Supplemental Appendix to

# MAXIMUM LIKELIHOOD ESTIMATION AND UNIFORM INFERENCE WITH SPORADIC IDENTIFICATION FAILURE

By

**Donald W. K. Andrews and Xu Cheng**

**October 2011**
**Revised October 2012**

**COWLES FOUNDATION DISCUSSION PAPER NO. 1824R**

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS**
**YALE UNIVERSITY**
**Box 208281**
**New Haven, Connecticut 06520-8281**

**http://cowles.econ.yale.edu/**

# Supplemental Appendices

for

# Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Xu Cheng
Department of Economics
University of Pennsylvania

First Draft: August, 2007
Revised: November 3, 2012

# 10. Outline

This Supplement includes five Supplemental Appendices (denoted A-E) to the paper "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure." Supplemental Appendix A proves Theorems 4.1, 4.2, 5.1, and 5.2 of that paper. Supplemental Appendix B provides (i) asymptotic size results for the robust QLR CS's, (ii) a sophisticated method for choosing $\kappa$ for type 2 robust CS's, (iii) statements of Assumptions V1 and V2, which concern the estimator of the variance matrix of $\widehat{\theta}_n$, and (iv) an extension of the sufficient conditions for Assumption S3$^*$(i) given in Section 9.1 of the paper for $\rho(w, \theta)$ functions of the form $\rho^*(w, a(x, \beta)h(x, \pi), \zeta)$. The extension is to the case where a parameter $\zeta$ appears. Supplemental Appendix C provides additional numerical results to those provided in the main paper for both the smooth transition autoregressive (STAR) model and the nonlinear binary choice model. Supplemental Appendix D verifies Assumptions S1-S4, B1, B2, C6, V1, and V2 for the nonlinear binary choice model. Supplemental Appendix E does likewise for the STAR model.

We let AC1 abbreviate the paper Andrews and Cheng (2012) "Estimation and Inference with Weak, Semi-strong, and Strong Identification," *Econometrica* 80, forthcoming.

# 11. Supplemental Appendix A: Proofs

This Appendix proves the results in Theorems 4.1, 4.2, 5.1, and 5.2. The method of proof is to show that Assumptions B1, B2, and S1-S3 imply certain high-level assumptions in AC1 (specifically, Assumptions A, B3, C1-C4, C8, and D1-D3 of AC1). In addition, it is straightforward that Assumptions S1 and S4 imply Assumption C5 of AC1. Given these results, Theorems 3.1, 3.2, 4.2, 4.3, and 4.4(b) of AC1 imply Theorems 4.1, 4.2, 5.1(a), 5.1(b), and 5.2, respectively, because the results of these theorems are the same, just the assumptions differ.

**Lemma 11.1.** *Suppose Assumptions* B1 *and* B2 *hold. Assumptions* S1-S3 *imply that*

*Assumptions* A, B3, C1-C4, C8, *and* D1-D3 *of* AC1 *hold with*

$$Q(\theta;\gamma_0)=E_{\gamma_0}\rho(W_i,\theta),\ D_\psi Q_n(\theta)=n^{-1}\sum_{i=1}^n\rho_\psi(W_i,\theta),\ D_{\psi\psi}Q_n(\theta)=n^{-1}\sum_{i=1}^n\rho_{\psi\psi}(W_i,\theta),$$
$$m(W_i,\theta)=\rho_\psi(W_i,\theta),\ \Omega(\pi_1,\pi_2;\gamma_0)=S_\psi V^\dagger((\psi_0,\pi_1),(\psi_0,\pi_2);\gamma_0)S_\psi',$$
$$H(\pi;\gamma_0)=E_{\gamma_0}\rho_{\psi\psi}(W_i,\psi_0,\pi),\ DQ_n(\theta)=n^{-1}\sum_{i=1}^n\rho_\theta(W_i,\theta),$$
$$D^2Q_n(\theta)=n^{-1}\sum_{i=1}^n\rho_{\theta\theta}(W_i,\theta),\ J(\gamma_0)=E_{\gamma_0}\rho_{\theta\theta}^\dagger(W_i,\theta_0),\ \textit{and}\ V(\gamma_0)=V^\dagger(\theta_0,\theta_0;\gamma_0).$$

We start by giving some general results that are useful in the proof of Lemma 11.1. Specifically, Lemma 11.2 is a uniform convergence result for non-stochastic functions, Lemma 11.3 is a uniform LLN, Lemma 11.4 is a stochastic equicontinuity result for empirical processes based on Theorem 3 of Hansen (1996), and Lemma 11.5 is a CLT. All of these results are for strong mixing triangular arrays. The proofs of Lemmas 11.2-11.5 are given below following those of Lemmas 11.1, 9.1, and 3.1.

**Lemma 11.2.** *Let* $\{q_n(\theta):n\geq 1\}$ *be non-stochastic functions on* $\Theta$. *Suppose* (i) $q_n(\theta)\to 0\ \forall\theta\in\Theta$, (ii) $||q_n(\theta_1)-q_n(\theta_2)||\leq C\delta\ \forall\theta_1,\theta_2\in\Theta$ *with* $||\theta_1-\theta_2||\leq\delta$, $\forall n\geq 1$, *for some* $C<\infty$ *and all* $\delta>0$, *and* (iii) $\Theta$ *is compact. Then,* $\sup_{\theta\in\Theta}||q_n(\theta)||\to 0$.

**Lemma 11.3.** *Suppose* (i) *Assumption* S1 *holds,* (ii) *for some function* $M_1(w):\mathcal{W}\to R^+$ *and all* $\delta>0$, $||s(w,\theta_1)-s(w,\theta_2)||\leq M_1(w)\delta$, $\forall\theta_1,\theta_2\in\Theta$ *with* $||\theta_1-\theta_2||\leq\delta$, $\forall w\in\mathcal{W}$, (iii) $E_\gamma\sup_{\theta\in\Theta}||s(W_i,\theta)||^{1+\varepsilon}+E_\gamma M_1(W_i)\leq C\ \forall\gamma\in\Gamma$ *for some* $C<\infty$ *and* $\varepsilon>0$, *and* (iv) $\Theta$ *is compact. Then,* $\sup_{\theta\in\Theta}||n^{-1}\sum_{i=1}^n s(W_i,\theta)-E_{\gamma_0}s(W_i,\theta)||\to_p 0$ *under* $\{\gamma_n\}\in\Gamma(\gamma_0)$ *and* $E_{\gamma_0}s(W_i,\theta)$ *is uniformly continuous on* $\Theta\ \forall\gamma_0\in\Gamma$.

**Comment.** Note that the centering term in Lemma 11.3 is $E_{\gamma_0}s(W_i,\theta)$, rather than $E_{\gamma_n}s(W_i,\theta)$.

**Lemma 11.4.** *Suppose* (i) *Assumption* S1 *holds,* (ii) *for some function* $M_1(w):\mathcal{W}\to R^+$ *and all* $\delta>0$, $||s(w,\theta_1)-s(w,\theta_2)||\leq M_1(w)\delta$, $\forall\theta_1,\theta_2\in\Theta$ *with* $||\theta_1-\theta_2||\leq\delta$, $\forall w\in\mathcal{W}$, *and* (iii) $E_\gamma\sup_{\theta\in\Theta}||s(W_i,\theta)||^q+E_\gamma M_1(W_i)^q\leq C\ \forall\gamma\in\Gamma$ *for some* $C<\infty$ *and* $q$ *as in Assumption* S1. *Then,* $\nu_n s(\theta)=n^{-1/2}\sum_{i=1}^n(s(W_i,\theta)-E_{\gamma_n}s(W_i,\theta))$ *is stochastically equicontinuous over* $\theta\in\Theta$ *under* $\{\gamma_n\}\in\Gamma(\gamma_0)$, *i.e.,* $\forall\varepsilon>0$ *and* $\eta>0$, $\exists\delta>0$ *such that* $\limsup_{n\to\infty}P[\sup_{\theta_1,\theta_2\in\Theta:||\theta_1-\theta_2||<\delta}||\nu_n s(\theta_1)-\nu_n s(\theta_2)||>\eta]<\varepsilon\ \forall\gamma_0\in\Gamma$.

**Lemma 11.5.** *Suppose* (i) *Assumption* S1 *holds,* (ii) $s(w) \in R$ *and* $E_\gamma|s(W_i)|^q \leq C$ $\forall \gamma \in \Gamma$ *for some* $C < \infty$ *and* $q$ *as in Assumption* S1. *Then,* $n^{-1/2}\sum_{i=1}^n (s(W_i) - E_{\gamma_n} s(W_i)) \to_d N(0, V_s(\gamma_0))$ *under* $\{\gamma_n\} \in \Gamma(\gamma_0)$ $\forall \gamma_0 \in \Gamma$, *where* $V_s(\gamma_0) = \sum_{m=-\infty}^{\infty} Cov_{\gamma_0}(s(W_i), s(W_{i+m}))$.

**Proof of Lemma 11.1.** The verification of Assumptions A and B3 are the same for the scalar $\beta$ and vector $\beta$ cases. Assumption A follows from Assumptions S2(i) and S2(ii).

Now we verify Assumption B3. In Assumption B3(i), $Q(\theta; \gamma_0) = E_{\gamma_0}\rho(W_i, \theta)$. Assumption B3(i) follows from Lemma 11.3 with $s(w, \theta) = \rho(w, \theta)$ by Assumptions S1, S2(i), S2(v), and S3(iii). Assumptions B3(ii) and B3(iii) can be verified by Assumption B3*. Assumption B3*(i) holds by Assumptions S2(i) and S3(iii). The remaining parts of Assumption B3* follow from Assumption S2 immediately.

We verify the quadratic expansions in Assumptions C1 and D1 using Lemma 11.5 in Appendix A of AC1-SM, which relies on Assumption Q1 of AC1-SM. Assumptions Q1(i) and Q1(ii) follow from Assumption S2(i). Assumption Q1(iii) follows from Lemma 11.3 with $s(w, \theta) = \rho_{\psi\psi}(w, \theta)$.

To verify Assumption Q1(iv) for $\theta \in \Theta_n(\delta_n)$, we write

$$
\begin{aligned}
& B^{-1}(\beta_n) n^{-1} \sum_{i=1}^n \rho_{\theta\theta}(W_i, \theta) B^{-1}(\beta_n) \\
= {} & B(\beta/\iota(\beta_n)) \left( n^{-1} \sum_{i=1}^n \left( \rho^{\dagger}_{\theta\theta}(W_i, \theta) + \iota^{-1}(\beta)\varepsilon(W_i, \theta) \right) \right) B(\beta/\iota(\beta_n)) \\
= {} & \left( n^{-1} \sum_{i=1}^n \rho^{\dagger}_{\theta\theta}(W_i, \theta) \right) (1 + o(1)) + \left( n^{-1/2} \sum_{i=1}^n \left( \varepsilon(W_i, \theta) - E_{\gamma_n}\varepsilon(W_i, \theta) \right) \right) \times \\
& (n^{1/2}\iota(\beta_n))^{-1}(1 + o(1)) + \left( E_{\gamma_n}\varepsilon(W_i, \theta)/\iota(\beta_n) \right)(1 + o(1)). \qquad (11.1)
\end{aligned}
$$

In (11.1), the first equality follows from (3.4) and the second equality holds because $B(\beta)$ only depends on $\beta$ through $\iota(\beta)$, $||\beta|| \leq ||\beta - \beta_n|| + ||\beta_n|| \leq (1 + \delta_n)||\beta_n||$, and $\delta_n = o(1)$. By (11.1) and the fact that $n^{1/2}||\beta_n|| \to \infty$ for $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, to verify Assumption Q1(iv), it suffices to establish the stochastic equicontinuity of $n^{-1}\sum_{i=1}^n \rho^{\dagger}_{\theta\theta}(W_i, \theta)$ and $n^{-1/2}\sum_{i=1}^n (\varepsilon(W_i, \theta) - E_{\gamma_n}\varepsilon(W_i, \theta))$ over $\theta \in \Theta_n(\delta_n)$ and the equicontinuity of $E_{\gamma_n}\varepsilon(W_i, \theta)/||\beta_n||$ over $\theta \in \Theta_n(\delta_n)$.

When $\beta$ is a scalar, the stochastic equicontinuity of $n^{-1}\sum_{i=1}^n \rho^{\dagger}_{\theta\theta}(W_i, \theta)$ follows from Lemma 11.3 using Assumptions S1, S3(ii), and S3(iii). The stochastic equicontinuity of $n^{-1/2}\sum_{i=1}^{n}(\varepsilon(W_i,\theta)-E_{\gamma_n}\varepsilon(W_i,\theta))$ follows Lemma 11.4 with $s(w,\theta)=\varepsilon(w,\theta)$ using Assumptions S3(ii) and S3(iii).

When $\beta$ is a vector, the stochastic equicontinuity of $n^{-1}\sum_{i=1}^{n}\rho_{\theta\theta}^{\dagger}(W_i,\theta^+)$ and $v_n\varepsilon(w,\theta^+)=n^{-1/2}\sum_{i=1}^{n}(\varepsilon(W_i,\theta^+)-E_{\gamma_n}\varepsilon(W_i,\theta^+))$ over $\theta^+\in\Theta^+$ hold by Lemmas 11.3 and 11.4 using Assumption S3 (vector $\beta$). By Andrews (1994, p. 2252), the stochastic equicontinuity of $v_n\varepsilon(w,\theta^+)$ over $\theta^+\in\Theta^+$ is equivalent to the following: for all sequences of random elements $\{\widehat{\theta}_{1,n}^{+}\in\Theta^+ : n\geq 1\}$ and $\{\widehat{\theta}_{2,n}^{+}\in\Theta^+ : n\geq 1\}$ that satisfy $||\widehat{\theta}_{1,n}^{+}-\widehat{\theta}_{12,n}^{+}||\to_p 0$, we have $||v_n\varepsilon(w,\widehat{\theta}_{1,n}^{+})-v_n\varepsilon(w,\widehat{\theta}_{2,n}^{+})||\to_p 0$. Note that $v_n\varepsilon(w,\theta^+)=v_n\varepsilon(w,\theta)$, where $\theta^+$ is the reparameterization of $\theta$. Hence, to show the stochastic equicontinuity of $v_n\varepsilon(w,\theta)$ over $\theta\in\Theta_n(\delta_n)$, it is sufficient to show that for all sequences of random elements $\{\widehat{\theta}_{1,n}\in\Theta_n(\delta_n) : n\geq 1\}$ and $\{\widehat{\theta}_{2,n}\in\Theta_n(\delta_n) : n\geq 1\}$, $||\widehat{\theta}_{1,n}-\widehat{\theta}_{12,n}||\to_p 0$ implies that $||\widehat{\theta}_{1,n}^{+}-\widehat{\theta}_{12,n}^{+}||\to_p 0$, where $\widehat{\theta}_{i,n}^{+}$ is the reparameterization of $\widehat{\theta}_{i,n}$ for $i=1$ and 2. The convergence related to $||\beta||,\zeta,$ and $\pi$ are straightforward. To show $||\widehat{\omega}_{1,n}-\widehat{\omega}_{2,n}||\to_p 0$, it is sufficient to show that $\widehat{\omega}_n\to_p\omega_0$ for all sequences of random elements $\{\widehat{\theta}_n\in\Theta_n(\delta_n) : n\geq 1\}$ under $\{\gamma_n\}\in\Gamma(\gamma_0,\infty,\omega_0)$ as in Assumption D1. By the definition of $\Theta_n(\delta_n)$, $||\beta_n||^{-1}(\widehat{\beta}_n-\beta_n)=o_p(1)$. This implies that $\widehat{\beta}_n=\beta_n+||\beta_n||o_p(1)$ and $||\widehat{\beta}_n||/||\beta_n||=1+o_p(1)$. Hence,

$$\widehat{\omega}_n=\frac{\widehat{\beta}_n}{||\widehat{\beta}_n||}=\frac{\widehat{\beta}_n-\beta_n}{||\beta_n||}\frac{||\beta_n||}{||\widehat{\beta}_n||}+\frac{\beta_n}{||\beta_n||}\frac{||\beta_n||}{||\widehat{\beta}_n||}\to_p\omega_0. \tag{11.2}$$

This completes the verification of the stochastic equicontinuity of $v_n\varepsilon(w,\theta)$ over $\theta\in\Theta_n(\delta_n)$ when $\beta$ is a vector. The stochastic equicontinuity of $n^{-1}\sum_{i=1}^{n}\rho_{\theta\theta}^{\dagger}(W_i,\theta)$ over $\theta\in\Theta_n(\delta_n)$ holds by the same reparameterization arguments above in the vector $\beta$ case.

