). □

Next, we impose conditions on the parameter spaces Θ and Γ. Recall $\Theta_\delta^* \equiv \{\theta \in \Theta^* : |\beta| < \delta\}$, where Θ^* is the true parameter space for θ. The "optimization parameter space" Θ satisfies the following.

ASSUMPTION B1 (AC13). (i) $\text{int}(\Theta) \supset \Theta^*$.

(ii) *For some* $\delta > 0$, $\Theta \supset \{\beta \in \mathbb{R}^{d_\beta} : |\beta| < \delta\} \times \mathcal{Z}^0 \times \Pi \supset \Theta^*_\delta$ *for some nonempty open set* $\mathcal{Z}^0 \subset \mathbb{R}^{d_\zeta}$ *and* Π.

(iii) Π *is compact.*

The following general results are useful in verifying B1 and B2 below: for a continuous function f, (i) if a set A is compact, then $f(A)$ is compact and (ii) $f^{-1}(\text{int}(A)) \subset \text{int}(f^{-1}(A))$ for any set A in the range of f, where the latter is necessary and sufficient for continuity. Also note that by definition, for a proper function f, if B is compact, then $f^{-1}(B)$ is compact. Lastly, for a function f, if $A \subset B$ then $f(A) \subset f(B)$.

VERIFICATION OF B1. TC3(ii) implies B1(i) since

$$\text{int}(\Theta) = \text{int}\big(\bar{h}^{-1}(\boldsymbol{\Theta})\big) \supset \bar{h}^{-1}\big(\text{int}(\boldsymbol{\Theta})\big) \supset \bar{h}^{-1}\big(\boldsymbol{\Theta}^*\big) = \Theta^*,$$

where the first $\supset$ is by the continuity of $\bar{h}$ and the second $\supset$ is by TC3(ii) and $\bar{h}^{-1}$ being a function. For B1(ii), first note that given TC3(iii),

$$\bar{h}^{-1}(\boldsymbol{\Theta}) \supset \bar{h}^{-1}\big(\big\{\beta \in \mathbb{R}^{d_\beta} : \|\beta\| < \delta\big\} \times \boldsymbol{\mathcal{Z}}^0 \times \Pi\big) \supset \bar{h}^{-1}\big(\boldsymbol{\Theta}^*_\delta\big).$$

But $\bar{h}^{-1}(\boldsymbol{\Theta}) = \Theta$ and

$$\begin{aligned}
\bar{h}^{-1}\big(\boldsymbol{\Theta}^*_\delta\big) &= \big\{\theta \in \Theta^* : \bar{h}(\theta) \in \boldsymbol{\Theta}^*_\delta\big\} \\
&= \big\{\theta \in \Theta^* : \bar{h}(\theta) \in \boldsymbol{\Theta}^*, \big|\bar{h}_1(\theta)\big| < \delta\big\} \\
&= \big\{\theta \in \Theta^* : \theta \in \Theta^*, |\beta| < \delta\big\} \\
&= \Theta^*_\delta,
\end{aligned}$$

where the third equality is by $\bar{h}$ being a homeomorphism and $\bar{h}_1(\theta) = \beta$ being the first element of $\bar{h}$. Also, with $B_\delta \equiv \{\beta \in \mathbb{R}^{d_\beta} : |\beta| < \delta\}$,

$$\begin{aligned}
\bar{h}^{-1}\big(B_\delta \times \boldsymbol{\mathcal{Z}}^0 \times \Pi\big) &= \big\{\theta \in \Theta^* : \bar{h}(\theta) \in B_\delta \times \boldsymbol{\mathcal{Z}}^0 \times \Pi\big\} \\
&= B_\delta \times \big\{\mu \in \mathcal{M}^* : h(\mu) \in \boldsymbol{\mathcal{Z}}^0 \times \Pi\big\} \\
&= B_\delta \times h^{-1}\big(\boldsymbol{\mathcal{Z}}^0 \times \Pi\big) \\
&\equiv B_\delta \times \mathcal{Z}^0 \times \Pi,
\end{aligned}$$

where $\mathcal{M}^* = \{\mu \in \mathbb{R}^{d_\mu} : \theta = (\beta, \mu) \text{ for some } \theta \in \Theta^*\}$, the second equality holds since $\bar{h}(\theta) = (\beta, h(\mu))$ and the last equality holds by TC3(iv). Lastly, B1(iii) holds by TC3(i). □

ASSUMPTION B2 (AC13). (i) Γ *is compact and* $\Gamma = \{\gamma = (\theta, \phi) : \theta \in \Theta^*, \phi \in \Phi^*(\theta)\}$.

(ii) $\forall \delta > 0, \exists \gamma = (\beta, \zeta, \pi, \phi) \in \Gamma$ *with* $0 < \|\beta\| < \delta$.