It remains to show $\sup_{\theta_1,\theta_2\in\Theta_n(\delta_n)}E_{\gamma_n}(\varepsilon(W_i,\theta_1)-\varepsilon(W_i,\theta_2))=o(||\beta_n||)$. When $\beta$ is a scalar, for any $\theta\in\Theta_n(\delta_n)$,

$$\begin{aligned}
&|\beta_n|^{-1}||E_{\gamma_n}\varepsilon(W_i,\theta)||\\
&=|\beta_n|^{-1}||E_{\gamma_n}\varepsilon(W_i,\theta)-E_{\gamma_n}\varepsilon(W_i,\psi_n,\pi)+E_{\gamma_n}\varepsilon(W_i,\psi_n,\pi)||\\
&\leq C_1|\beta_n|^{-1}|\beta_n|\delta_n+C_2||\pi-\pi_n||\leq(C_1+C_2)\delta_n
\end{aligned} \tag{11.3}$$

for some $C_1,C_2<\infty$ and any constants $\delta_n\to 0$, where the first inequality follows from Assumptions S3(i)-S3(iii) (scalar $\beta$) with $\delta=|\beta_n|\delta_n$ in Assumption S3(ii) and $||\theta-(\psi_n,\pi)||=||\psi-\psi_n||\leq|\beta_n|\delta_n$ by the definition of $\Theta_n(\delta_n)$, and the second inequality holds because $\theta\in\Theta_n(\delta_n)$.

When $\beta$ is a vector, we reparameterize $\theta$ as $\theta^+$. For any $\theta \in \Theta_n(\delta_n)$, we have $(1-\delta_n)||\beta_n|| \leq ||\beta|| \leq (1+\delta_n)||\beta_n||$ and

$$\begin{aligned}||\omega - \omega_n|| &= \left\| \frac{\beta}{||\beta||} - \frac{\beta_n}{||\beta_n||} \right\| = \left\| \frac{\beta - \beta_n}{||\beta_n||} \frac{||\beta_n||}{||\beta||} + \frac{\beta_n}{||\beta_n||} \left( \frac{||\beta_n||}{||\beta||} - 1 \right) \right\| \\ &\leq \delta_n (1-\delta_n)^{-1} + (1-\delta_n)^{-1} - 1 = 2(1-\delta_n)^{-1}\delta_n \leq 4\delta_n, \end{aligned} \tag{11.4}$$

for $n$ large enough that $\delta_n \leq 1/2$, where the first inequality uses the triangle inequality, $||\beta - \beta_n||/||\beta_n|| \leq \delta_n$, and $||\beta_n||/||\beta|| \leq (1-\delta_n)^{-1}$ and the equalities are straightforward.

Let $\theta^{++} = (||\beta_n||, \omega, \zeta_n, \pi)$. For $\theta \in \Theta_n(\delta_n)$ and large $n$,

$$\begin{aligned} & ||\beta_n||^{-1} ||E_{\gamma_n} \varepsilon(W_i, \theta)|| \\ = & ||\beta_n||^{-1} ||E_{\gamma_n} \varepsilon(W_i, \theta^+)|| \\ = & ||\beta_n||^{-1} ||E_{\gamma_n} \varepsilon(W_i, \theta^+) - E_{\gamma_n} \varepsilon(W_i, \theta^{++}) + E_{\gamma_n} \varepsilon(W_i, \theta^{++})|| \\ \leq & C_1 ||\beta_n||^{-1} ||\beta_n|| \delta_n + C_2 (||\pi - \pi_n|| + ||\omega - \omega_n||) \leq (C_1 + 5C_2)\delta_n \end{aligned} \tag{11.5}$$

for some $C_1, C_2 < \infty$, where the first inequality follows from Assumption S3(i)-S3(iii) (vector $\beta$) with $\delta = ||\beta_n||\delta_n$ and $||\theta^+ - \theta^{++}|| \leq ||\beta - \beta_n|| + ||\zeta - \zeta_n|| \leq 2||\beta_n||\delta_n$ by the definition of $\Theta_n(\delta_n)$, and the third inequality holds by $\theta \in \Theta_n(\delta_n)$ and (11.4). This completes the verification of Assumptions C1 and D1 with the stochastic partial derivatives given in Lemma 11.1.

Assumption C2(i) holds with $m(w, \theta) = \rho_\psi(w, \theta)$ by Lemma 11.5(a) of AC1-SM given the verification above of Assumption Q1. Assumptions C2(ii) and C2(iii) follow from Assumptions S2(iii) and S2(iv) given that the true parameter $\theta_0$ lies in the interior of $\Theta$ by Assumption B1(i).

The verifications for Assumptions C3, C4, and C8 below are the same for cases with scalar $\beta$ and vector $\beta$ because $\rho_\psi(W_i, \theta)$ and $\rho_{\psi\psi}(W_i, \theta)$ do not involve re-scaling with $B(\beta)$.

We now verify Assumption C3. To this end, it is sufficient to show that $\nu_n \rho_\psi(\theta) = n^{-1/2} \sum_{i=1}^n (\rho_\psi(W_i, \theta) - E_{\gamma_n} \rho_\psi(W_i, \theta))$ converges weakly to a Gaussian process on $\Theta$ with covariance kernel $S_\psi V^\dagger(\theta_1, \theta_2; \gamma_0) S_\psi'$. The finite-dimensional convergence holds by Lemma 11.5 and the Cramer-Wold device under Assumptions S3(ii) and S3(iii). Note that $\rho_\psi(w, \theta) = S_\psi \rho_\theta^\dagger(w, \theta)$ by the structure of $B(\beta)$. This yields the form of $\Omega(\pi_1, \pi_2; \gamma_0)$ given in Lemma 11.1. The stochastic equicontinuity of $\nu_n \rho_\psi(\theta)$ on $\theta \in \Theta$ follows from Lemma 11.4 with $s(w, \theta) = \rho_\psi(w, \theta)$ under Assumptions S3(ii) and S3(iii). The parameter space $\Theta \subset R^{d_\theta}$ is compact. Hence, the weak convergence of $\upsilon_n \rho_\psi(\theta)$ holds by the Proposition in Andrews (1994, p. 2251).

To show Assumption C4(i) holds with $H(\pi;\gamma_0) = E_{\gamma_0}\rho_{\psi\psi}(W_i,\psi_0,\pi)$, we have

$$\begin{aligned} &\sup_{\pi\in\Pi} \|n^{-1}\sum_{i=1}^n \rho_{\psi\psi}(W_i,\psi_{0,n},\pi) - E_{\gamma_0}\rho_{\psi\psi}(W_i,\psi_0,\pi)\| \\ \leq &\sup_{\theta\in\Theta} \|n^{-1}\sum_{i=1}^n \rho_{\psi\psi}(W_i,\theta) - E_{\gamma_0}\rho_{\psi\psi}(W_i,\theta)\| + \\ &\sup_{\pi\in\Pi} \|E_{\gamma_0}\rho_{\psi\psi}(W_i,\psi_{0,n},\pi) - E_{\gamma_0}\rho_{\psi\psi}(W_i,\psi_0,\pi)\| \end{aligned} \tag{11.6}$$

by the triangle inequality. The first term on the rhs of (11.6) is $o_p(1)$ by Lemma 11.3 with $s(w,\theta) = \rho_{\psi\psi}(w,\theta)$ using Assumptions S1, S2(v), and S3(iii). The second term on the rhs of (11.6) is $o(1)$ because $E_{\gamma_0}\rho_{\psi\psi}(W_i,\theta)$ is continuous in $\psi$ uniformly over $\pi \in \Pi$ by Lemma 11.3. Hence, the rhs of (11.6) is $o_p(1)$, which is the desired result. Assumption C4(ii) holds by Assumptions S3(iii) and S3(iv).

To verify Assumption C8, we have

$$\begin{aligned} &\|\frac{\partial}{\partial\psi'}E_{\gamma_n}\rho_\psi(W_i,\psi_n,\pi_n) - E_{\gamma_0}\rho_{\psi\psi}(W_i,\theta_0)\| = \|E_{\gamma_n}\rho_{\psi\psi}(W_i,\psi_n,\pi_n) - E_{\gamma_0}\rho_{\psi\psi}(W_i,\theta_0)\| \\ \leq &\sup_{\theta\in\Theta}\|E_{\gamma_n}\rho_{\psi\psi}(W_i,\theta) - E_{\gamma_0}\rho_{\psi\psi}(W_i,\theta)\| + \|E_{\gamma_0}\rho_{\psi\psi}(W_i,\theta_n) - E_{\gamma_0}\rho_{\psi\psi}(W_i,\theta_0)\|, \end{aligned} \tag{11.7}$$

where the equality follows from $(\partial/\partial\psi')E_{\gamma_n}\rho_\psi(W_i,\psi_n,\pi_n) = E_{\gamma_n}\rho_{\psi\psi}(W_i,\psi_n,\pi_n)$ using $E_{\gamma_0}\sup_{\theta\in\Theta}\|\rho_{\psi\psi}(W_i,\theta)\| \leq C \ \forall\gamma_0 \in \Gamma$, and the inequality holds by the triangle inequality. The first term in the second line of (11.7) converges to 0 by Lemma 11.2. The conditions for Lemma 11.2 hold by the arguments in the second paragraph of the proof of Lemma 11.3 with $s(w,\theta)$ replaced by $\rho_{\psi\psi}(w,\theta)$. The required conditions are provided in Assumptions S3(ii) and S3(iii). The second term in the second line of (11.7) converges to 0 by the continuity of $E_{\gamma_0}\rho_{\psi\psi}(W_i,\theta)$ in $\theta$, which holds by Lemma 11.3.

To verify Assumption D2, we have

$$J_n = n^{-1}\sum_{i=1}^n \rho^\dagger_{\theta\theta}(W_i,\theta_n) + (n^{1/2}\iota(\beta_n))^{-1}n^{-1/2}\sum_{i=1}^n \varepsilon(W_i,\theta_n) \tag{11.8}$$

by (3.4). When $\beta$ is a scalar, by applying Lemma 11.3 with $s(w,\theta) = \rho^{\dagger}_{\theta\theta}(w,\theta)$ and invoking the continuity of $E_{\gamma_0}\rho^{\dagger}_{\theta\theta}(W_i,\theta)$ in $\theta$, we obtain $n^{-1}\sum_{i=1}^{n}\rho^{\dagger}_{\theta\theta}(W_i,\theta_n) \to_p E_{\gamma_0}\rho^{\dagger}_{\theta\theta}(W_i,\theta_0) = J(\gamma_0)$. Because $n^{1/2}|\beta_n| \to \infty$, the second summand in (11.8) is $o_p(1)$ provided $n^{-1/2}\sum_{i=1}^{n}\varepsilon(W_i,\theta_n) = O_p(1)$. This is verified by applying the triangular array CLT in Lemma 11.5 with $s(w) = \varepsilon(w,\theta_n)$ using $E_{\gamma_n}\varepsilon(W_i,\theta_n) = 0$ by Assumption S3(i). The above results combine to give $J_n \to_p J(\gamma_0)$ as desired. The matrix $J(\gamma_0)$ is positive definite by Assumption S3(iv) and symmetric by the construction of $\rho^{\dagger}_{\theta\theta}(W_i,\theta_0)$ in (3.4).

When $\beta$ is a vector, the verification of Assumption D2 is the same as above by reparameterizing $\theta$ as $\theta^{+}$, replacing Assumption S3 (scalar $\beta$) with Assumption S3 (vector $\beta$), and using the fact that $\theta^{+}_n \to \theta^{+}_0$ under $\{\gamma_n\} \in \Gamma(\gamma_0,\infty,\omega_0)$, where $\theta^{+}_n$ and $\theta^{+}_0$ are the counterparts of $\theta_n$ and $\theta_0$ after reparameterization.

To verify Assumption D3, we have

$$n^{-1/2}B^{-1}(\beta_n)DQ_n(\theta_n) = n^{-1/2}\sum_{i=1}^{n}\rho^{\dagger}_{\theta}(W_i,\theta_n), \tag{11.9}$$

where the equality follows from (3.4). By Assumptions B1(i), S2(iii), and S2(iv), $(\partial/\partial\theta)$ $E_{\gamma_0}\rho(W_i,\theta_0) = 0\ \forall\gamma_0 \in \Gamma$. Under Assumption S3(iii), we have $E_{\gamma_0}\sup_{\theta\in\Theta}||\rho_\theta(W_i,\theta)|| < \infty$ because the parameter space of $\beta$ is bounded. Hence, an exchange of $\partial$ and $E$ yields $E_{\gamma_0}\rho_\theta(W_i,\theta_0) = 0$, which implies that $E_{\gamma_0}\rho^{\dagger}_{\theta}(W_i,\theta_0) = 0$ by (3.4). Because $E_{\gamma_n}\rho^{\dagger}_{\theta}(W_i,\theta_n) = 0$, $n^{-1/2}\sum_{i=1}^{n}\rho^{\dagger}_{\theta}(W_i,\theta_n) = \nu_n\rho^{\dagger}_{\theta}(\theta_n)$.

When $\beta$ is a scalar, $\nu_n\rho^{\dagger}_{\theta}(\theta)$ converges weakly to a Gaussian process with covariance $V^{\dagger}(\theta_1,\theta_2;\gamma_0)$ on $\theta \in \Theta$ by the arguments given in the verification of Assumption C3. Hence, $\nu_n\rho^{\dagger}_{\theta}(\theta_n)$ converges in distribution to a normally distributed random variable with variance $V^{\dagger}(\theta_0,\theta_0;\gamma_0)$. Assumption D3(ii) holds by Assumption S3(v).

When $\beta$ is a vector, the weak convergence above holds by replacing $\theta$ and $\Theta$ with $\theta^{+}$ and $\Theta^{+}$, respectively, using Assumption S3 (vector $\beta$) and the convergence in distribution holds because $\theta^{+}_n \to \theta^{+}_0$ under $\{\gamma_n\} \in \Gamma(\gamma_0,\infty,\omega_0)$. This completes the verification of Assumption D3. □

**Proof of Lemma 9.1.** Assumption S3*(i) and (9.2) imply that $E_{\gamma_0}\varepsilon(W_i,\theta_0) = 0$.

Let $D(x,\theta)$ denote the matrix in the third line of (9.2) so that $\varepsilon(w,\theta)=\rho'(\cdot)D(x,\theta)$. To verify the Lipschitz condition in Assumption S3(i), we have

$$\begin{aligned}
&||\beta_0||^{-1}||E_{\gamma_0}\varepsilon(W_i,\psi_0,\pi)|| = ||\beta_0||^{-1}||E_{\gamma_0}\rho'(W_i,a(X_i,\beta_0)h(X_i,\pi))D(X_i,\theta)|| \\
&= ||E_{\gamma_0}\rho''(W_i,a(X_i,\beta_0)h(X_i,\widetilde{\pi}))\frac{a(X_i,\beta_0)}{||\beta_0||}h_\pi(X_i,\widetilde{\pi})'(\pi-\pi_0)D(X_i,\theta)|| \\
&\leq E_{\gamma_0}|\rho''(W_i,a(X_i,\beta_0)h(X_i,\widetilde{\pi}))|\cdot||a_\beta(X_i,\widetilde{\beta})||\cdot||h_\pi(X_i,\widetilde{\pi})||\cdot||D(X_i,\theta)||\cdot||\pi-\pi_0|| \\
&\leq C||\pi-\pi_0||, \qquad (11.10)
\end{aligned}$$

where the first equality holds by (9.2), the second equality follows from a mean-value expansion of $\rho'(W_i,a(X_i,\beta_0)h(X_i,\pi))$ in $\pi$ around $\pi_0$ with $\widetilde{\pi}$ between $\pi$ and $\pi_0$ and uses Assumption S3*(i), the first inequality follows from a mean-value expansion of $a(X_i,\beta_0)$ in $\beta_0$ around 0 with $\widetilde{\beta}$ between $\beta_0$ and 0, and the second inequality follows from Assumption S3*(ii) and $|a(X_i,\beta)|/||\beta||\leq||a_\beta(X_i,\widetilde{\beta})||$ with $\widetilde{\beta}$ between $\beta$ and 0 by a mean-value expansion.

This completes the proof when $\beta$ is a scalar.

When $\beta$ is a vector, $||E_{\gamma_0}\varepsilon(W_i,\theta^+)||\leq||E_{\gamma_0}\varepsilon(W_i,\theta^+)-E_{\gamma_0}\varepsilon(W_i,||\beta_0||,\omega_0,\pi)||+||E_{\gamma_0}\varepsilon(W_i,||\beta_0||,\omega_0,\pi)||$, where $\theta^+=(||\beta_0||,\omega,\pi)$. By (11.10), it is sufficient to show $||E_{\gamma_0}\varepsilon(W_i,||\beta_0||,\omega,\pi)-E_{\gamma_0}\varepsilon(W_i,||\beta_0||,\omega_0,\pi)||\leq C||\beta_0||(||\omega-\omega_0||+||\pi-\pi_0||)$ for some $C<\infty$. According to (9.2), $\varepsilon(\cdot)=\rho'(\cdot)D(\cdot)$. For notational simplicity, we let $\rho'(\omega)$ and $D(\omega)$ be defined such that $\varepsilon(W_i,||\beta_0||,\omega,\pi)=\rho'(\omega)D(\omega)$. By the triangle inequality,

$$\begin{aligned}
&||E_{\gamma_0}\rho'(\omega)D(\omega)-E_{\gamma_0}\rho'(\omega_0)D(\omega_0)|| \\
&\leq||E_{\gamma_0}\left(\rho'(\omega)-\rho'(\omega_0)\right)D(\omega)||+||E_{\gamma_0}\rho'(\omega_0)(D(\omega)-D(\omega_0)||. \qquad (11.11)
\end{aligned}$$

Because $\rho'(\cdot)$ does not involve $\iota(\beta)$, the reparameterization inside of $\rho'(\cdot)$ simply replaces all $\beta$ with $||\beta||\omega$. Hence, any partial derivative of $\rho'(\omega)$ wrt $\omega$ is equivalent to the partial derivative wrt $\beta$ in the original parametrization multiplied by $||\beta||$.