(iii) $\forall \gamma = (\beta, \zeta, \pi, \phi) \in \Gamma$ *with* $0 < \|\beta\| < \delta$ *for some* $\delta > 0$, $\gamma_a = (a\beta, \zeta, \pi, \phi) \in \Gamma$ $\forall a \in [0, 1]$.

Verification of B2. Consider B2(i). Under TC4(i), define $\Phi^*(\theta)$ as $\Phi^*(\theta) \equiv \boldsymbol{\Phi}^*(\bar{h}(\theta))$. Since $\boldsymbol{\Gamma}$ is compact, $\boldsymbol{\Theta}^*$ and $\boldsymbol{\Phi}^*(\boldsymbol{\theta})$ are compact for $\boldsymbol{\theta} \in \boldsymbol{\Theta}^*$. Thus, $\Theta^* = \bar{h}^{-1}(\boldsymbol{\Theta}^*)$ is compact by the properness of $\bar{h}$. Also given (B.4), we have

$$\Phi^*(\theta) \equiv \boldsymbol{\Phi}^*\big(\bar{h}(\theta)\big) = \boldsymbol{\Phi}^* = [0.01, 0.99],$$

which is compact and, therefore, Γ is also compact. Next, TC4(ii) implies B2(ii). This is because, $\forall \delta > 0$, for $\boldsymbol{\gamma} = (\beta, \boldsymbol{\mu}, \phi)$ that satisfies TC4(ii), let γ in B2(ii) be $\gamma = (\beta, h^{-1}(\boldsymbol{\mu}), \phi)$, which is in Γ since $(\beta, \boldsymbol{\mu}) \in \boldsymbol{\Theta}^*$ implies $(\beta, h^{-1}(\boldsymbol{\mu})) = \bar{h}^{-1}(\beta, \boldsymbol{\mu}) \in \Theta^*$. To show that TC4(iii) implies B2(iii), note that for any $\gamma = (\beta, \zeta, \pi, \phi) \in \Gamma$ with $0 < |\beta| < \delta$ for some $\delta > 0$, $\boldsymbol{\gamma} = (\beta, h(\zeta, \pi), \phi) \in \boldsymbol{\Gamma}$. By TC4(iii), this implies that $\boldsymbol{\gamma}_a = (a\beta, h(\zeta, \pi), \phi) \in \boldsymbol{\Gamma}$ $\forall a \in [0, 1]$. Therefore, $\gamma_a = (a\beta, h^{-1}(h(\zeta, \pi)), \phi) \in \Gamma$. □

Define a "weighted noncentral chi-square" process $\{\xi(\pi; \gamma_0, b) : \pi \in \Pi\}$ by

$$\xi(\pi; \gamma_0, b) \equiv -\frac{1}{2}\big(G(\pi; \gamma_0) + K(\pi; \gamma_0)b\big)' H^{-1}(\pi; \gamma_0)\big(G(\pi; \gamma_0) + K(\pi; \gamma_0)b\big),$$

where $G(\pi; \gamma_0)$ is defined such that $G_n(\cdot) \Rightarrow G(\cdot; \gamma_0)$, where " $\Rightarrow$" denotes weak convergence, with

$$G_n(\pi) \equiv n^{-1/2} \sum_{i=1}^{n} \big(\rho_\psi(W_i; \psi_{0,n}, \pi) - E_{\gamma_n} \rho_\psi(W_i; \psi_{0,n}, \pi)\big)$$

and

$$H(\pi; \gamma_0) \equiv E_{\gamma_0} \rho_{\psi\psi}(W_i; \psi_0, \pi).$$

Assumption C6. *Each sample path of the stochastic process* $\{\xi(\pi; \gamma_0, b) : \pi \in \Pi\}$ *in some set* $A(\gamma_0, b)$ *with* $\Pr_{\gamma_0}(A(\gamma_0, b)) = 1$ *is minimized over* Π *at a unique point* (*which may depend on the sample path*), *denoted* $\pi^*(\gamma_0, b)$, $\forall \gamma_0 \in \Gamma$ *with* $\beta_0 = 0$, $\forall b$ *with* $\|b\| < \infty$.

In Assumption C6, $\pi^*(\gamma_0, b)$ is random. The following is a primitive sufficient condition for Assumption C6 for the case where β is scalar. Let $\rho_\psi(w, \theta) \equiv (\rho_\beta(w, \theta)', \rho_\zeta(w, \theta)')'$. When $\beta = 0$, $\rho_\zeta(w, \theta)'$ does not depend on π by Assumption S2(ii) and is denoted by $\rho_\zeta(w, \psi)'$. For $\beta_0 = 0$, define

$$\rho^*_\psi(W_i, \psi_0, \pi_1, \pi_2)' \equiv \big(\rho_\beta(W_i, \psi_0, \pi_1)', \rho_\beta(W_i, \psi_0, \pi_2)', \rho_\zeta(W_i, \psi_0)'\big)',$$
$$\Omega_G(\pi_1, \pi_2; \psi_0) \equiv \mathrm{Cov}_{\gamma_0}\big(\rho^*_\psi(W_i, \psi_0, \pi_1, \pi_2)', \rho^*_\psi(W_i, \psi_0, \pi_1, \pi_2)'\big).$$

Assumption C6†. (i) $d_\beta = 1$

(ii) $\Omega_G(\pi_1, \pi_2; \gamma_0)$ *is positive definite* $\forall \pi_1, \pi_2 \in \Pi$ *with* $\pi_1 \neq \pi_2$, $\forall \gamma_0 \in \Gamma$ *with* $\beta_0 = 0$.