The first summand on the right-hand side of (11.11) satisfies

$$\begin{aligned}
&||E_{\gamma_0}(\rho'(\omega_0)-\rho'(\omega))D(\omega)|| \\
&\leq E_{\gamma_0}||\rho''(\cdot)a_\beta(\cdot)h(\cdot)||\cdot||D(\omega)||\cdot||\beta_0||\cdot||\omega-\omega_0|| \\
&\leq C||\beta_0||\cdot||\omega-\omega_0||, \qquad (11.12)
\end{aligned}$$

where the first inequality holds by a mean-value expansion wrt $\omega$ and uses $\beta = ||\beta||\omega$ and the second inequality holds by Assumption S3*(ii). The arguments of the functions are suppressed for brevity.

To bound the second summand on the right-hand side of (11.11), note that $\rho'(\omega_0)$ differs from $\rho'(W_i, a(X_i, \beta_0)h(X_i, \pi_0))$ in Assumption S3*(i) by having $\pi$ in the place of $\pi_0$. Using Assumption S3*(i), we have

$$\begin{aligned}
&||E_{\gamma_0}\rho'(\omega_0)(D(\omega) - D(\omega_0))|| \\
\leq\ & E_{\gamma_0}\left\|\rho''(\cdot)a(\cdot)h_\pi(\cdot)(D(\omega) - D(\omega_0))\right\| \cdot ||\pi - \pi_0|| \\
=\ & E_{\gamma_0}\left\|\rho''(\cdot)a_\beta(\cdot)h_\pi(\cdot)(D(\omega) - D(\omega_0))\right\| \cdot ||\beta_0|| \cdot ||\pi - \pi_0|| \\
\leq\ & C||\beta_0|| \cdot ||\pi - \pi_0||, \qquad (11.13)
\end{aligned}$$

where the first inequality follows from a mean-value expansion wrt $\pi$ around $\pi_0$, the equality follows from a mean-value expansion wrt $\beta$ around 0 and uses $a(x, 0) = 0$, and the second inequality follows from the moment conditions in Assumption S3*(ii). The desired result follows from (11.11)-(11.13). □

**Proof of Lemma 3.1.** We verify Assumption C6 for the sample average estimator using Assumption C6** and Lemma 4.1 of AC1. Because $\beta$ is a scalar, it remains to show Assumption C6**(ii). By Lemma 11.1,

$$\Omega(\pi_1, \pi_2; \gamma_0) = \sum_{m=-\infty}^{\infty} Cov_{\gamma_0}(\rho_\psi(W_i, \psi_0, \pi_1), \rho_\psi(W_{i+m}, \psi_0, \pi_2)). \qquad (11.14)$$

This implies that the covariance matrix $\Omega_G(\pi_1, \pi_2; \gamma_0)$ in Assumption C6**(ii) takes the form $\Omega_G(\pi_1, \pi_2; \gamma_0)$ in Assumption C6†(ii). Hence, Assumption C6**(ii) is implied by Assumption C6†(ii). □

**Proof of Lemma 11.2.** For any given $\varepsilon > 0$, let $\delta^* = \min\{\delta, \frac{\varepsilon}{2C}\}$. Using the compactness of $\Theta$, let $\{B(\theta_j, \delta^*) : j = 1, ..., J\}$ be a finite cover of $\Theta$, where $B(\theta_j, \delta^*)$ denote a closed ball in $\Theta$ of radius $\delta^* \geq 0$ centered at $\theta_j$. Because $q_n(\theta)$ converges to 0 $\forall \theta \in \Theta$, there exists $N_j$ such that $||q_n(\theta_j)|| \leq \varepsilon/2$ for any $n \geq N_j$, for $j = 1, ..., J$. Let

$N = \max_{j \leq J} N_j$. Then, $\max_{j \leq J} ||q_n(\theta_j)|| \leq \varepsilon/2$ for any $n \geq N$. For any $n \geq N$,

$$\sup_{\theta \in \Theta} ||q_n(\theta)|| \leq \max_{j \leq J} \left( \sup_{\theta' \in B(\theta_j, \delta^*)} ||q_n(\theta') - q_n(\theta_j)|| + ||q_n(\theta_j)|| \right)$$
$$\leq \sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta^*)} ||q_n(\theta') - q_n(\theta)|| + \max_{j \leq J} ||q_n(\theta_j)|| \leq C\delta^* + \varepsilon/2 \leq \varepsilon, \quad (11.15)$$

where the first inequality uses the property of the finite cover and the triangle inequality, the second inequality is straightforward, the third inequality uses condition (ii) of Lemma 11.2 and $\delta^* \leq \delta$, and the fourth inequality follows from $\delta^* \leq \varepsilon/(2C)$. $\square$

**Proof of Lemma 11.3.** First, we establish the result of the lemma with $E_{\gamma_n} s(W_i, \theta)$ in place of $E_{\gamma_0} s(W_i, \theta)$. We use the uniform LLN given in Theorem 4 of Andrews (1992) employing Assumption TSE-1B with $q_t(z, \theta) = s(w, \theta)$. Now we verify Assumptions TSE-1B, DM, BD, and P-WLLN of Andrews (1992). Assumption TSE-1B(a) holds because $s(w, \theta)$ is continuous in $\theta$ and $\Theta$ is compact. Assumption TSE-1B(b) holds because $\{W_i : i \leq n\}$ is strictly stationary and $E_{\gamma_n} 1(W_i \in A) \to E_{\gamma_0} 1(W_i \in A)$ for all measurable sets $A \subset \mathcal{W}$ under $\{\gamma_n\} \in \Gamma(\gamma_0)$ by $\gamma_n \to \gamma_0$ and the weak convergence of $W_i$ under $\gamma_n$ to $W_i$ under $\gamma_0$, which holds by the definition of the metric on $\Gamma$, see Section 2.1 of the paper. Assumption DM holds by condition (iii) of Lemma 11.3. Assumption BD holds because $\Theta$ is compact. Assumption P-WLLN holds, i.e., $n^{-1} \sum_{i=1}^n s(W_i, \theta) - E_{\gamma_n} s(W_i, \theta) \to_p 0 \ \forall \theta \in \Theta$ under $\{\gamma_n\} \in \Gamma(\gamma_0)$, by the WLLN for dependent triangular arrays of strong mixing random variables in Example 4 of Andrews (1988) given that $\sup_{\gamma \in \Gamma} E_\gamma ||s(W_i, \theta)||^{1+\delta} < \infty$ for some $\delta > 0$. Theorem 4 of Andrews (1992) gives $\sup_{\theta \in \Theta} ||n^{-1} \sum_{i=1}^n s(W_i, \theta) - E_{\gamma_n} s(W_i, \theta)|| \to 0$ under $\{\gamma_n\} \in \Gamma(\gamma_0)$. Note that the same proof holds whether $\{W_i : i \geq 1\}$ are strong mixing or i.i.d. in Assumption S1.

To obtain the desired result, it remains to show $\sup_{\theta \in \Theta} ||E_{\gamma_n} s(W_i, \theta) - E_{\gamma_0} s(W_i, \theta)|| \to 0$ under $\{\gamma_n\} \in \Gamma(\gamma_0)$. The pointwise convergence holds for any $\theta \in \Theta$ by (i) the weak convergence induced by $\gamma_n \to \gamma_0$ and the definition of the metric on $\Gamma$ and (ii) $E_\gamma \sup_{\theta \in \Theta} ||s(w, \theta)||^{1+\delta} \leq C \ \forall \gamma \in \Gamma$ for some $\delta > 0$. Because $\Theta$ is compact and pointwise convergence holds, we apply Lemma 11.2 with $q_n(\theta) = E_{\gamma_n} s(W_i, \theta) - E_{\gamma_0} s(W_i, \theta)$.

Condition (ii) of Lemma 11.2 holds because for any $\theta_1, \theta_2 \in \Theta$ with $||\theta_1 - \theta_2|| \leq \delta$,

$$\begin{aligned} &||q_n(\theta_1) - q_n(\theta_2)|| = ||E_{\gamma_n}\left(s(W_i, \theta_1) - s(W_i, \theta_2)\right) - E_{\gamma_0}\left(s(W_i, \theta_1) - s(W_i, \theta_2)\right)|| \\ &\leq E_{\gamma_n}||s(W_i, \theta_1) - s(W_i, \theta_2)|| + E_{\gamma_0}||s(W_i, \theta_1) - s(W_i, \theta_2)|| \\ &\leq (E_{\gamma_n} M_1(W_i) + E_{\gamma_0} M_1(W_i))\delta \leq C\delta, \end{aligned} \tag{11.16}$$

where the first inequality follows from the triangle inequality and Jensen's inequality, the second inequality holds by condition (ii) of Lemma 11.3, and the third inequality holds by condition (iii) of Lemma 11.3.

The uniform continuity of $E_{\gamma_0} s(W_i, \theta)$ on $\Theta$ holds by the dominated convergence theorem and the compactness of $\Theta$. This completes the proof. □

**Proof of Lemma 11.4.** For the case that Assumption S1 holds with $\{W_i : i \geq 1\}$ being strong mixing, we show the stochastic equicontinuity (SE) of the empirical process $v_n s(\theta)$ using Theorem 3 of Hansen (1996), which is suitable for strong mixing arrays. When $s(w, \theta)$ is a vector, the SE of $\nu_n s(\theta)$ is implied by the SE of each entry of $\nu_n s(\theta)$. Hence, without loss of generality, we assume $s(w, \theta) \in R$ as in Hansen (1996). We now verify (11)-(13) in Assumption 4 of Hansen (1996). The condition in (11) holds provided $\alpha_m \leq Cm^{-A}$ for some $A > (1/p - 1/q)^{-1}$ and $d_\theta < p < q$. This is implied by Assumption S1. Conditions (12) and (13) hold because $E_\gamma \sup_{\theta \in \Theta} ||s(W_i, \theta)||^q \leq C$ and $E_\gamma M_1(W_i)^q \leq C$ $\forall \gamma \in \Gamma$ and $\{W_i : i \geq 1\}$ is strictly stationary. Applying Theorem 3 of Hansen (1996) with $a = d_\theta$ and $\lambda = 1$ yields: for each $\varepsilon > 0$, there exists a $\delta_1 > 0$ such that

$$\limsup_{n\to\infty} \left\| \sup_{\rho(\theta_1,\theta_2)<\delta_1} |\nu_n s(\theta_1) - \nu_n s(\theta_2)| \right\|_p < \varepsilon, \tag{11.17}$$

where $\rho(\theta_1, \theta_2) = \limsup_{n\to\infty}(E_{\gamma_n}|s(W_i, \theta_1) - s(W_i, \theta_2)|^q)^{1/q}$ and $|| \cdot ||_p$ is the $L^p$-norm for some $d_\theta < p < q$. By conditions (ii) and (iii) of Lemma 11.4, for each $\delta_1 > 0$, there exists $\delta > 0$, such that $||\theta_1 - \theta_2|| < \delta$ implies that $\rho(\theta_1, \theta_2) < \delta_1$. This, (11.17), and Markov's inequality yield the SE of $\nu_n s(\theta)$ over $\theta \in \Theta$.

For the case that Assumption S1 holds with $\{W_i : i \geq 1\}$ being i.i.d., the stochastic equicontinuity (SE) of the empirical process $v_n s(\theta)$ holds by Theorems 1 and 2 of Andrews (1994) using the type II class. For this result, the envelope function and the Lipschitz function must have $q = 2 + \delta$ moments finite, which holds by Assumption S1 and condition (iii) of the Lemma. □

**Proof of Lemma 11.5.** First, we consider the case in which $\{W_i : i \geq 1\}$ are strong mixing. We show that $V_s(\gamma_0)$ exists and

$$\lim_{n\to\infty} Var_{\gamma_n}(n^{-1/2}\sum_{i=1}^{n} s(W_i)) = V_s(\gamma_0) \tag{11.18}$$

under $\{\gamma_n\} \in \Gamma(\gamma_0)$. By change of variables, we have

$$\begin{aligned} & Var_{\gamma_n}(n^{-1/2}\sum_{i=1}^{n} s(W_i)) \\ = & \sum_{m=-n+1}^{n-1} Cov_{\gamma_n}(s(W_i), s(W_{i+m})) - \sum_{m=-n+1}^{n-1} \frac{|m|}{n} Cov_{\gamma_n}(s(W_i), s(W_{i+m})). \end{aligned} \tag{11.19}$$

By a standard strong mixing inequality, e.g., see Davidson (1994, p. 212), and Assumption S1,

$$|Cov_{\gamma}(s(W_i), s(W_{i+m}))| \leq C_1 \alpha_m^{1-2/q} \leq C_2 m^{-A(1-2/q)}, \text{ where } A(1-2/q) > d_\theta \frac{q-2}{q-d_\theta} \geq 2 \tag{11.20}$$

using $d_\theta \geq 2$, for some $C_1, C_2 < \infty$ $\forall \gamma \in \Gamma$. Hence, $V_s(\gamma_0)$ exists and the second term on the rhs of (11.19) converges to 0.

It remains to show that the first term on the rhs of (11.19) converges to $V_s(\gamma_0)$. Because the metric on $\Gamma$ induces weak convergence under $\gamma_n \to \gamma_0$ and $E_\gamma|s(W_i)|^{2+\delta} \leq C$ $\forall \gamma \in \Gamma$ for some $\delta > 0$, we have

$$Cov_{\gamma_n}(s(W_i), s(W_{i+m})) \to Cov_{\gamma_0}(s(W_i), s(W_{i+m})) \tag{11.21}$$

under $\gamma_n \in \Gamma(\gamma_0)$ for any $m \in R$ (e.g., see Theorem 2.20 and Example 2.21 of van der Vaart (1998)). By the DCT, (11.20), and (11.21), we have

$$\begin{aligned} \lim_{n\to\infty} \sum_{m=-n+1}^{n-1} Cov_{\gamma_n}(s(W_i), s(W_{i+m})) &= \lim_{n\to\infty} \sum_{m=-\infty}^{\infty} 1(|m| < n-1) Cov_{\gamma_n}(s(W_i), s(W_{i+m})) \\ &\to V_s(\gamma_0). \end{aligned} \tag{11.22}$$

This and (11.19) yield (11.18).

When $V_s(\gamma_0) = 0$, we have $\lim_{n\to\infty} Var_{\gamma_n}(n^{-1/2}\sum_{i=1}^{n} s(W_i)) = 0$, which implies that $n^{-1/2}\sum_{i=1}^{n}(s(W_i) - E_{\gamma_n}s(W_i)) \to_d N(0, V_s(\gamma_0)) = 0$. When $V_s(\gamma_0) > 0$, we assume

$Var_{\gamma_n}(n^{-1/2}\sum_{i=1}^n s(W_i)) > 0\ \forall n \geq 1$ without loss of generality. To show the triangular array CLT in Lemma 11.5, we apply Corollary 1 of de Jong (1997) with $\beta = \gamma = 0$, $c_{ni} = n^{-1/2}(n^{-1/2}||\sum_{i=1}^n \Delta_n s(W_i)||_2)^{-1}$, and $X_{ni} = n^{-1/2}\Delta_n s(W_i)(||n^{-1/2}\sum_{i=1}^n \Delta_n s(W_i)||_2)^{-1}$, where $\Delta_n s(W_i) = s(W_i) - E_{\gamma_n} s(W_i)$. Now we verify conditions (a)-(c) of Assumption 2 of de Jong (1997). Condition (a) holds automatically. Condition (b) holds because $c_{ni} > 0$ and $E_{\gamma_n}|X_{ni}/c_{ni}|^q = E_{\gamma_n}|\Delta_n s(W_i)|^q \leq C\ \forall \gamma_n \in \Gamma$ for some $C < \infty$. Condition (c) holds by taking $V_{ni} = X_{ni}$, $d_{ni} = 0$, and using Assumption S1 because $\alpha_m \leq Cm^{-A}$ and $A > q/(q-2)$. By Corollary 1 of de Jong (1997), we have $X_{ni} \to_d N(0,1)$. This and (11.18) lead to the desired result.

When Assumption S1 holds with $\{W_i : i \geq 1\}$ being i.i.d. under $\gamma_0 \in \Gamma$, a standard triangular array CLT gives the desired result because $2+\delta$ moments of $s(W_i)$ are finite and uniformly bounded over $\gamma_0 \in \Gamma$ by Assumption S1. $\square$

# 12. Supplemental Appendix B: Miscellaneous Results

This Appendix provides (i) the asymptotic size results for the robust QLR CS's, (ii) a sophisticated method for choosing $\kappa$ for type 2 robust CS's, (iii) statements of Assumptions V1 and V2, which concern the estimator of the variance matrix of $\widehat{\theta}_n$, and (iv) an extension of the sufficient conditions for Assumption S3*(i) given in Section 9.1 in the Appendix of the paper for $\rho(w,\theta)$ functions of the form $\rho^*(w, a(x,\beta)h(x,\pi),\zeta)$. The extension is to the case where a parameter $\zeta$ appears.

## 12.1. Asymptotic Size of Robust QLR CS's

Here, we show that the LF and type 2 robust QLR CS's defined in the text of the paper have correct asymptotic size.

For the null-imposed (NI) critical values, we use the following notation: $H(v) = \{h = (b,\gamma_0) \in H : ||b|| < \infty, r(\theta_0) = v\}$, $V_r = \{v_0 : r(\theta_0) = v_0$ for some $h = (b,\gamma_0) \in H\}$, and the NI-LF critical value is $c_{QLR,1-\alpha}^{LF}(v) = \max\{\sup_{h \in H(v)} c_{QLR,1-\alpha}(h), \chi^2_{d_r,1-\alpha}\}$.

The asymptotic size results for the LF QLR CS's rely on the following df continuity conditions, which are not restrictive in most examples.

**Assumption LF.** (i) The df of $QLR(h)$ is continuous at $c_{QLR,1-\alpha}(h)\ \forall h \in H$.

(ii) If $c^{LF}_{QLR,1-\alpha} > \chi^2_{d_r,1-\alpha}$, $c^{LF}_{QLR,1-\alpha}$ is attained at some $h_{\max} \in H$.

**Assumption NI-LF.** (i) The df of $QLR(h)$ is continuous at $c_{QLR,1-\alpha}(h)$ $\forall h \in H(v)$, $\forall v \in V_r$.

(ii) For some $v \in V_r$, $c^{LF}_{QLR,1-\alpha}(v) = \chi^2_{d_r,1-\alpha}$ or $c^{LF}_{QLR,1-\alpha}(v)$ is attained at some $h_{\max} \in H$.

For $h \in H$, define

$$\widehat{c}_{QLR,1-\alpha}(h) \tag{12.1}$$
$$= \begin{cases} c^{LF}_{QLR,1-\alpha} + \Delta_1 & \text{if } A(h) \leq \kappa \\ \chi^2_{d_r,1-\alpha} + \Delta_2 + [c^{LF}_{QLR,1-\alpha} + \Delta_1 - \chi^2_{d_r,1-\alpha} - \Delta_2] \cdot s(A(h) - \kappa) & \text{if } A(h) > \kappa. \end{cases}$$

Note that $\widehat{c}_{QLR,1-\alpha}(h)$ equals $\widehat{c}_{QLR,1-\alpha,n}$ with $A(h)$ in place of $A_n$. The asymptotic distribution of $\widehat{c}_{QLR,1-\alpha,n}$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $||b|| < \infty$ is the distribution of $\widehat{c}_{QLR,1-\alpha}(h)$.

Define $\widehat{c}_{QLR,1-\alpha}(h, v)$ analogously to $\widehat{c}_{QLR,1-\alpha}(h)$, but with $c^{LF}_{QLR,1-\alpha}$, $\Delta_1$, and $\Delta_2$ replaced by $c^{LF}_{QLR,1-\alpha}(v)$, $\Delta_1(v)$, and $\Delta_2(v)$, respectively, for $v \in V_r$. The asymptotic distribution of $\widehat{c}_{QLR,1-\alpha,n}(v)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $||b|| < \infty$ is the distribution of $\widehat{c}_{QLR,1-\alpha}(h, v)$.

The asymptotic size results for the type 2 robust QLR CS's rely on the following df continuity conditions, which are not restrictive in most examples.

**Assumption Rob2.** (i) $P(QLR(h) = \widehat{c}_{QLR,1-\alpha}(h)) = 0$ $\forall h \in H$.

(ii) If $\Delta_2 > 0$, $NRP(\Delta_1, \Delta_2; h^*) = \alpha$ for some point $h^* \in H$.

**Assumption NI-Rob2.** (i) $P(QLR(h) = \widehat{c}_{QLR,1-\alpha}(h, v)) = 0$ $\forall h \in H(v)$, $\forall v \in V_r$.

(ii) For some $v \in V_r$, $\Delta_2(v) = 0$ or $NRP(\Delta_1(v), \Delta_2(v); h^*) = \alpha$ for some point $h^* \in H(v)$.

The correct asymptotic size properties of LF and robust type 2 QLR CS's are established in the following Theorem.

**Theorem 12.1.** *Suppose Assumptions* S1-S4, B1, B2, C7, RQ1-RQ3, *and* RQ4(i) *hold. Then, the nominal* $1-\alpha$ *robust* QLR *CS has* $AsySz = 1-\alpha$ *when based on the following critical values*: (a) LF, (b) NI-LF, (c) *type* 2 *robust, and* (d) *type* 2 NI *robust, provided the following additional Assumptions hold, respectively*: (a) LF, (b) NI-LF, (c) C6, Rob2, V1, *and* V2, *and* (d) C6, NI-Rob2, V1, *and* V2.

**Comments. 1.** Plug-in versions of the robust QLR CS's considered in Theorem 12.1 also have asymptotically correct size under continuity assumptions on $c_{QLR,1-\alpha}(h)$ that typically are not restrictive. For brevity, we do not provide formal results here.

**2.** If part (ii) of Assumption LF, NI-LF, Rob2, or NI-Rob2 does not hold, then the corresponding part of Theorem 12.1 still holds, but with $AsySz \geq 1 - \alpha$.
**3.** The proof of Theorem 12.1 is as follows. Theorem 12.1 holds by Theorem 5.1(b) of AC1, plus the proof given in Supplemental Appendix A above that Assumptions B1, B2, and S1-S4 imply Assumptions A, B3, C1-C5, C8, and D1-D3 of AC1. The reason is that the results of the Theorem 12.1 and Theorem 5.1(b) of AC1 are the same, just the assumptions differ.

## 12.2. Choice of $\kappa$ for Type 2 Robust Confidence Sets

For type 2 robust CS's, a sophisticated method for choosing $\kappa$ is to minimize the average asymptotic FCP of the robust CS at a chosen set of points.[32] Of interest is a robust CS for $r(\theta)$. Let $\mathcal{K}$ denote the set of $\kappa$ values from which one selects. First, for given $h \in H$, one chooses a null value $v_{H_0}(h)$ that differs from the true value $v_0 = r(\theta_0)$ (where $h = (b, \gamma_0)$ and $\gamma_0 = (\theta_0, \phi_0)$). The null value $v_{H_0}(h)$ is selected such that the robust CS based on a reasonable choice of $\kappa$, such as $\kappa = 1.5$ or 2, has a FCP that is in a range of interest, such as close to 0.50.[33,34] Second, one computes the FCP of the value $v_{H_0}(h)$ for each robust CS with $\kappa \in \mathcal{K}$. Third, one repeats steps one and two for each $h \in \mathcal{H}$, where $\mathcal{H}$ is a representative subset of $H$.[35] The optimal choice of $\kappa$ is the value that minimizes over $\mathcal{K}$ the average FCP at $v_{H_0}(h)$ over $h \in \mathcal{H}$.

## 12.3. Assumptions V1 and V2

Here we state Assumptions V1 and V2, which concern estimators of the asymptotic variance matrix of $\widehat{\theta}_n$. These assumptions are used with the standard $t$ tests and CS's, as well as with the robust $t$ and QLR CS's, which employ variance matrix estimators in the

[32] For $t$ and Wald CS's, asymptotic FCP's follow from the results in this paper, AC1, and/or Andrews and Cheng (2008). For QLR CI's, asymptotic FCP results only cover restrictions involving $\pi$, see Comment 5 to Theorem 4.2 of AC1. For other restrictions, one can use a large finite sample size when determining $\kappa$.

[33] For reasonable choices, the value of $\kappa$ used to obtain $v_{H_0}(h)$ typically has very little effect on the final comparison across different values of $\kappa$. For example, this is true in the binary choice and STAR models considered here, and in the ARMA(1, 1) model considered in AC1.

[34] When $b$ is close to 0, the FCP may be larger than 0.50 for all admissible $v$ due to weak identification. In such cases, $v_{H_0}(h)$ is taken to be the admissible value that minimizes the FCP for the selected value of $\kappa$ that is being used to obtain $v_{H_0}(h)$.

[35] When $r(\theta) = \pi$, we do not include $h$ values in $\mathcal{H}$ for which $b = 0$ because when $b = 0$ there is no information about $\pi$ and it is not necessarily desirable to have a small FCP.

identification category selection procedure. These assumptions are not very restrictive.

Assumption V1 has two forms depending on whether $\beta$ is a scalar or a vector.

**Assumption V1 (scalar $\beta$).** (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 $\Theta$ that satisfy $\sup_{\theta \in \Theta} ||\widehat{J}_n(\theta) - J(\theta;\gamma_0)|| \rightarrow_p 0$ and $\sup_{\theta \in \Theta} ||\widehat{V}_n(\theta) - V(\theta;\gamma_0)|| \rightarrow_p 0$ 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 $\theta$ on $\Theta$ $\forall \gamma_0 \in \Gamma$ 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 $\beta_0 = 0$.

**Assumption V1 (vector $\beta$).** (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 $\Theta^+$ that satisfy $\sup_{\theta^+ \in \Theta^+} ||\widehat{J}_n(\theta^+) - J(\theta^+;\gamma_0)|| \rightarrow_p 0$ and $\sup_{\theta^+ \in \Theta^+} ||\widehat{V}_n(\theta^+) - V(\theta^+;\gamma_0)|| \rightarrow_p 0$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$.[36]
(ii) $J(\theta^+;\gamma_0)$ and $V(\theta^+;\gamma_0)$ are continuous in $\theta^+$ on $\Theta^+$ $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.
(iii) $\lambda_{\min}(\Sigma(\pi,\omega;\gamma_0)) > 0$ and $\lambda_{\max}(\Sigma(\pi,\omega;\gamma_0)) < \infty$ $\forall \pi \in \Pi$, $\forall \omega \in R^{d_\beta}$ with $||\omega|| = 1$, $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.
(iv) $P(\tau_\beta(\pi^*(\gamma_0, b);\gamma_0, b) = 0) = 0$ $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$ and $\forall b$ with $||b|| < \infty$.[37]

**Assumption V2.** Under $\Gamma(0, \infty, \omega_0)$, $\widehat{J}_n \rightarrow_p J(\gamma_0)$ and $\widehat{V}_n \rightarrow_p V(\gamma_0)$.

## 12.4. Adjustment for $\zeta$

Here we provide sufficient conditions for Assumption S3*(i) when $\rho(w,\theta) = \rho^*(w, a(x,\beta)h(x,\pi), \zeta)$, as in (3.2), and a parameter $\zeta$ appears. (Section 9.1 in the Appendix of the paper provides analogous results when no parameter $\zeta$ appears.) For simplicity, we assume $a(x,\beta)$ and $h(x,\pi)$ are both scalars. Let $\rho'(\cdot)$ and $\rho''(\cdot)$ denote the first and second order partial derivatives of $\rho^*(w, a(x,\beta)h(x,\pi), \zeta)$ wrt $a(x,\beta)h(x,\pi)$. Let $\rho_\zeta(\cdot)$ and $\rho_{\zeta\zeta}(\cdot)$ denote the first and second order partial derivatives of $\rho^*(w, a(x,\beta)h(x,\pi), \zeta)$ wrt $\zeta$. Let $\rho_{12}(\cdot) \in R^{d_\zeta}$ denote the partial derivative of $\rho'(\cdot)$ wrt $\zeta$. The partial derivatives

[36] The functions $J(\theta^+;\gamma_0)$ and $V(\theta^+;\gamma_0)$ do not depend on $\omega_0$, only $\gamma_0$.

[37] Assumption V1 (vector $\beta$) differs from Assumption V1 (scalar $\beta$) because in the vector $\beta$ case Assumption V1(ii) (scalar $\beta$) (i.e., continuity in $\theta$) often fails, but Assumption V1(ii) (vector $\beta$) (i.e., continuity in $\theta^+$) holds.

in (9.1) are the same when $\zeta$ appears in $\rho(w,\theta)$. The partial derivatives wrt $\zeta$ are

$$\begin{aligned}
\rho_{\zeta}(w,\theta) &= \rho_{\zeta}(\cdot) \in R^{d_{\zeta}\times 1},\ \rho_{\zeta\zeta}(w,\theta) = \rho_{\zeta\zeta}(\cdot) \in R^{d_{\zeta}\times d_{\zeta}}, \\
\rho_{\beta\zeta}(w,\theta) &= a_{\beta}(x,\beta)h(x,\pi)\rho_{12}(\cdot) \in R^{d_{\beta}\times d_{\zeta}}, \\
\rho_{\pi\zeta}(w,\theta) &= a(x,\beta)h_{\pi}(x,\pi)\rho_{12}(\cdot) \in R^{d_{\pi}\times d_{\zeta}}.
\end{aligned} \tag{12.2}$$

In this case, we define

$$\begin{aligned}
\rho_{\theta}^{\dagger}(w,\theta) &= \rho'(\cdot)a^{\dagger}(x,\theta) + \overline{\rho}_{\zeta}(\cdot),\ \rho_{\theta\theta}^{\dagger}(w,\theta) = \rho''(\cdot)a^{\dagger}(x,\theta)a^{\dagger}(x,\theta)' + \overline{\rho}_{\zeta\zeta}(\cdot),\ \text{where} \\
a^{\dagger}(x,\theta) &= (a_{\beta}(x,\beta)'h(x,\pi), 0_{d_{\zeta}}, \frac{a(x,\beta)}{\iota(\beta)}h_{\pi}(x,\pi)')', \\
\overline{\rho}_{\zeta}(\cdot) &= (0_{d_{\beta}}, \rho_{\zeta}(\cdot), 0_{d_{\pi}})', \\
\overline{\rho}_{\zeta\zeta}(\cdot) &= \begin{bmatrix} 0_{d_{\beta}\times d_{\beta}} & a_{\beta}(x,\beta)h(x,\pi)\rho_{12}(\cdot) & 0_{d_{\beta}\times d_{\pi}} \\ (a_{\beta}(x,\beta)h(x,\pi)\rho_{12}(\cdot))' & \rho_{\zeta\zeta}(\cdot) & \left(\frac{a(x,\beta)}{\iota(\beta)}h_{\pi}(x,\pi)\rho_{12}(\cdot)\right)' \\ 0_{d_{\pi}\times d_{\beta}} & \frac{a(x,\beta)}{\iota(\beta)}h_{\pi}(x,\pi)\rho_{12}(\cdot) & 0_{d_{\pi}\times d_{\pi}} \end{bmatrix}, \\
\varepsilon(w,\theta) &= \rho'(\cdot) \begin{bmatrix} a_{\beta\beta}(x,\beta)h(x,\pi) & 0_{d_{\beta}\times d_{\zeta}} & a_{\beta}(x,\beta)h_{\pi}(x,\pi)' \\ 0_{d_{\zeta}\times d_{\beta}} & 0_{d_{\zeta}\times d_{\zeta}} & 0_{d_{\zeta}\times d_{\pi}} \\ h_{\pi}(x,\pi)a_{\beta}(x,\beta)' & 0_{d_{\pi}\times d_{\zeta}} & \frac{a(x,\beta)}{\iota(\beta)}h_{\pi\pi}(x,\pi) \end{bmatrix}.
\end{aligned} \tag{12.3}$$

Comparing the definition of $\varepsilon(w,\theta)$ in (12.3) with that in (9.2), it is clear that, if $\rho(w,\theta)$ takes the form in (3.2) and a parameter $\zeta$ appears, then Assumption S3$^*$ still implies Assumption S3(i) provided $\rho'(\cdot)$ and $\rho''(\cdot)$ in Assumption S3$^*$ are adjusted to include $\zeta$, evaluated at $\zeta_0$.

# 13. Supplemental Appendix C: Numerical Results

Table S-1 compares the finite-sample ($n = 500$) coverage probabilities of the null-imposed robust CI's for $\beta$ in the STAR model with true and estimated values of $\zeta$. (See the end of the STAR-model numerical-results section in the main paper for further discussion.)

Figures S-1 and S-2 report asymptotic and finite-sample ($n = 500$) densities of the estimators for $\beta$ and $\pi$ in the STAR model when $\pi_0 = -3.0$. Figures S-3 to S-6 report asymptotic and finite-sample (n=500) densities of the $t$ and QLR statistics for $\beta$ and $\pi$ in the STAR model when $\pi_0 = -1.5$. Figures S-7 and S-8 report CP's of nominal 0.95

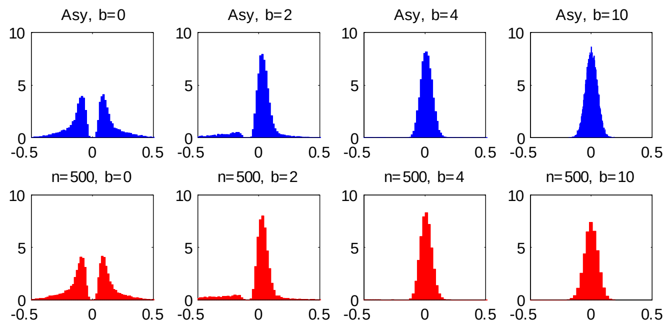


Figure S-1. Asymptotic and Finite-Sample ($n = 500$) Densities of the Estimator of $\beta$ in the STAR Model when $\pi_0 = -3.0$.