Note that Assumptions S1–S3 and C6† imply C6; see Lemma 3.1 of AC13.

Verification of C6†(ii). Noting that $D_{\zeta} p_{yd,z}(\psi_0, \pi)$ does not depend on π when $\beta_0 = 0$ so that we may denote it $D_{\zeta} p_{yd,z}(\psi_0)$, define

$$D_{\psi} p^*_{yd,z}(\psi_0, \pi_1, \pi_2) \equiv \big(D_{\beta} p_{yd,z}(\psi_0, \pi_1)', D_{\beta} p_{yd,z}(\psi_0, \pi_2)', D_{\zeta} p_{yd,z}(\psi_0)'\big)'.$$

Then

$$\begin{aligned}\Omega_G(\pi_1, \pi_2; \psi_0) &= E_{\gamma_0} \rho^*_{\psi}(W_i, \psi_0, \pi_1, \pi_2) \rho^*_{\psi}(W_i, \psi_0, \pi_1, \pi_2)' \\ &= \sum_{y,d,z=0,1} \frac{\phi_{z,0}}{p^0_{yd}} D_{\psi} p^*_{yd,z}(\psi_0, \pi_1, \pi_2) D_{\psi} p^*_{yd,z}(\psi_0, \pi_1, \pi_2)',\end{aligned}$$

where the second equality follows from (B.6) and $D_{\psi} p^*_{yd,z}(\psi_0, \pi_1, \pi_2)$ can be expressed as

$$D_{\psi} p^*_{11,0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \qquad D_{\psi} p^*_{10,0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \qquad D_{\psi} p^*_{01,0} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \qquad D_{\theta} p^*_{00,0} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \\ 0 \end{bmatrix},$$

$$D_{\psi} p^*_{11,1} = \begin{bmatrix} C_2\big(h_3(\zeta_{10}, \zeta_{30}, \pi_1), \zeta_{10}; \pi_1\big) \\ C_2\big(h_3(\zeta_{10}, \zeta_{30}, \pi_2), \zeta_{10}; \pi_2\big) \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

$$D_{\psi} p^*_{10,1} = \begin{bmatrix} -C_2\big(h_2(\zeta_{10}, \zeta_{20}, \pi_1), \zeta_{10}; \pi_1\big) \\ -C_2\big(h_2(\zeta_{10}, \zeta_{20}, \pi_2), \zeta_{10}; \pi_2\big) \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

and

$$D_{\psi} p^*_{01,1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - D_{\psi} p^*_{11,1}, \qquad D_{\psi} p^*_{00,1} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ 0 \\ 0 \end{bmatrix} - D_{\psi} p^*_{10,1},$$

using (B.23)–(B.26). The remaining arguments are similar to those used in the verification of S3(iv)(a): Let $M^*_{yd,z} \equiv D_{\psi} p^*_{yd,z}(\psi_0, \pi_1, \pi_2) \times D_{\psi} p^*_{yd,z}(\psi_0, \pi_1, \pi_2)'$ and $\tilde{M}^*_{yd,z} \equiv M^*_{yd,z}/p^0_{yd}$. Then

$$\Omega_G(\pi_1, \pi_2; \psi_0) = \phi_{1,0} \sum_{y,d=0,1} \tilde{M}^*_{yd,1} + \phi_{0,0} \sum_{y,d=0,1} \tilde{M}^*_{yd,0}. \tag{B.35}$$

Let $c \equiv C_2(h_3(\zeta_{10}, \zeta_{30}, \pi_1), \zeta_{10}; \pi_1)$, $\tilde{c} \equiv C_2(h_3(\zeta_{10}, \zeta_{30}, \pi_2), \zeta_{10}; \pi_2)$, $d \equiv C_2(h_2(\zeta_{10}, \zeta_{20}, \pi_1), \zeta_{10}; \pi_1)$, and $\tilde{d} \equiv C_2(h_2(\zeta_{10}, \zeta_{20}, \pi_2), \zeta_{10}; \pi_2)$ for notational simplicity. Then

$$M^*_{11,1} = \begin{bmatrix} c^2 & c\tilde{c} & 0 & 0 & c \\ c\tilde{c} & \tilde{c}^2 & 0 & 0 & \tilde{c} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ c & \tilde{c} & 0 & 0 & 1 \end{bmatrix},$$