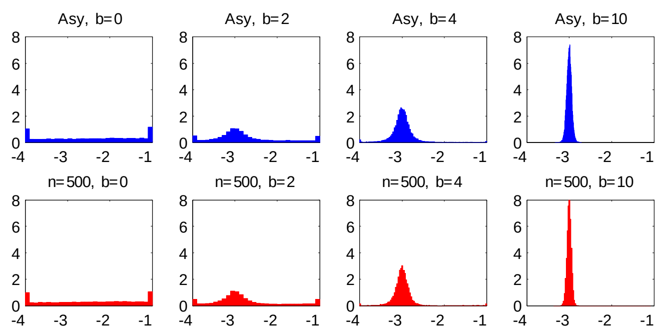


Figure S-2. Asymptotic and Finite-Sample ($n = 500$) Densities of the Estimator of $\pi$ in the STAR Model when $\pi_0 = -3.0$.

standard and robust $|t|$ and QLR CI's for $\beta$ and $\pi$ in the STAR model when $\pi_0 = -3.0$.

Figures S-9 to S-16 are analogous to Figures S-1 to S-8 but for the binary choice model. The true values of $\pi$ considered are $\pi_0 = 1.5$ and $\pi_0 = 2.0$.

Table S-1. Finite-Sample Coverage Probabilities of Null-Imposed Robust CI's for $\beta$ in the STAR Model with True and Estimated Values of $\zeta$, $n = 500$, $\pi_0 = -1.5$[38]

| $b$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t\_\zeta_0$ | 0.939 | 0.950 | 0.946 | 0.947 | 0.948 | 0.947 | 0.944 | 0.946 | 0.949 | 0.949 | 0.950 | 0.947 | 0.956 |
| $t\_\widehat{\zeta}$ | 0.936 | 0.951 | 0.946 | 0.947 | 0.947 | 0.947 | 0.949 | 0.944 | 0.947 | 0.947 | 0.947 | 0.947 | 0.957 |
| $QLR\_\zeta_0$ | 0.923 | 0.932 | 0.930 | 0.925 | 0.923 | 0.924 | 0.916 | 0.921 | 0.923 | 0.926 | 0.929 | 0.932 | 0.933 |
| $QLR\_\widehat{\zeta}$ | 0.920 | 0.935 | 0.927 | 0.924 | 0.926 | 0.919 | 0.915 | 0.908 | 0.915 | 0.922 | 0.925 | 0.924 | 0.929 |

[38] The simulation is conducted with the null value of $b$ and the true value of $\pi$ imposed so that the asymptotic CP is 0.95 for all $b$ values, which serves as a good benchmark. The finite-sample CP's in Table S-1 sometimes differ noticeably from 0.95 due to the small scale of the simulation, i.e., only 1000 simulations repetitions are employed to compute the CP's, as described in footnote 28.

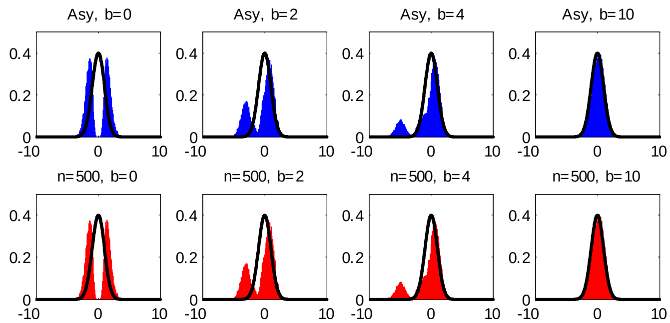


Figure S-3. Asymptotic and Finite-Sample ($n = 500$) Densities of the $t$ Statistic for $\beta$ in the STAR Model when $\pi_0 = -1.5$ and the Standard Normal Density (Black Line).

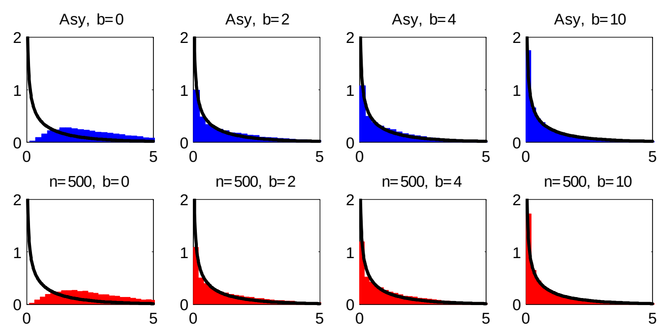


Figure S-4. Asymptotic and Finite-Sample (n=500) Densities of the QLR Statistic for $\beta$ in the STAR Model when $\pi_0 = -1.5$ and the $\chi_1^2$ Density (Black Line).

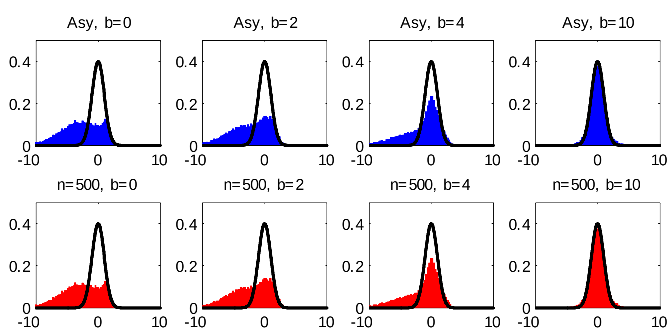


Figure S-5. Asymptotic and Finite-Sample ($n = 500$) Densities of the $t$ Statistic for $\pi$ in the STAR Model when $\pi_0 = -1.5$ and the Standard Normal Density (Black Line).

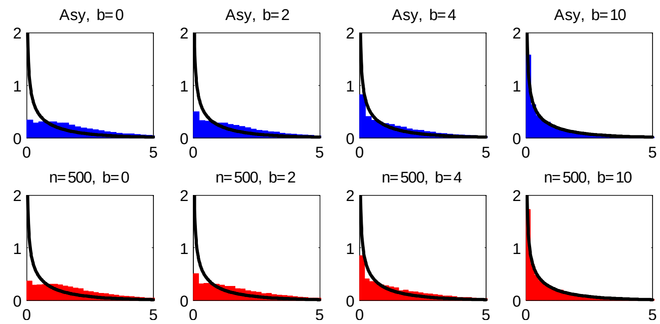


Figure S-6. Asymptotic and Finite-Sample (n=500) Densities of the QLR Statistic for $\pi$ in the STAR Model when $\pi_0 = -1.5$ and the $\chi^2_1$ Density (Black Line).

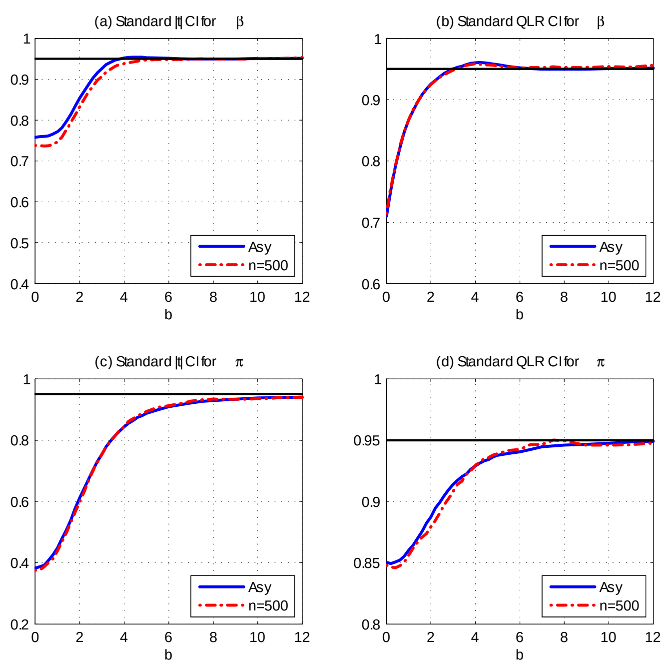


Figure S-7. Coverage Probabilities of Standard $|t|$ and QLR CI's for $\beta$ and $\pi$ in the STAR Model when $\pi_0 = -3.0$.

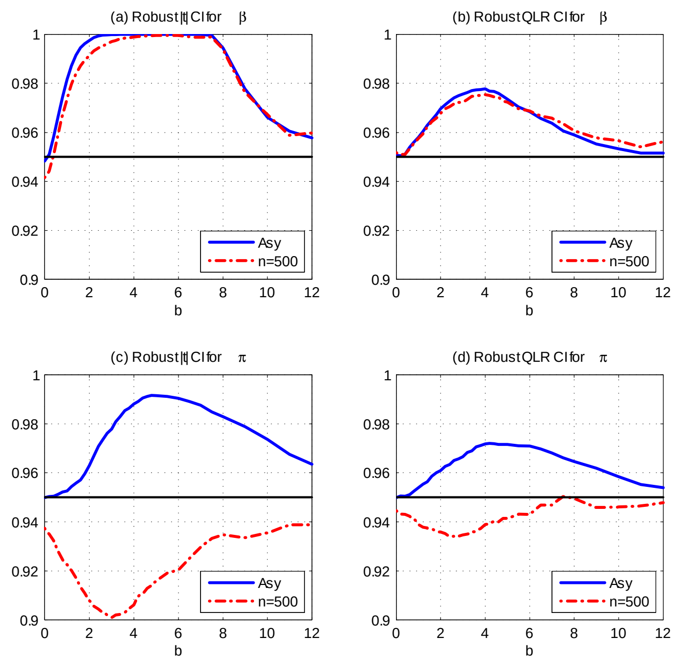


Figure S-8. Coverage Probabilities of Robust $|t|$ and QLR CI's for $\beta$ and $\pi$ in the STAR Model when $\pi_0 = -3.0$, $\kappa = 2.5$, $D = 1$, and $s(x) = \exp(-2x)$.

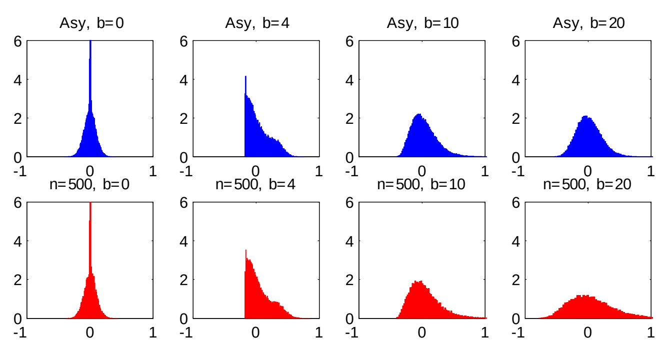


Figure S-9. Asymptotic and Finite-Sample (n=500) Densities of the Estimator of $\beta$ in the Binary Choice Model when $\pi_0 = 2.0$.

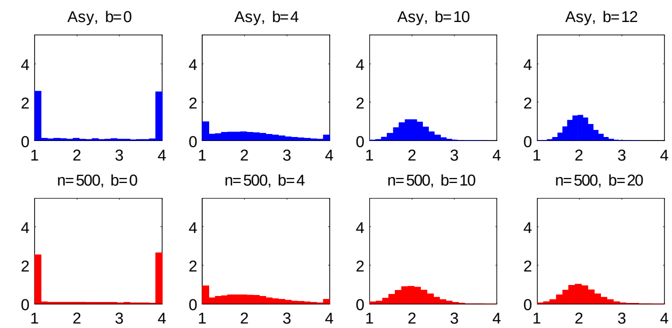


Figure S-10. Asymptotic and Finite-Sample (n=500) Densities of the Estimator of $\pi$ in the Binary Choice Model when $\pi_0 = 2.0$.

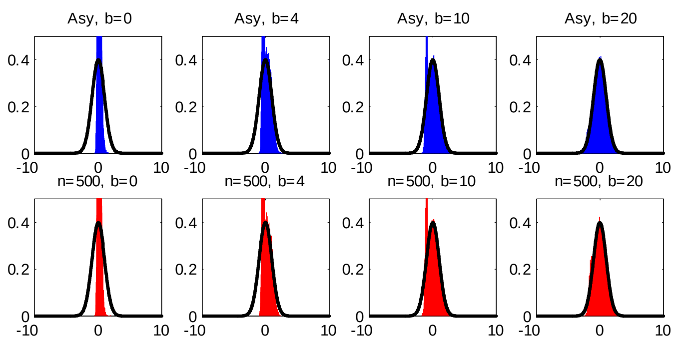


Figure S-11. Asymptotic and Finite-Sample (n=500) Densities of the $t$ Statistic for $\pi$ in the Binary Choice Model when $\pi_0 = 1.5$ and the Standard Normal Density (Black Line).

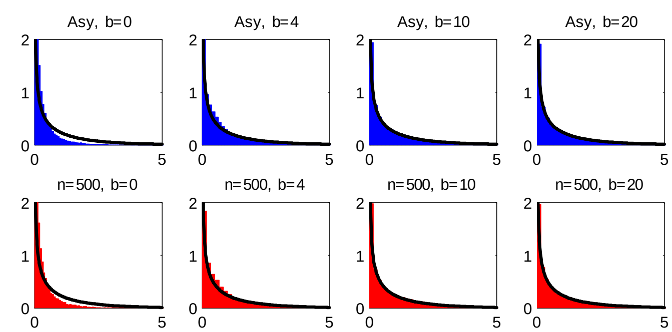


Figure S-12. Asymptotic and Finite-Sample (n=500) Densities of the QLR Statistic for $\pi$ in the Binary Choice Model when $\pi_0 = 1.5$ and the $\chi^2_1$ Density (Black Line).

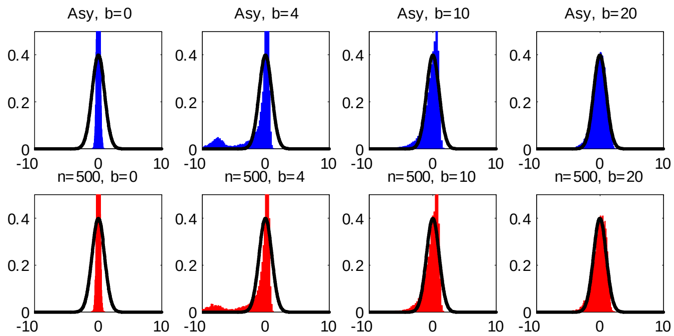


Figure S-13. Asymptotic and Finite-Sample (n=500) Densities of the $t$ Statistic for $\beta$ in the Binary Choice Model when $\pi_0 = 1.5$ and the Standard Normal Density (Black Line).

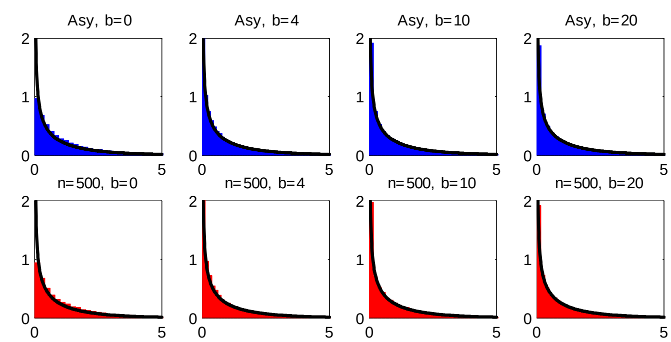


Figure S-14. Asymptotic and Finite-Sample (n=500) Densities of the QLR Statistic for $\beta$ in the Binary Choice Model when $\pi_0 = 1.5$ and the $\chi^2_1$ Density (Black Line).

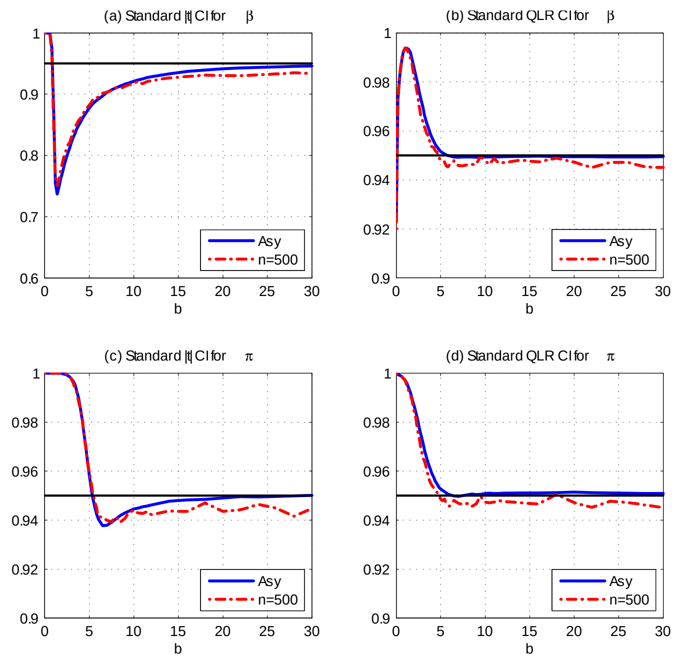


Figure S-15. Coverage Probabilities of Standard $|t|$ and QLR CI's for $\beta$ and $\pi$ in the Binary Choice Model when $\pi_0 = 2.0$.

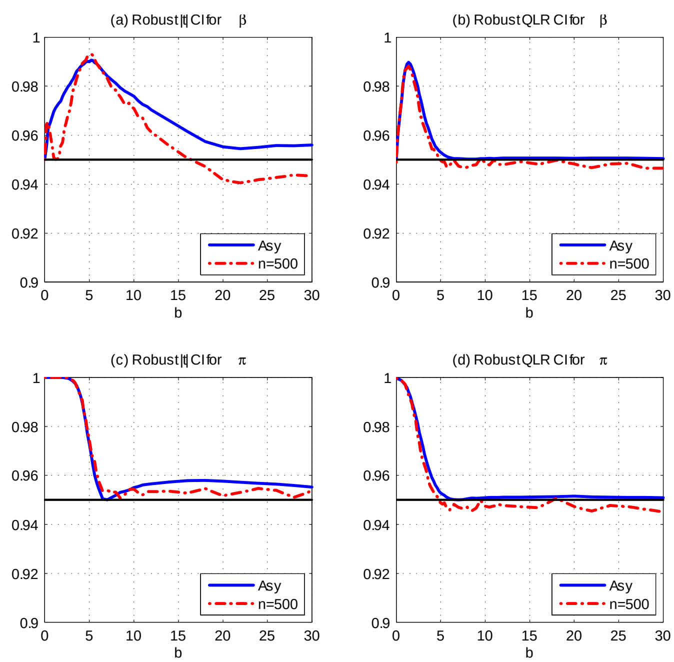


Figure S-16. Coverage Probabilities of Robust $|t|$ and QLR CI's for $\beta$ and $\pi$ in the Binary Choice Model when $\pi_0 = 2.0$, $\kappa = 1.5$, $D = 1$, and $s(x) = \exp(-x/2)$.

# 14. Supplemental Appendix D: Nonlinear Binary Choice Model, Verification of Assumptions

We start by deriving the formulae for the key quantities specified in (3.13). Next, we verify Assumptions S1-S4. Then, we verify Assumptions B1 and B2. Finally, we verify the remaining Assumptions C6, V1, and V2. (Note that Assumption C7 is verified in Section 3.5 of the main paper.)

## 14.1. Derivation of Key Quantities

Here we calculate the key quantities $\Omega(\pi_1, \pi_2; \gamma_0)$, $H(\pi; \gamma_0)$, $J(\gamma_0)$, and $V(\gamma_0)$ that are specified in (3.13).

By (2.4),

$$\begin{aligned} E_{\gamma_0}(Y_i - L_i(\theta_0)|X_i, Z_i) &= 0 \text{ a.s. and} \\ E_{\gamma_0}((Y_i - L_i(\theta_0))^2|X_i, Z_i) &= L_i(\theta_0)(1 - L_i(\theta_0)) \text{ a.s.} \end{aligned} \tag{14.1}$$

For $\gamma_0$ with $\beta_0 = 0$, we have $g_i(\psi_0, \pi) = g_i(\theta_0)$, $L_i(\psi_0, \pi) = L_i(\theta_0)$, $L_i'(\psi_0, \pi) = L_i'(\theta_0)$, and $w_{j,i}(\psi_0, \pi) = w_{j,i}(\theta_0)$ for $j = 1, 2$, $\forall \pi \in \Pi$. In consequence,

$$\begin{aligned} \Omega(\pi_1, \pi_2; \gamma_0) &= S_\psi V^\dagger((\psi_0, \pi_1), (\psi_0, \pi_2); \gamma_0) S_\psi' \\ &= E_{\gamma_0} \frac{L_i'^2(\theta_0)}{L_i(\theta_0)(1 - L_i(\theta_0))} d_{\psi,i}(\pi_1) d_{\psi,i}(\pi_2)', \end{aligned} \tag{14.2}$$

where $S_\psi = [I_{d_\psi} : 0_{d_\psi \times d_\pi}]$, the first equality holds by Lemma 11.1, and the second equality holds by independence across $i$ of $\{W_i : i \leq n\}$ and (14.1).

Now, we have

$$\begin{aligned} \rho_{\psi\psi}(W_i, \psi_0, \pi) &= [w_{1,i}^2(\theta_0)(Y_i - L_i(\theta_0))^2 + w_{2,i}(\theta_0)(Y_i - L_i(\theta_0))] d_{\psi,i}(\pi) d_{\psi,i}(\pi)' \text{ and} \\ H(\pi; \gamma_0) &= E_{\gamma_0} \rho_{\psi\psi}(W_i, \psi_0, \pi) = E_{\gamma_0} \frac{L_i'^2(\theta_0)}{L_i(\theta_0)(1 - L_i(\theta_0))} d_{\psi,i}(\pi) d_{\psi,i}(\pi)', \end{aligned} \tag{14.3}$$

where the first equality uses (3.7), the second equality holds by Lemma 11.1, and the third equality uses (14.1).

In addition, we have

$$
\begin{aligned}
V(\gamma_0) &= V^\dagger(\theta_0, \theta_0; \gamma_0) = Var_{\gamma_0}(\rho_\theta^\dagger(W_i, \theta_0)) \\
&= E_{\gamma_0} w_{1,i}^2(\theta_0)(Y_i - L_i(\theta_0))^2 d_i(\pi_0) d_i(\pi_0)' \\
&= E_{\gamma_0} \frac{L_i'^2(\theta_0)}{L_i(\theta_0)(1 - L_i(\theta_0))} d_i(\pi_0) d_i(\pi_0)', \qquad (14.4)
\end{aligned}
$$

where the first equality holds by (3.5) and the second equality holds by independence across $i$ of $\{W_i : i \leq n\}$ and (14.1).

Next, we have

$$
\begin{aligned}
J(\gamma_0) &= E_{\gamma_0} \rho_{\theta\theta}^\dagger(W_i, \theta_0) \\
&= E_{\gamma_0}[w_{1,i}^2(\theta_0)(Y_i - L_i(\theta_0))^2 + w_{2,i}(\theta_0)(Y_i - L_i(\theta_0))] d_i(\pi) d_i(\pi)' \\
&= E_{\gamma_0} \frac{L_i'^2(\theta_0)}{L_i(\theta_0)(1 - L_i(\theta_0))} d_i(\pi) d_i(\pi)', \qquad (14.5)
\end{aligned}
$$

where the first equality holds by Lemma 11.1, the second equality holds using (3.8), and the third equality holds by (14.1).

The matrix $K(\pi; \gamma_0)$ is derived in Section 14.7 below.

## 14.2. Verification of Assumptions S1 and S2

Given that $\{W_i : i \geq 1\}$ are i.i.d. under $\gamma_0$ $\forall \gamma_0 \in \Gamma$, Assumption S1 holds with $q = 2 + \delta$ for $\delta > 0$.

Assumption S2(i) holds with

$$
\rho(W_i, \theta) = -[Y_i \log L_i(\theta) + (1 - Y_i) \log(1 - L_i(\theta))]. \qquad (14.6)
$$

When $\beta = 0$, $L_i(\theta) = L(\beta h(X_i, \pi) + Z_i'\zeta)$ does not depend on $\pi$ and, hence, $\rho(W_i, \theta)$ does not depend on $\pi$. This verifies Assumption S2(ii).

To verify Assumptions S2(iii) and S2(iv), we have

$$
E_{\gamma_0}(\rho(W_i, \theta)|X_i, Z_i) = -[L_i(\theta_0) \log L_i(\theta) + (1 - L_i(\theta_0)) \log(1 - L_i(\theta))] \qquad (14.7)
$$

because $E_{\gamma_0}(Y_i|X_i, Z_i) = L_i(\theta_0)$ by (2.4). Now we view $E_{\gamma_0}(\rho(W_i, \theta)|X_i, Z_i)$ as a function of $L_i(\theta)$. The first- and second-order derivatives of $E_{\gamma_0}(\rho(W_i,\theta)|X_i,Z_i)$ wrt $L_i(\theta)$ are

$$\frac{\partial}{\partial L_i(\theta)}E_{\gamma_0}(\rho(W_i,\theta)|X_i,Z_i) = \frac{L_i(\theta)-L_i(\theta_0)}{L_i(\theta)(1-L_i(\theta))} \text{ and}$$
$$\frac{\partial^2}{\partial L_i^2(\theta)}E_{\gamma_0}(\rho(W_i,\theta)|X_i,Z_i) = \frac{L_i(\theta_0)+L_i^2(\theta)-2L_i(\theta)L_i(\theta_0)}{L_i^2(\theta)(1-L_i(\theta))^2}, \tag{14.8}$$

see (14.48) below. The second-order derivative is positive for all $\theta \in \Theta$ because its numerator is greater than $(L_i(\theta_0)-L_i(\theta))^2 \geq 0$. When $L_i(\theta) = L_i(\theta_0)$, the first-order derivative is 0. Hence, $E_{\gamma_0}(\rho(W_i,\theta)|X_i,Z_i)$, viewed as a function of $L_i(\theta)$, has a unique global minima at $L_i(\theta_0)$. Because $L'(u) > 0$, $E_{\gamma_0}\rho(W_i,\theta)$ is minimized at $\theta$ if and only if $P_{\gamma_0}(g_i(\theta) = g_i(\theta_0)) = 1$.

When $\beta_0 = 0$, $g_i(\theta) - g_i(\theta_0) = \beta h(X_i,\pi) + (\zeta-\zeta_0)'Z_i$. Because $P_{\gamma_0}(a'(h(X_i,\pi),Z_i) = 0) < 1$ for all $a \in R^{d_\zeta+1}$ with $a \neq 0$ (by the definition of $\Phi^*$ in (3.12)), $P_{\gamma_0}(g_i(\theta)-g_i(\theta_0) = 0) = 1$ if and only if $\beta = 0$ and $\zeta = \zeta_0$. This implies Assumption S2(iii).

When $\beta_0 \neq 0$, $g_i(\theta) - g_i(\theta_0) = \beta h(X_i,\pi) - \beta_0 h(X_i,\pi_0) + (\zeta-\zeta_0)'Z_i$. Because $P_{\gamma_0}(a'(h(X_i,\pi),h(X_i,\pi_0),Z_i) = 0) < 1$ for all $a \in R^{d_\zeta+2}$ with $a \neq 0$ and $\pi \neq \pi_0$, $P_{\gamma_0}(g_i(\theta) - g_i(\theta_0) = 0) < 1$ when $\pi \neq \pi_0$. When $\pi = \pi_0$, $g_i(\theta) - g_i(\theta_0) = (\beta - \beta_0)h(X_i,\pi v) + (\zeta-\zeta_0)'Z_i$. Because $P_{\gamma_0}(a'(h(X_i,\pi),Z_i) = 0) < 1$ for all $a \in R^{d_\zeta+1}$ with $a \neq 0$, $P_{\gamma_0}(g_i(\theta) - g_i(\theta_0) = 0) = 1$ if and only if $\zeta = \zeta_0$, $\beta = \beta_0$, and $\pi = \pi_0$. This verifies Assumption S2(iv).

Assumption S2(v) holds because $\Psi(\pi)$ does not depend on $\pi$ and $\Psi$, $\Pi$, and $\Theta$ are all compact. Assumption S2(vi) holds automatically because $\Psi(\pi)$ does not depend on $\pi$.

## 14.3. Verification of Assumption S3(i)

To verify Assumption S3(i), note that $E_{\gamma_0}(Y_i|X_i,Z_i) = P_{\gamma_0}(Y_i = 1|X_i,Z_i) = L_i(\theta_0)$ by (2.4). Hence, $E_{\gamma_0}(Y_i - L_i(\theta_0)|X_i,Z_i) = 0$ implies $E_{\gamma_0}\varepsilon(W_i,\theta_0) = 0$ by the law of iterated expectations (LIE).

Let $\overline{L}_i' = \sup_{\theta\in\Theta}|L_i'(\theta)|$ and $\overline{L}_i'' = \sup_{\theta\in\Theta}|L_i''(\theta)|$.

A mean-value expansion of $L_i(\psi_0,\pi)$ wrt $\pi$ around $\pi_0$ yields

$$\begin{aligned} L_i(\psi_0,\pi) - L_i(\theta_0) &= L_i'(\psi_0,\widetilde{\pi})\frac{\partial g_i(\psi_0,\widetilde{\pi})}{\partial \pi'}(\pi-\pi_0) \\ &= L_i'(\psi_0,\widetilde{\pi})h_\pi(X_i,\psi_0,\widetilde{\pi})'\beta_0(\pi-\pi_0), \end{aligned} \tag{14.9}$$

where $\widetilde{\pi}$ is between $\pi$ and $\pi_0$. To verify the second part of Assumption S3(i), we have

$$\begin{aligned}
&||E_{\gamma_0}(w_{1,i}(\psi_0,\pi)[Y_i - L_i(\psi_0,\pi)]h_\pi(X_i,\pi))|| \\
&= ||E_{\gamma_0}(w_{1,i}(\psi_0,\pi)[L_i(\theta_0) - L_i(\psi_0,\pi)]h_\pi(X_i,\pi))|| \\
&\leq |\beta_0| \cdot ||\pi - \pi_0|| E_{\gamma_0}(\overline{w}_{1,i}\overline{L}_i'\overline{h}_{\pi,i}^2) \leq C|\beta_0| \cdot ||\pi - \pi_0||, \qquad (14.10)
\end{aligned}$$

for some $C < \infty$, where the equality holds by LIE, the first inequality holds using (14.9), and the second inequality holds by the Cauchy-Schwarz inequality and the moment conditions in (3.12).

Similarly, we have

$$||E_{\gamma_0}(w_{1,i}(\psi_0,\pi)[Y_i - L_i(\psi_0,\pi)]h_{\pi\pi}(X_i,\pi))|| \leq C|\beta_0| \cdot ||\pi - \pi_0|| \qquad (14.11)$$

for some $C < \infty$. By (3.8), (14.10), and (14.11), the second part of Assumption S3(i) holds.

## 14.4. Verification of Assumption S3(ii)

Next, we verify Assumption S3(ii). We use the following generic results in the calculations below. Let $A = aa'$, where $a = (a_1', ..., a_p')' \in R^{d_a}$ and $a_1, ..., a_p$ are vectors (possibly of different dimensions). Then,

$$||A|| = \left(\sum_{j=1}^{p}\sum_{k=1}^{p}||a_j a_k'||^2\right)^{1/2} = \sum_{j=1}^{p}||a_j||^2, \qquad (14.12)$$

where the first equality holds by the definition of $||A||$ and the second equality holds because $||ab'|| = ||a|| \cdot ||b||$ for vectors $a$ and $b$. Similarly, let $A^* = a^*a^{*\prime}$, where $a_1^*, ..., a_p^*$ are sub-vectors of $a^*$ that are conformable with $a_1, ..., a_p$. Then,

$$\begin{aligned}
&||A - A^*|| = ||aa' - a^*a^{*\prime}|| \leq ||a(a-a^*)'|| + ||(a-a^*)a^{*\prime}|| \\
&= (||a|| + ||a^*||)||a - a^*|| \leq \sum_{j=1}^{p}(||a_j|| + ||a_j^*||)\sum_{k=1}^{p}||a_k - a_k^*||, \qquad (14.13)
\end{aligned}$$

where the first inequality holds by triangle inequality, the second equality holds because $||ab'|| = ||a|| \cdot ||b||$, and the last inequality holds because $(x^2 + y^2)^{1/2} \leq x + y$ for non-negative scalars $x$ and $y$.

Define $v_{1,i}(\theta) = w_{1,i}(\theta)(Y_i - L_i(\theta))$, $v_{2,i}(\theta) = w_{2,i}(\theta)(Y_i - L_i(\theta))$, and $\overline{\beta} = \max\{b_1, b_2\}$. Below, let $\theta_1, \theta_2 \in \Theta$ with $||\theta_1 - \theta_2|| \leq \delta$ for $\delta > 0$.