$$M^*_{01,1} = \begin{bmatrix} (1-c)^2 & (1-c)(1-\tilde{c}) & 1-c & 0 & -(1-c) \\ (1-c)(1-\tilde{c}) & (1-\tilde{c})^2 & 1-\tilde{c} & 0 & -(1-\tilde{c}) \\ 1-c & 1-\tilde{c} & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ -(1-c) & -(1-\tilde{c}) & -1 & 0 & 1 \end{bmatrix},$$

$$M^*_{10,1} = \begin{bmatrix} d^2 & d\tilde{d} & 0 & d & 0 \\ d\tilde{d} & \tilde{d}^2 & 0 & \tilde{d} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ d & \tilde{d} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$M^*_{00,1} = \begin{bmatrix} (1-d)^2 & (1-d)(1-\tilde{d}) & 1-d & 1-d & 0 \\ (1-d)(1-\tilde{d}) & (1-\tilde{d})^2 & 1-\tilde{d} & 1-\tilde{d} & 0 \\ 1-d & 1-\tilde{d} & 1 & 1 & 0 \\ 1-d & 1-\tilde{d} & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$M^*_{11,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad M^*_{01,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$M^*_{10,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad M^*_{00,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

By Lemma B.1 and similar arguments to those used to verify S3(iv)(a), since $\sum_{y,d=0,1} \lambda_{\min}(\tilde{M}^*_{yd,0}) = \lambda_{\min}(\tilde{M}^*_{00,1}) = 0$, we consider the rest of the sum in (B.35) and apply (B.27). Let $a \equiv p^0_{01}/p^0_{10}$ and $b \equiv p^0_{01}/p^0_{11}$. Then, $\tilde{M}^*_{01,1} + \tilde{M}^*_{10,1} + \tilde{M}^*_{11,1} = (M^*_{01,1} + aM^*_{10,1} + bM^*_{11,1})/p^0_{01}$, and

$$M^* \equiv M^*_{01,1} + aM^*_{10,1} + bM^*_{11,1}$$

$$= \begin{bmatrix} ad^2+(1-c)^2+bc^2 & ad\tilde{d}+(1-c)(1-\tilde{c})+bc\tilde{c} & 1-c & ad & -(1-c)+bc \\ ad\tilde{d}+(1-c)(1-\tilde{c})+bc\tilde{c} & a\tilde{d}^2+(1-\tilde{c})^2+b\tilde{c}^2 & 1-\tilde{c} & a\tilde{d} & -(1-\tilde{c})+b\tilde{c} \\ 1-c & 1-\tilde{c} & 1 & 0 & -1 \\ ad & a\tilde{d} & 0 & a & 0 \\ -(1-c)+bc & -(1-\tilde{c})+b\tilde{c} & -1 & 0 & 1+b \end{bmatrix}.$$

For the kth leading principal minor $|M_k^*|$ and determinant $|M^*|$ of M^*,

$$\left|M_1^*\right| = ad^2 + (1-c)^2 > 0,$$

$$\left|M_2^*\right| = a\left\{\tilde{d}(1-c) - d(1-\tilde{c})\right\}^2 + b\left\{c(1-\tilde{c}) - \tilde{c}^2(1-c)\right\}^2 + ab(\tilde{d}c - d\tilde{c})^2 > 0,$$

$$\left|M_3^*\right| = ab(\tilde{d}c - d\tilde{c})^2 + ab(\tilde{d}c + d\tilde{c})^2 + 4bc\tilde{c}(1-c)(1-\tilde{c}) > 0,$$

$$\begin{aligned}\left|M_4^*\right| &= a\left\{a\tilde{d}^2(1-c)^2 + (1-c)^2(1-\tilde{c})^2\right\} \\ &\quad + ab\left\{(1-\tilde{c})^2c^2 + bc^2\tilde{c}^2 + ad^2\tilde{c}^2 + (1-c)^2\tilde{c}^2\right\} \\ &> 0,\end{aligned}$$

$$\begin{aligned}\left|M^*\right| &= ab\Big[a(\tilde{d}c - d\tilde{c})^2 + \left\{c(1-\tilde{c}) - \tilde{c}(1-c)\right\}^2 + a\left\{\left(\tilde{d}(1-c) - d\tilde{c}\right)^2 + (1-b)d^2\tilde{c}^2\right\} \\ &\quad + a^2d^2\tilde{d}^2 + (1-c)^2(1-\tilde{c})^2 + b(1-\tilde{c})^2c^2 + b^2c^2\tilde{c}^2 + b\tilde{c}^2(1-c)^2 + 2a\tilde{d}^2c(1-c)\Big] \\ &> 0.\end{aligned}$$

Therefore, $\tilde{M}_{01,1}^* + \tilde{M}_{10,1}^* + \tilde{M}_{11,1}^*$ is positive definite and by (B.27), we can easily show that $\lambda_{\min}(\Omega_G(\pi_1, \pi_2; \psi_0)) > 0$. □