By the triangle inequality, we have

$$\begin{aligned} & ||\rho_{\psi\psi}(W_i, \theta_1) - \rho_{\psi\psi}(W_i, \theta_2)|| \\ \leq & \left(||v_{1,i}^2(\theta_1) - v_{1,i}^2(\theta_2)|| + ||v_{2,i}(\theta_1) - v_{2,i}(\theta_2)||\right) \cdot ||d_{\psi,i}(\pi_1) d_{\psi,i}(\pi_1)'|| \\ & + \left(||v_{1,i}^2(\theta_2)|| + ||v_{2,i}(\theta_2)||\right) \cdot ||d_{\psi,i}(\pi_1) d_{\psi,i}(\pi_1)' - d_{\psi,i}(\pi_2) d_{\psi,i}(\pi_2)'||. \end{aligned} \quad (14.14)$$

Note that

$$\begin{aligned} ||v_{1,i}^2(\theta_1) - v_{1,i}^2(\theta_2)|| &= ||v_{1,i}(\theta_1) - v_{1,i}(\theta_2)|| \cdot ||v_{1,i}(\theta_1) + v_{1,i}(\theta_2)||, \text{ where} \\ ||v_{1,i}(\theta_1) - v_{1,i}(\theta_2)|| &\leq ||w_{1,i}(\theta_1) - w_{1,i}(\theta_2)|| \cdot ||Y_i - L_i(\theta_1)|| + ||w_{1,i}(\theta_2)|| \cdot ||L_i(\theta_1) - L_i(\theta_2)|| \\ &\leq \left(M_1(W_i) + \overline{w}_{1,i} \overline{L}_i'(\overline{h}_i + ||Z_i|| + \overline{\beta} \cdot \overline{h}_{\pi,i})\right) \delta, \text{ and} \\ ||v_{1,i}(\theta_1) + v_{1,i}(\theta_2)|| &\leq 2\overline{w}_{1,i}, \end{aligned} \quad (14.15)$$

where the first inequality follows from the triangle inequality, the second inequality holds by (i) $||w_{1,i}(\theta_1) - w_{1,i}(\theta_2)|| \leq M_1(W_i)\delta$, (ii) $||Y_i - L_i(\theta)|| \leq 1$, and (iii) $||L_i(\theta_1) - L_i(\theta_2)|| \leq \overline{L}_i'(\overline{h}_i + ||Z_i|| + \overline{\beta} \cdot \overline{h}_{\pi,i})\delta$ by a mean-value expansion of $L_i(\theta) = L(g_i(\theta))$ wrt $\theta$, and the third inequality follows from the triangle inequality and $||Y_i - L_i(\theta)|| \leq 1$. Similarly,

$$\begin{aligned} & ||v_{2,i}(\theta_1) - v_{2,i}(\theta_2)|| \leq \left(M_2(W_i) + \overline{w}_{2,i} \overline{L}_i'(\overline{h}_i + ||Z_i|| + \overline{\beta} \cdot \overline{h}_{\pi,i})\right) \delta, \\ & ||v_{1,i}^2(\theta_2)|| \leq \overline{w}_{1,i}^2, \text{ and } ||v_{2,i}(\theta_2)|| \leq \overline{w}_{2,i}. \end{aligned} \quad (14.16)$$

Applying the inequality in (14.12) with $a = d_{\psi,i}(\pi_1) = (h(X_i, \pi_1), Z_i')'$, we have

$$||d_{\psi,i}(\pi_1) d_{\psi,i}(\pi_1)'|| \leq \overline{h}_i^2 + ||Z_i||^2. \quad (14.17)$$

Applying the inequality in (14.13) with $a = d_{\psi,i}(\pi_1)$, $a^* = d_{\psi,i}(\pi_2)$, $||a_1 - a_1^*|| \leq \overline{h}_{\pi,i}||\pi_1 - \pi_2||$, and $||a_2 - a_2^*|| = 0$, we have

$$||d_{\psi,i}(\pi_1) d_{\psi,i}(\pi_1)' - d_{\psi,i}(\pi_2) d_{\psi,i}(\pi_2)'|| \leq 2(\overline{h}_i + ||Z_i||)\overline{h}_{\pi,i}||\pi_1 - \pi_2||. \quad (14.18)$$

Equations (14.14)-(14.18) combine to yield

$$||\rho_{\psi\psi}(W_i, \theta_1) - \rho_{\psi\psi}(W_i, \theta_2)|| \leq M_\psi(W_i)\delta, \quad (14.19)$$

where

$$M_\psi(W_i) = \Big[2\overline{w}_{1,i}\Big(M_1(W_i) + \overline{w}_{1,i}\overline{L}_i'(\overline{h}_i + ||Z_i|| + \overline{\beta}\cdot\overline{h}_{\pi,i})\Big) + M_2(W_i) \\ + \overline{w}_{2,i}\overline{L}_i'(\overline{h}_i + ||Z_i|| + \overline{\beta}\cdot\overline{h}_{\pi,i})\Big](\overline{h}_i^2 + ||Z_i||^2) + 2\left(\overline{w}_{1,i}^2 + \overline{w}_{2,i}\right)(\overline{h}_i + ||Z_i||)\overline{h}_{\pi,i}. \quad (14.20)$$

To show $||\rho_{\theta\theta}^\dagger(\theta_1) - \rho_{\theta\theta}^\dagger(\theta_2)|| \leq M_{\theta\theta}(W_i)\delta$ for some function $M_{\theta\theta}(W_i)$, the calculation is the same as that above with $d_{\psi,i}(\pi)$ replaced by $d_i(\pi)$. The inequalities in (14.17) and (14.18) become

$$||d_i(\pi_1)d_i(\pi_1)'|| \leq \overline{h}_i^2 + ||Z_i||^2 + \overline{h}_{\pi,i}^2 \text{ and} \\ ||d_i(\pi_1)d_i(\pi_1)' - d_i(\pi_2)d_i(\pi_2)'|| \leq 2(\overline{h}_i + ||Z_i|| + \overline{h}_{\pi,i})(\overline{h}_{\pi,i} + \overline{h}_{\pi\pi,i})\cdot||\pi_1 - \pi_2||. \quad (14.21)$$

By the same arguments as those used in (14.14)-(14.20), we have

$$M_{\theta\theta}(W_i) = \Big[2\overline{w}_{1,i}\Big(M_1(W_i) + \overline{w}_{1,i}\overline{L}_i'(\overline{h}_i + ||Z_i|| + \overline{\beta}\cdot\overline{h}_{\pi,i})\Big) \\ + M_2(W_i) + \overline{w}_{2,i}\overline{L}_i'(\overline{h}_i + ||Z_i|| + \overline{\beta}\cdot\overline{h}_{\pi,i})\Big] \times (\overline{h}_i^2 + ||Z_i||^2 + \overline{h}_{\pi,i}^2) \\ + 2\left(\overline{w}_{1,i}^2 + \overline{w}_{2,i}\right)(\overline{h}_i + ||Z_i|| + \overline{h}_{\pi,i})(\overline{h}_{\pi,i} + \overline{h}_{\pi\pi,i}). \quad (14.22)$$

Next, we show $||\rho_\theta^\dagger(\theta_1) - \rho_\theta^\dagger(\theta_2)|| \leq M_\theta(W_i)\delta$ for some function $M_\theta(W_i)$. To this end, note that

$$||\rho_\theta^\dagger(\theta_1) - \rho_\theta^\dagger(\theta_2)|| \\ \leq ||v_{1,i}(\theta_1) - v_{1,i}(\theta_2)||\cdot(\overline{h}_i + ||Z_i|| + \overline{h}_{\pi,i}) + ||v_{1,i}(\theta_2)||\cdot(\overline{h}_{\pi,i} + \overline{h}_{\pi\pi,i})\delta, \quad (14.23)$$

where $||v_{1,i}(\theta_1) - v_{1,i}(\theta_2)||$ satisfies the inequality in (14.15) and $||v_{1,i}(\theta_2)|| \leq \overline{w}_{1,i}$. Hence,

$$M_\theta(W_i) = \Big(M_1(W_i) + \overline{w}_{1,i}\overline{L}_i'(\overline{h}_i + ||Z_i|| + \overline{\beta}\cdot\overline{h}_{\pi,i})\Big)(\overline{h}_i + ||Z_i|| + \overline{h}_{\pi,i}) + \overline{w}_{1,i}(\overline{h}_{\pi,i} + \overline{h}_{\pi\pi,i}). \quad (14.24)$$

Next, we show $||\varepsilon(W_i,\theta_1) - \varepsilon(W_i,\theta_2)|| \leq M_\varepsilon(W_i)\delta$ for some function $M_\varepsilon(W_i)$. To this end, note that

$$\begin{aligned} ||\varepsilon(W_i,\theta_1)-\varepsilon(W_i,\theta_2)|| &\le 2||v_{1,i}(\theta_1)h_\pi(X_i,\pi_1)-v_{1,i}(\theta_2)h_\pi(X_i,\pi_2)|| \\ &\quad +||v_{1,i}(\theta_1)h_{\pi\pi}(X_i,\pi_1)-v_{1,i}(\theta_2)h_{\pi\pi}(X_i,\pi_2)|| \\ &\le 2||v_{1,i}(\theta_1)-v_{1,i}(\theta_2)||\overline{h}_{\pi,i}+2||v_{1,i}(\theta_2)||\overline{h}_{\pi\pi,i}\delta \qquad (14.25) \\ &\quad +||v_{1,i}(\theta_1)-v_{1,i}(\theta_2)||\overline{h}_{\pi\pi,i}+||v_{1,i}(\theta_2)||M_h(W_i)\delta, \end{aligned}$$

where $||v_{1,i}(\theta_1)-v_{1,i}(\theta_2)||$ satisfies the inequality in (14.15), $||v_{1,i}(\theta_2)||\le\overline{w}_{1,i}$, the first inequality follows from a mean-value expansion of $h_\pi(X_i,\pi)$ wrt $\pi$ and the second inequality follows from $||h_{\pi\pi}(X_i,\pi_1)-h_{\pi\pi}(X_i,\pi_2)||\le M_h(X_i)\cdot||\pi_1-\pi_2||$. By (14.25), we have

$$M_\varepsilon(W_i)=\left(M_1(W_i)+\overline{w}_{1,i}\overline{L}_i'(\overline{h}_i+||Z_i||+\overline{\beta}\cdot\overline{h}_{\pi,i})\right)(2\overline{h}_{\pi,i}+\overline{h}_{\pi\pi,i})+\overline{w}_{1,i}(2\overline{h}_{\pi\pi,i}+M_h(W_i)). \qquad (14.26)$$

Hence, Assumption S3(ii) holds with

$$M_1(W_i)=M_\psi(W_i)+M_{\theta\theta}(W_i) \text{ and } M_2(W_i)=M_\theta(W_i)+M_\varepsilon(W_i). \qquad (14.27)$$

## 14.5. Verification of Assumption S3(iii)

The condition $E_{\gamma_0}M_2(W_i)^q\le C_1$ for some $C_1<\infty$ holds if $E_{\gamma_0}M_\theta(W_i)^q\le C_2$ and $E_{\gamma_0}M_\varepsilon(W_i)^q\le C_2$ for some $C_2<\infty$. Because $\overline{L}_i'\le\overline{w}_{1,i}$, $E_{\gamma_0}M_\theta(W_i)^q\le C_2$ and $E_{\gamma_0}M_\varepsilon(W_i)^q\le C_2$ hold provided, for some $C<\infty$, (i) $E_{\gamma_0}M_1^q(W_i)(\overline{h}_i^q+||Z_i||^q+\overline{h}_{\pi,i}^q+\overline{h}_{\pi\pi,i}^q)\le C$, (ii) $E_{\gamma_0}\overline{w}_{1,i}^{2q}(\overline{h}_i^q+||Z_i||^q+\overline{h}_{\pi,i}^q)(\overline{h}_i^q+||Z_i||^q+\overline{h}_{\pi,i}^q+\overline{h}_{\pi\pi,i}^q)\le C$, and (iii) $E_{\gamma_0}\overline{w}_{1,i}^q(\overline{h}_{\pi,i}^q+\overline{h}_{\pi\pi,i}^q+M_h(W_i)^q)\le C$. Condition (i) holds by conditions in (3.12) using Hölder's inequality to give $E_{\gamma_0}M_1^q(W_i)\overline{h}_i^q\le(E_{\gamma_0}M_1^{4q/3})^{3/4}(E_{\gamma_0}\overline{h}_i^{4q})^{1/4}\le C$ and likewise with $||Z_i||$, $\overline{h}_{\pi,i}$, and $\overline{h}_{\pi\pi,i}$ in place of $\overline{h}_i$. Condition (ii) holds by $E_{\gamma_0}\overline{w}_{1,i}^{2q}\overline{h}_i^q||Z_i||^q\le(E_{\gamma_0}\overline{w}_{1,i}^{4q})^{1/2}(E_{\gamma_0}\overline{h}_i^{4q})^{1/4}(E_{\gamma_0}||Z_i||^{4q})^{1/4}\le C$ and likewise with $||Z_i||^q$ and $\overline{h}_{\pi,i}^q$ in place of $\overline{h}_i^q$ and $\overline{h}_i^q$, $\overline{h}_{\pi,i}^q$, and $\overline{h}_{\pi\pi,i}^q$ in place of $||Z_i||^q$. Condition (iii) holds by $E_{\gamma_0}\overline{w}_{1,i}^qM_h(W_i)^q\le(E_{\gamma_0}\overline{w}_{1,i}^{4q})^{1/4}(E_{\gamma_0}M_h(W_i)^{4q/3})^{3/4}\le C$ and likewise with $\overline{h}_{\pi,i}^q$ and $\overline{h}_{\pi\pi,i}^q$ in place of $M_h(W_i)^q$.

The condition $E_{\gamma_0}M_1(W_i)\le C_1$ for some $C_1<\infty$ holds if $E_{\gamma_0}M_\psi(W_i)\le C_2$ and $E_{\gamma_0}M_{\theta\theta}(W_i)\le C_2$ for some $C_2<\infty$. Because $\overline{L}_i'\le\overline{w}_{1,i}$, $E_{\gamma_0}M_\psi(W_i)\le C_2$ and $E_{\gamma_0}M_{\theta\theta}(W_i)\le C_2$ hold provided, for some $C<\infty$, (i) $E_{\gamma_0}M_1(W_i)\overline{w}_{1,i}(\overline{h}_i^2+||Z_i||^2+\overline{h}_{\pi,i}^2)\le C$, (ii) $E_{\gamma_0}\overline{w}_{1,i}^3(\overline{h}_i+||Z_i||+\overline{h}_{\pi,i})(\overline{h}_i^2+||Z_i||^2+\overline{h}_{\pi,i}^2)\le C$, (iii) $E_{\gamma_0}M_2(W_i)(\overline{h}_i^2+||Z_i||^2+\overline{h}_{\pi,i}^2)\le C$, (iv) $E_{\gamma_0}\overline{w}_{1,i}\overline{w}_{2,i}(\overline{h}_i+||Z_i||+\overline{h}_{\pi,i})(\overline{h}_i^2+||Z_i||^2+\overline{h}_{\pi,i}^2)\le C$, (v)

$E_{\gamma_0}\left(\overline{w}_{1,i}^2+\overline{w}_{2,i}\right)(\overline{h}_i+||Z_i||+\overline{h}_{\pi,i})(\overline{h}_{\pi,i}+\overline{h}_{\pi\pi,i})\leq C$. Condition (i) holds by conditions in (3.12) using the Cauchy-Schwarz inequality and $q>2$ to give $E_{\gamma_0}M_1(W_i)\overline{w}_{1,i}\overline{h}_i^2\leq(E_{\gamma_0}M_1(W_i)^2)^{1/2}(E_{\gamma_0}\overline{w}_{1,i}^4)^{1/4}(E_{\gamma_0}\overline{h}_i^8)^{1/4}\leq C$ and likewise with $||Z_i||^2$ and $\overline{h}_{\pi,i}^2$ in place of $\overline{h}_i^2$. Condition (ii) holds by $E_{\gamma_0}\overline{w}_{1,i}^3\overline{h}_i\overline{h}_{\pi,i}^2\leq(E_{\gamma_0}\overline{w}_{1,i}^6)^{1/2}(E_{\gamma_0}\overline{h}_i^4)^{1/4}(E_{\gamma_0}\overline{h}_{\pi,i}^8)^{1/4}\leq C$ and likewise with $||Z_i||$ and $\overline{h}_{\pi,i}$ in place of $\overline{h}_i$ and with $\overline{h}_i^2$ and $||Z_i||^2$ in place of $\overline{h}_{\pi,i}^2$. Condition (iii) holds by $E_{\gamma_0}M_2(W_i)\overline{h}_i^2\leq(E_{\gamma_0}M_2(W_i)^{4/3})^{3/4}(E_{\gamma_0}\overline{h}_i^8)^{1/4}\leq C$ and likewise with $||Z_i||^2$ and $\overline{h}_{\pi,i}^2$ in place of $\overline{h}_i^2$. Condition (iv) holds by $E_{\gamma_0}\overline{w}_{1,i}\overline{w}_{2,i}\overline{h}_i\overline{h}_{\pi,i}^2\leq(E_{\gamma_0}\overline{w}_{1,i}^8)^{1/8}(E_{\gamma_0}\overline{w}_{2,i}^2)^{1/2}(E_{\gamma_0}\overline{h}_i^8)^{1/8}(E_{\gamma_0}\overline{h}_{\pi,i}^8)^{1/4}\leq C$ and likewise with $||Z_i||$ and $\overline{h}_{\pi,i}$ in place of $\overline{h}_i$ and with $\overline{h}_i^2$ and $||Z_i||^2$ in place of $\overline{h}_{\pi,i}^2$. Condition (v) holds by $E_{\gamma_0}\overline{w}_{2,i}\overline{h}_i\overline{h}_{\pi,i}\leq(E_{\gamma_0}\overline{w}_{2,i}^2)^{1/2}(E_{\gamma_0}\overline{h}_i^4)^{1/4}(E_{\gamma_0}\overline{h}_{\pi,i}^4)^{1/4}\leq C$ and likewise with $\overline{w}_{1,i}^2$ in place of $\overline{w}_{2,i}$, $||Z_i||$ and $\overline{h}_{\pi,i}$ in place of $\overline{h}_i$, and $\overline{h}_{\pi\pi,i}$ in place of $\overline{h}_{\pi,i}$.

By (14.6),

$$E_{\gamma_0}\sup_{\theta\in\Theta}|\rho(W_i,\theta)|^{1+\delta}\leq E_{\gamma_0}(\sup_{\theta\in\Theta}|\log L_i(\theta)|+\sup_{\theta\in\Theta}|\log(1-L_i(\theta))|)^{1+\delta}\leq C,\qquad(14.28)$$

for some $C<\infty$, where the first inequality holds because $Y_i$ is 0 or 1 and the second inequality holds by conditions in (3.12).