Define a nonstochastic function $\{\eta(\pi; \gamma_0) : \pi \in \Pi\}$ by

$$\eta(\pi; \gamma_0) \equiv -\frac{1}{2}K(\pi; \gamma_0)' H^{-1}(\pi; \gamma_0) K(\pi; \gamma_0).$$

Assumption C7. *The nonstochastic function $\eta(\pi; \gamma_0)$ is uniquely minimized over $\pi \in \Pi$ at π_0 $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.*

For $\beta_0 = 0$, by (B.15)–(B.16) we can write

$$K(\pi; \gamma_0) = -\sum_{y,d,z=0,1} \frac{\phi_{z,0}}{p_{yd}^0} \frac{\partial p_{yd,z}(\theta_0)}{\partial \beta_0} D_\psi p_{yd,z}(\psi_0, \pi),$$

$$H(\pi; \gamma_0) = \sum_{y,d,z=0,1} \frac{\phi_{z,0}}{p_{yd}^0} D_\psi p_{yd,z}(\psi_0, \pi) D_\psi p_{yd,z}(\psi_0, \pi)'.$$

Note that we can partition $H(\pi)$ and $K(\pi)$, suppressing γ_0, as

$$H(\pi) = \begin{bmatrix} H_{11}(\pi) & H_{12}(\pi) \\ H_{21}(\pi) & H_{22} \end{bmatrix} \begin{matrix} \}d_\beta \\ \}d_\zeta \end{matrix} \quad \text{and} \quad K(\pi) = \begin{pmatrix} K_1(\pi) \\ K_2 \end{pmatrix} \begin{matrix} \}d_\beta \\ \}d_\zeta \end{matrix},$$

and note that $K(\pi_0) = [-H_{11}(\pi_0) : -H_{21}(\pi_0)']'$ by the expressions for $K(\pi; \gamma_0)$ and $H(\pi; \gamma_0)$.

Verification of C7. We first show that, for any $\pi \in \Pi$,

$$\eta(\pi) \geq \eta(\pi_0).$$

For matrices A and B, let $A \leq B$ denote $B - A$ being p.s.d. Then we can show that

$$K(\pi)' H^{-1}(\pi) K(\pi) \leq H_{11}(\pi_0) = K(\pi_0)' H^{-1}(\pi_0) K(\pi_0), \tag{B.36}$$

where the inequality is an application of the matrix Cauchy–Schwarz inequality (Proposition B.1 below) and the equality holds because $K(\pi_0) = [-H_{11}(\pi_0) : -H_{21}(\pi_0)']'$; see below for the proof. Lastly, the weak inequality in (B.36) holds as an equality if and only if $\rho_\beta(W_i, \psi_0, \pi_0)a + \rho_\psi(W_i, \psi_0, \pi)'b = 0$ with probability 1 for some $a \in \mathbb{R}$ and $b \in \mathbb{R}^{d_\psi}$ with $(a, b') \neq 0$. Let $D_\beta p^0_{yd,z} \equiv D_\beta p_{yd,z}(\psi_0, \pi_0)$ and $D_\psi p_{yd,z}(\pi) \equiv D_\psi p_{yd,z}(\psi_0, \pi)$ for simplicity. Then, when $\beta_0 = 0$,

$$\rho_\beta(W_i, \psi_0, \pi_0)a + \rho_\psi(W_i, \psi_0, \pi)'b = \sum_{y,d,z=0,1} \frac{\mathbf{1}_{ydz}(W_i)}{p^0_{yd}} \big[D_\beta p^0_{yd,z} a + D_\psi p_{yd,z}(\pi)' b\big].$$

But, it is easy to see that a $(1 + d_\psi) \times 8$ matrix (suppressing π in $D_\psi p_{yd,z}(\pi)$ and letting $h_{3,0} \equiv h_3(\pi_0)$ and $h_{2,0} \equiv h_2(\pi_0)$)

$$\begin{bmatrix} D_\beta p^0_{11,1} & D_\beta p^0_{10,1} & D_\beta p^0_{01,1} & D_\beta p^0_{00,1} & D_\beta p^0_{11,0} & D_\beta p^0_{10,0} & D_\beta p^0_{01,0} & D_\beta p^0_{00,0} \\ D_\psi p_{11,1} & D_\psi p_{10,1} & D_\psi p_{01,1} & D_\psi p_{00,1} & D_\psi p_{11,0} & D_\psi p_{10,0} & D_\psi p_{01,0} & D_\psi p_{00,0} \end{bmatrix}$$
$$= \begin{bmatrix} C_2(h_{3,0}; \zeta_{10}, \pi_0) & -C_2(h_{2,0}, \zeta_{10}; \pi_0) & 1 - C_2(h_{3,0}, \zeta_{10}; \pi_0) & -1 + C_2(h_{2,0}, \zeta_{10}; \pi_0) & 0 & 0 & 0 & 0 \\ C_2(h_3; \zeta_{10}, \pi_0) & -C_2(h_2, \zeta_{10}; \pi_0) & 1 - C_2(h_3, \zeta_{10}; \pi_0) & -1 + C_2(h_2, \zeta_{10}; \pi_0) & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \end{bmatrix}$$

has full row rank (i.e., rank of $1 + d_\psi$) except when $\pi = \pi_0$, since

$$C_2\big(h_3(\pi), \zeta_{10}; \pi\big) \neq C_2\big(h_3(\pi_0), \zeta_{10}; \pi_0\big),$$
$$C_2\big(h_2(\pi), \zeta_{10}; \pi\big) \neq C_2\big(h_2(\pi_0), \zeta_{10}; \pi_0\big)$$

for $\pi \neq \pi_0$. This can be shown by modifying the proof of Lemmas 3.1 and 4.1 of HV17 under Assumption TC2, which yields

$$\partial C_2\big(h_3(\pi), \zeta_1; \pi\big)/\partial \pi = C_{\pi 2}\big(h_3(\pi), \zeta_1; \pi\big) + C_{12}\big(h_3(\pi), \zeta_1; \pi\big) h_{3,\pi}(\pi) < 0$$

and

$$C_{\pi 2}\big(h_2(\pi), \zeta_1; \pi\big) + C_{12}\big(h_2(\pi), \zeta_1; \pi\big) h_{2,\pi}(\pi) < 0.$$

In fact, h_2 or h_3 can be seen as u_1^* in Lemma 4.1 of HV17. Therefore, there is no (a, b') with $(a, b') \neq 0$ such that $D_\beta p^0_{yd,z} a + D_\psi p'_{yd,z}(\pi) b = 0$ for all $(y, d, z) \in \{0, 1\}^3$, which implies that there is no (a, b') with $(a, b') \neq 0$ such that $\rho_\beta(W_i, \psi_0, \pi_0)a + \rho_\psi(W_i, \psi_0, \pi)'b = 0$ with probability 1. In other words, the equality holds uniquely at $\pi = \pi_0$ so that for any $\pi \neq \pi_0$, $\Pr[c'(\rho_\beta(W_i, \psi_0, \pi_0), \rho_\psi(W_i, \psi_0, \pi)')' = 0] < 1$ for all $c \in \mathbb{R}^{d_\beta + d_\psi}$ with $c \neq 0$, and thus the inequality in (B.36) is strict. □

PROPOSITION B.1. *Let $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$ be random vectors such that $E\|x\|^2 < \infty$, $E\|y\|^2 < \infty$, and Eyy' is nonsingular. Then*

$$\left(Exy'\right)\left(Eyy'\right)^{-1}\left(Eyx'\right) \leq Exx'.$$

For our verification proof, taking $x = \rho_\beta(W_i, \psi_0, \pi_0)$ and $y = \rho_\psi(W_i, \psi_0, \pi)$, we have

$$\begin{aligned} E_{\gamma_0} yy' &= H(\pi), \\ E_{\gamma_0} xx' &= H_{11}(\pi_0), \\ -E_{\gamma_0} xy' &= -\left(E_{\gamma_0} yx'\right)' = K(\pi). \end{aligned}$$

PROOF OF $H_{11}(\pi_0) = K(\pi_0)' H^{-1}(\pi_0) K(\pi_0)$. Define a 4×4 block-diagonalizing matrix

$$A(r) = \begin{bmatrix} 1 & -H_{12}(r) H_{22}^{-1} \\ 0_3 & I_3 \end{bmatrix}.$$

Then

$$\begin{aligned} K(r_0)' H^{-1}(r_0) K(r_0) &= K(r_0)' A(r) \left[A(r) H(r_0) A(r)\right]^{-1} A(r) K(r_0) \\ &= (-1)^2 \left[H_{11}(r_0) : H_{21}(r_0)'\right] A(r) \left[A(r) H(r_0) A(r)\right]^{-1} A(r) \begin{bmatrix} H_{11}(r_0) \\ H_{21}(r_0) \end{bmatrix} \\ &= \left[H_{11}(r_0) - H_{12}(r_0) H_{22}^{-1} H_{21}(r_0) : H_{21}(r_0)'\right] \begin{bmatrix} H_{11}^*(r_0)^{-1} & 0 \\ 0 & H_{22}^{-1} \end{bmatrix} \\ &\quad \times \begin{bmatrix} H_{11}(r_0) - H_{12}(r_0) H_{22}^{-1} H_{21}(r_0) \\ H_{21}(r_0) \end{bmatrix} \\ &= \left[1 : H_{21}(r_0)' H_{22}^{-1}\right] \begin{bmatrix} H_{11}(r_0) - H_{12}(r_0) H_{22}^{-1} H_{21}(r_0) \\ H_{21}(r_0) \end{bmatrix} \\ &= H_{11}(r_0), \end{aligned}$$

where the second equality is due to the fact that $K(r_0) = [-H_{11}(r_0) : -H_{21}(r_0)']'$ and $H_{11}^*(r_0)$ is implicitly defined. We also use the symmetry of $H(r)$ in this derivation. □

Define the following quantities that arise in the asymptotic distribution of $\hat{\theta}_n$ and the test statistics we consider. Letting $S_\psi \equiv [I_{d_\psi} : 0_{d_\psi \times 1}]$ denote the $d_\psi \times d_\theta$ selector matrix that selects ψ out of θ:

$$\begin{aligned} \Omega(\pi_1, \pi_2; \gamma_0) &\equiv S_\psi V^\dagger\left((\psi_0, \pi_1), (\psi_0, \pi_2); \gamma_0\right) S_\psi', \\ J(\theta; \gamma_0) &\equiv E_{\gamma_0} \rho_{\theta\theta}^\dagger(W_i; \theta), \\ V(\theta; \gamma_0) &= V^\dagger(\theta, \theta; \gamma_0), \end{aligned}$$

and

$$J(\gamma_0) \equiv J(\theta_0; \gamma_0),$$
$$V(\gamma_0) \equiv V(\theta_0; \gamma_0).$$

Note that

$$J(\gamma_0) = V(\gamma_0)$$

by (B.33). Define

$$\Sigma(\theta; \gamma_0) \equiv J^{-1}(\theta; \gamma_0) V(\theta; \gamma_0) J^{-1}(\theta; \gamma_0)$$

and

$$\Sigma(\pi; \gamma_0) \equiv \Sigma(\psi_0, \pi; \gamma_0).$$

ASSUMPTION V1. (i) *$\widehat{J}_n = \widehat{J}_n(\widehat{\theta}_n)$ and $\widehat{V}_n = \widehat{V}_n(\widehat{\theta}_n)$ for some (stochastic) $d_\theta \times d_\theta$ matrix-valued functions $\widehat{J}_n(\theta)$ and $\widehat{V}_n(\theta)$ on Θ that satisfy* $\sup_{\theta \in \Theta} \|\widehat{J}_n(\theta) - J(\theta; \gamma_0)\| \to_p 0$ *and* $\sup_{\theta \in \Theta} \|\widehat{V}_n(\theta) - V(\theta; \gamma_0)\| \to_p 0$ *under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$.*

(ii) *$J(\theta; \gamma_0)$ and $V(\theta; \gamma_0)$ are continuous in θ on Θ $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.*

(iii) *$\lambda_{\min}(\Sigma(\pi; \gamma_0)) > 0$ and $\lambda_{\max}(\Sigma(\pi; \gamma_0)) < \infty$ $\forall \pi \in \Pi$, $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.*

VERIFICATION OF V1(i). We define the following:

$$\begin{aligned}\widehat{J}_n(\theta) &\equiv \frac{1}{n}\sum_{i=1}^n \rho^\dagger_{\theta\theta}(W_i, \theta) = \frac{1}{n}\sum_{i=1}^n \rho^\dagger_\theta(W_i, \theta)\rho^\dagger_\theta(W_i, \theta)' \\ &= \frac{1}{n}\sum_{i=1}^n \sum_{y,d,z=0,1} \mathbf{1}_{ydz}(W_i) \frac{D_\theta p^\dagger_{yd,z}(\theta) D_\theta p^\dagger_{yd,z}(\theta)'}{p_{yd,z}(\theta)^2},\end{aligned}$$

where $D_\theta p^\dagger_{yd,z}(\theta_0)$ are defined above. Also,

$$\begin{aligned}\widehat{V}_n(\theta) &\equiv \frac{1}{n}\sum_{i=1}^n \rho^\dagger_\theta(W_i, \theta)\rho^\dagger_\theta(W_i, \theta)' \\ &= \frac{1}{n}\sum_{i=1}^n \sum_{y,d,z=0,1} \mathbf{1}_{ydz}(W_i) \frac{D_\theta p^\dagger_{yd,z}(\theta) D_\theta p^\dagger_{yd,z}(\theta)'}{p_{yd,z}(\theta)^2} = \widehat{J}_n(\theta).\end{aligned}$$

The rest of the proof follows from the uniform law of large numbers in Lemma 9.3 of ACMLwp with Assumptions S1 and S3 and Θ being compact. □

VERIFICATION OF V1(ii). The continuity follows from the fact that the first and second derivatives of $p_{yd,z}(\theta)$ are continuous by Lemma B.1(vi). □

Verification of V1(iii). Note that

$$\Sigma(\pi;\gamma_0)=J^{-1}(\psi_0,\pi;\gamma_0)V(\psi_0,\pi;\gamma_0)J^{-1}(\psi_0,\pi;\gamma_0)=V^{-1}(\psi_0,\pi;\gamma_0)$$

since $V(\psi_0,\pi;\gamma_0)=J(\psi_0,\pi;\gamma_0)$. This is because

$$\begin{aligned}V(\psi_0,\pi;\gamma_0)&=\mathrm{Cov}_{\gamma_0}\big(\rho^\dagger_\theta(W_i,\psi_0,\pi),\rho^\dagger_\theta(W_i,\psi_0,\pi)\big)=E_{\gamma_0}\rho^\dagger_\theta(W_i;\psi_0,\pi)\rho^\dagger_\theta(W_i;\psi_0,\pi)'\\&=E_{\gamma_0}\rho^\dagger_{\theta\theta}(W_i;\theta),\end{aligned}$$

where the last equality holds since $\rho^\dagger_{\theta\theta}(w,\theta)=\rho^\dagger_\theta(w,\theta)\rho^\dagger_\theta(w,\theta)'$, and the second-to-last equality holds since

$$\begin{aligned}E_{\gamma_0}\rho^\dagger_\theta(W_i,\psi_0,\pi)&=-\sum_{y,d,z=0,1}\phi_{z,0}D_\theta p^\dagger_{yd,z}(\psi_0,\pi)\\&=-\sum_{y,d=0,1}\phi_{0,0}D_\theta p^\dagger_{yd,0}(\psi_0,\pi)-\sum_{y,d=0,1}\phi_{1,0}D_\theta p^\dagger_{yd,1}(\psi_0,\pi)\\&=0.\end{aligned}$$

Now, for the first part of V1(iii), note that since each element of the vectors in (B.28)–(B.31) are bounded by TC2(iii) and B2(i), the elements of the matrix

$$\begin{aligned}V(\psi_0,\pi;\gamma_0)&=E_{\gamma_0}\rho^\dagger_\theta(W_i;\psi_0,\pi)\rho^\dagger_\theta(W_i;\psi_0,\pi)'\\&=\sum_{y,d,z=0,1}\frac{\phi_{z,0}}{p_{yd,0}}D_\theta p^\dagger_{yd,z}(\psi_0,\pi)D_\theta p^\dagger_{yd,z}(\psi_0,\pi)'\end{aligned}$$

are bounded. For a $d\times d$ matrix A, $\sum_{i=1}^d|\lambda_i|\le\sum_{i,j=1}^d|A_{ij}|$ where the λ_i's are A's eigenvalues and the A_{ij}'s are A's elements. Therefore, $\lambda_{\max}(V(\psi_0,\pi;\gamma_0))<\infty$. This implies that $\lambda_{\min}(V^{-1}(\psi_0,\pi;\gamma_0))>0$. By Lemma B.1, the proof of the second part is similar to the proofs of S3(iv)(b) and S3(v) and we can show that $\lambda_{\min}(V(\psi_0,\pi;\gamma_0))>0$, which implies that $\lambda_{\max}(V^{-1}(\psi_0,\pi;\gamma_0))<\infty$. □

References

Andrews, D. W. K. and X. Cheng (2012), "Supplement to 'Estimation and inference with weak, semi-strong and strong identification'." *Econometrica Supplementary Material*, 80. [9]

Andrews, D. W. K. and X. Cheng (2013a), "Maximum likelihood estimation and uniform inference with sporadic identification failure." *Journal of Econometrics*, 173, 36–56. [9]

Andrews, D. W. K. and X. Cheng (2013b), "Supplemental appendices for 'Maximum likelihood estimation and uniform inference with sporadic identification failure'." Addendum to Cowles Foundation Discussion Paper No. 1824R, Yale University. [9]

Andrews, D. W. K. and X. Cheng (2014), "Supplementary material on 'GMM estimation and uniform subvector inference with possible identification failure'." *Econometric Theory Supplement*, 30. [9]

Hadamard, J. (1906a), "Sur les Transformations Planes." *Comptes Rendus des Seances de l'Academie des Sciences, Paris*, 74, 142. [3]

Hadamard, J. (1906b), "Sur les Transformations Ponctuelles." *Bulletin de la Societe Mathematique de France*, 34, 71–84. [3]

Han, S. and A. McCloskey (2019), "Estimation and inference with a (nearly) singular Jacobian." *Quantitative Economics*, 10, 1019–1068. [1, 8]

Han, S. and E. Vytlacil (2017), "Identification in a generalization of bivariate probit models with dummy endogenous regressors." *Journal of Econometrics*, 199, 63–73. [9]

Lindelöf, E. (1894), "Sur l'application des méthodes d'approximations successives à l'étude des intégrales réelles des équations différentielles ordinaires." *Journal de Mathématiques Pures et Appliquées*, 117–128. [3]

McCloskey, A. (2017), "Bonferroni-based size-correction for nonstandard testing problems." *Journal of Econometrics*, 200, 17–35. [7, 8]

Picard, E. (1893), "Sur l'application des méthodes d'approximations successives à l'étude de certaines équations différentielles ordinaires." *Journal de Mathématiques Pures et Appliquées*, 217–272. [3]

Weyl, H. (1912), "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)." *Mathematische Annalen*, 71 (4), 441–479. [23]

Co-editor Andres Santos handled this manuscript.

Manuscript received 27 September, 2017; final version accepted 27 November, 2018; available online 12 February, 2019.