By (3.7),

$$E_{\gamma_0}\sup_{\theta\in\Theta}||\rho_{\psi\psi}(W_i,\theta)||^{1+\delta}\leq E_{\gamma_0}(\overline{w}_{1,i}^2+\overline{w}_{2,i})^{1+\delta}\sup_{\pi\in\Pi}||d_{\psi,i}(\pi)d_{\psi,i}(\pi)'||^{1+\delta}\leq C\qquad(14.29)$$

for some $C<\infty$, where the first inequality holds by $|Y_i-L_i(\theta)|\leq 1$ and the triangle inequality and the second inequality holds (14.17) and conditions in (3.12). Similarly, we can show $E_{\gamma_0}\sup_{\theta\in\Theta}||\rho_{\theta\theta}^{\dagger}(W_i,\theta)||^{1+\delta}\leq C$ with $d_{\psi,i}(\pi)$ in (14.29) replaced by $d_i(\pi)$ and (14.17) replaced by (14.21).

By (3.8),

$$\begin{aligned}E_{\gamma_0}\sup_{\theta\in\Theta}||\rho_{\theta}^{\dagger}(W_i,\theta)||^q&\leq E_{\gamma_0}(\overline{w}_{1,i}\sup_{\pi\in\Pi}||d_i(\pi)||)^q\\&\leq E_{\gamma_0}\overline{w}_{1,i}^q(\overline{h}_i+||Z_i||+\overline{h}_{\pi,i})^q\leq C\qquad(14.30)\end{aligned}$$

for some $C<\infty$, where the first inequality holds because $|Y_i-L_i(\theta)|\leq 1$, the second inequality holds because $||d_i(\pi)||\leq||h(X_i,\pi)||+||Z_i||+||h_\pi(X_i,\pi)||$, and the third inequality holds by conditions in (3.12).

By (3.8),

$$E_{\gamma_0} \sup_{\theta \in \Theta} ||\varepsilon(W_i, \theta)||^q \leq E_{\gamma_0} \overline{w}_{1,i}^q (2\overline{h}_{\pi,i} + \overline{h}_{\pi\pi,i})^q \leq C \tag{14.31}$$

for some $C < \infty$, where the the first inequality follows from $|Y_i - L_i(\theta)| \leq 1$ and the second inequality holds by conditions in (3.12).

This completes the verification of Assumption S3(iii).

## 14.6. Verification of Assumptions S3(iv) and S3(v)

To verify Assumption S3(iv), we apply the LIE and obtain

$$\begin{aligned} E_{\gamma_0} \rho_{\psi\psi}(W_i, \theta) &= E_{\gamma_0} [w_{1,i}^2(\theta) e_{1,i}(\theta) + w_{2,i}(\theta) e_{2,i}(\theta)] d_{\psi,i}(\pi) d_{\psi,i}(\pi)', \text{ where} \\ e_{1,i}(\theta) &= E_{\gamma_0}((Y_i - L_i(\theta))^2 | X_i, Z_i) \text{ and } e_{2,i}(\theta) = E_{\gamma_0}(Y_i - L_i(\theta) | X_i, Z_i). \end{aligned} \tag{14.32}$$

When $\beta_0 = 0$, $g_i(\psi_0, \pi) = Z_i' \zeta_0$ and $L_i(\psi_0, \pi) = L(g_i(\psi_0, \pi)) = L(Z_i' \zeta_0)$, $\forall \pi \in \Pi$. By (2.4),

$$e_{1,i}(\psi_0, \pi) = L(Z_i' \zeta_0)(1 - L(Z_i' \zeta_0)) \text{ and } e_{2,i}(\psi_0, \pi) = 0. \tag{14.33}$$

Hence, when $\beta_0 = 0$,

$$E_{\gamma_0} \rho_{\psi\psi}(W_i, \psi_0, \pi) = E_{\gamma_0} \frac{L'^2(Z_i' \zeta_0)}{L(Z_i' \zeta_0)(1 - L(Z_i' \zeta_0))} d_{\psi,i}(\pi) d_{\psi,i}(\pi)'. \tag{14.34}$$

The quantity $E_{\gamma_0} \rho_{\psi\psi}(W_i, \psi_0, \pi)$ is continuous in $\pi$ on $\Pi$ by the DCT using (14.19), (14.20), and the discussion following (14.27). Hence, $\lambda_{\min}(E_{\gamma_0} \rho_{\psi\psi}(W_i, \psi_0, \pi))$ also is continuous on the compact set $\Pi$ and attains its minimum at some point $\pi_{\min} \in \Pi$. Its minimum is zero only if the positive semi-definite matrix $E_{\gamma_0} \rho_{\psi\psi}(W_i, \psi_0, \pi_{\min})$ is not positive definite. The latter is ruled out by the fact that $L'^2(Z_i' \zeta_0)/(L(Z_i' \zeta_0)(1 - L(Z_i' \zeta_0)))$ is positive a.s. and the condition in (3.12) that $P_{\gamma_0}(a'(h(X_i, \pi), Z_i) = 0) < 1$, $\forall \pi \in \Pi$, $\forall a \in R^{d_\zeta + 1}$ with $a \neq 0$. Thus, $\inf_{\pi \in \Pi} \lambda_{\min}(E_{\gamma_0} \rho_{\psi\psi}(W_i, \psi_0, \pi)) > 0$ when $\beta_0 = 0$ and the first part of Assumption S3(iv) holds.

As in (14.32)-(14.34), we can show

$$E_{\gamma_0} \rho_{\theta\theta}^{\dagger}(W_i, \theta_0) = E_{\gamma_0} \frac{L_i'^2(\theta_0)}{L_i(\theta_0)(1 - L_i(\theta_0))} d_i(\pi_0) d_i(\pi_0)' \tag{14.35}$$

by replacing $(\psi_0, \pi)$ with $\theta_0$ and $d_{\psi,i}(\pi)$ with $d_i(\pi_0)$ in the arguments above. Because $L_i'(\theta_0) > 0$ and $0 < L_i(\theta_0) < 1$, $E_{\gamma_0}\rho^{\dagger}_{\theta\theta}(W_i, \theta_0)$ is positive definite because $E_{\gamma_0}d_i(\pi_0)d_i(\pi_0)'$ is positive definite as specified in (3.12). Hence, the second part of Assumption S3(iv) holds.

By (14.4) and (14.35), $V^{\dagger}(\theta_0, \theta_0; \gamma_0) = E_{\gamma_0}\rho^{\dagger}_{\theta\theta}(W_i, \theta_0)$. Hence, $V^{\dagger}(\theta_0, \theta_0; \gamma_0)$ is positive definite.

## 14.7. Verification of Assumption S4

Because $m(W_i, \theta) = \rho_{\psi}(W_i, \theta)$ by Lemma 11.1,

$$\begin{aligned} E_{\gamma_0}m(W_i, \theta) &= E_{\gamma_0}\rho_{\psi}(W_i, \theta) = E_{\gamma_0}w_{1,i}(\theta)(Y_i - L_i(\theta))d_{\psi,i}(\pi) \\ &= E_{\gamma_0}w_{1,i}(\theta)(L_i(\theta_0) - L_i(\theta))d_{\psi,i}(\pi), \end{aligned} \tag{14.36}$$

where $\gamma_0 = (\beta_0, \zeta_0, \pi_0, \phi_0)$, the second equality holds by (3.7), and the third equality holds by iterated expectations and (2.4). In (14.36), $E_{\gamma_0}m(W_i, \theta)$ depends on $\beta_0$ only through $L_i(\theta_0)$. Hence,

$$\begin{aligned} K(\theta; \gamma_0) &= (\partial/\partial\beta_0)E_{\gamma_0}w_{1,i}(\theta)(L_i(\theta_0) - L_i(\theta))d_{\psi,i}(\pi) \\ &= E_{\gamma_0}w_{1,i}(\theta)L_i'(\theta_0)h(X_i, \pi_0)d_{\psi,i}(\pi), \end{aligned} \tag{14.37}$$

where the first equality holds because the observations are identically distributed and the second equality holds by an exchange of $E$ and $\partial$ because $E_{\gamma_0}\sup_{\theta\in\Theta,\theta_0\in\Theta_0}||w_{1,i}(\theta)L_i'(\theta_0)$ $h(X_i, \pi_0)d_{\psi,i}(\pi)|| < \infty$ by conditions in (3.12) and $(\partial/\partial\beta_0)g_i(\theta_0) = h(X_i, \pi_0)$. Hence, Assumption S4(i) holds.

Now we show that Assumptions S4(ii) holds with

$$K(\pi; \gamma_0) = K(\psi_0, \pi; \gamma_0) = E_{\gamma_0}w_{1,i}(\psi_0, \pi)L_i'(\theta_0)h(X_i, \pi_0)d_{\psi,i}(\pi). \tag{14.38}$$

Define $a_i(\theta, \theta_0) = w_{1,i}(\theta)L_i'(\theta_0)h(X_i, \pi_0)d_{\psi,i}(\pi)$. It suffices to show that $E_{\gamma_n}a_i(\theta, \theta^*) \to E_{\gamma_0}a_i(\theta, \theta^*)$ uniformly over $(\theta, \theta^*) \in \Theta \times \Theta^*$ as $\gamma_n \to \gamma_0$ and $E_{\gamma_0}a_i(\theta, \theta^*)$ is continuous in $(\theta, \theta^*)$. The continuity holds by the continuity of $a_i(\theta, \theta^*)$ in $(\theta, \theta^*)$, $E_{\gamma_0}\sup_{(\theta,\theta^*)\in\Theta\times\Theta^*}$ $||a_i(\theta, \theta^*)|| < \infty$ by conditions in (3.12), and the dominated convergence theorem. By Lemma 11.3, the uniform convergence follows from the pointwise convergence and the equicontinuity of $E_{\gamma_0}a_i(\theta, \theta^*)$ in $(\theta, \theta^*)$ over $\gamma_0 \in \Gamma$. The pointwise convergence $E_{\gamma_n}a_i(\theta, \theta^*) \to E_{\gamma_0}a_i(\theta, \theta^*)$ holds because (i) the expectations $E_{\gamma_n}a_i(\theta, \theta^*)$ and $E_{\gamma_0}a_i(\theta, \theta^*)$

depend on $\phi_n$ and $\phi_0$, respectively, but not on $\theta_n$ and $\theta_0$, (ii) $\phi_n \to \phi_0$ implies convergence in distribution by the metric on $\Phi^*$, and (iii) the $L^{1+\delta}$ boundedness of $a_i(\theta, \theta^*)$, i.e., $E_{\gamma_0}||a_i(\theta, \theta^*)||^{1+\delta} \leq C < \infty$ for any $\gamma_0 \in \Gamma$. Equicontinuity holds because for any $(\theta_1, \theta_1^*)$ and $(\theta_2, \theta_2^*)$ with $||(\theta_1, \theta_1^*) - (\theta_2, \theta_2^*)|| \leq \delta$,

$$
\begin{aligned}
& E_{\gamma_0}||a_i(\theta_1, \theta_1^*) - a_i(\theta_2, \theta_2^*)|| \\
\leq\ & E_{\gamma_0}||w_{1,i}(\theta_1) - w_{1,i}(\theta_2)|| \cdot ||L_i'(\theta_1^*)h(X_i, \pi_1^*)d_{\psi,i}(\pi_1)|| \\
& + E_{\gamma_0}||w_{1,i}(\theta_2)|| \cdot ||L_i'(\theta_1^*)h(X_i, \pi_1^*)d_{\psi,i}(\pi_1) - L_i'(\theta_2^*)h(X_i, \pi_2^*)d_{\psi,i}(\pi_2)|| \\
\leq\ & E_{\gamma_0}M_1(W_i)\overline{L}_i'\overline{h}_i \sup_{\pi \in \Pi}||d_{\psi,i}(\pi)||\delta \qquad (14.39) \\
& + E_{\gamma_0}\overline{w}_{1,i}\left[\left(\overline{L}_i''\overline{h}_i + \overline{L}_i\overline{h}_{\pi,i}\right)\sup_{\pi \in \Pi}||d_{\psi,i}(\pi)|| + \overline{L}_i'\overline{h}_i||\sup_{\pi \in \Pi}(\partial/\partial\pi')d_{\psi,i}(\pi)||\right]\delta \leq C\delta
\end{aligned}
$$

for some $C < \infty$ for all $\gamma_0 \in \Gamma$, where the first inequality holds by the triangle inequality, the second inequality follows from $||w_{1,i}(\theta_1) - w_{1,i}(\theta_2)|| \leq M_1(W_i)\delta$ and a mean-value expansion of $L_i'(\theta_1^*)h(X_i, \pi_1^*)d_{\psi,i}(\pi_1)$ wrt $(\theta_1, \theta_1^*)$ around $(\theta_2, \theta_2^*)$, and the third inequality holds by the Cauchy-Schwarz inequality and conditions in (3.12). This completes the verification of Assumption S4.

## 14.8. Verification of Assumptions B1 and B2

Given the definitions in Section 3.2 of the paper, Assumptions B1(i) and B1(iii) follow immediately. Assumption B1(ii) holds by taking $\delta < \min\{b_1^*, b_2^*\}$ and $\mathcal{Z}^0 = int(\mathcal{Z})$.

Given the definitions in Section 3.2 of the paper, the true parameter space $\Gamma$ is of the form in (2.6). Thus, Assumption B2(i) holds immediately. Assumption B2(ii) follows from the form of $\mathcal{B}^*$ given in (2.9). Assumption B2(iii) follows from the form of $\mathcal{B}^*$ and the fact that $\Theta^*$ is a product space and $\Phi^*(\theta_0)$ does not depend on $\beta_0$. Hence, the true parameter space $\Gamma$ satisfies Assumption B2.

## 14.9. Verification of Assumption C6

Assumption C6 holds by Lemma 3.1 under Assumptions S1-S3 and C6$^\dagger$. We now verify Assumption C6. Assumption C6$^\dagger$(i) holds because $\beta$ is a scalar. To verify Assumption C6$^\dagger$(ii), we have

$$\rho_\beta(W_i, \theta) = w_{1,i}(\theta)(Y_i - L_i(\theta))h(X_i, \pi) \text{ and } \rho_\zeta(W_i, \theta) = w_{1,i}(\theta)(Y_i - L_i(\theta))Z_i. \quad (14.40)$$

When $\beta_0 = 0$,

$$\begin{aligned}\rho^*_\psi(W_i, \psi_0, \pi_1, \pi_2) &= w_{1,i}(\psi_0)(Y_i - L_i(\psi_0))h_{Z,i}(\pi_1, \pi_2), \text{ where} \\ w_{1,i}(\psi_0) &= \frac{L'(Z_i'\zeta_0)}{L(Z_i'\zeta_0)(1 - L(Z_i'\zeta_0))}, \ L_i(\psi_0) = L(Z_i'\zeta_0), \text{ and} \\ h_{Z,i}(\pi_1, \pi_2) &= (h(X_i, \pi_1), h(X_i, \pi_2), Z_i')'. \end{aligned} \tag{14.41}$$

The covariance matrix in Assumption C6$^\dagger$(ii) is

$$\begin{aligned}\Omega_G(\pi_1, \pi_2; \gamma_0) &= Cov_{\gamma_0}(\rho^*_\psi(W_i, \psi_0, \pi_1, \pi_2), \rho^*_\psi(W_i, \psi_0, \pi_1, \pi_2)) \\ &= E_{\gamma_0} w_{1,i}^2(\psi_0)(Y_i - L_i(\psi_0))^2 h_{Z,i}(\pi_1, \pi_2) h_{Z,i}(\pi_1, \pi_2)' \\ &= E_{\gamma_0} \frac{L'^2(Z_i'\zeta_0)}{L(Z_i'\zeta_0)(1 - L(Z_i'\zeta_0))} h_{Z,i}(\pi_1, \pi_2) h_{Z,i}(\pi_1, \pi_2)', \end{aligned} \tag{14.42}$$

where the first equality holds because the observations are independent and identically distributed, the second equality follows from $E\rho^*_\psi(W_i, \psi_0, \pi_1, \pi_2) = 0$, which in turn holds by the LIE and (2.4), and the third equality holds by (14.1). Because $L'(Z_i'\zeta_0) > 0$ and $0 < L(Z_i'\zeta_0) < 1$, $\Omega_G(\pi_1, \pi_2; \gamma_0)$ is positive definite because $P(a'h_{Z,i}(\pi_1, \pi_2) = 0) < 1$ for all $a \in R^{d_z+2}$ with $a \neq 0$ by the conditions in (3.12).

## 14.10. Verification of Assumptions V1 and V2

Here we verify Assumptions V1 (scalar $\beta$) and V2, which are stated in Supplemental Appendix B above.

For the binary choice model, the matrices $J(\gamma_0)$ $(= V(\gamma_0))$ and $\widehat{J}_n(\theta)$ $(= \widehat{V}_n(\theta))$ are defined in (3.15) and (5.12), respectively. Define

$$J(\theta; \gamma_0) = E_{\gamma_0} \frac{L_i'^2(\theta)}{L_i(\theta)(1 - L_i(\theta))} d_i(\pi) d_i(\pi)'. \tag{14.43}$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0)$, $\sup_{\theta \in \Theta} ||\widehat{J}_n(\theta) - J(\theta; \gamma_0)|| \to_p 0$ and $J(\theta; \gamma_0)$ is continuous in $\theta$ on $\Theta$ by the uniform law of large numbers in Lemma 11.3, where the smoothness and moment conditions hold by conditions in (3.12). In addition, $J(\theta_0; \gamma_0) = J(\gamma_0)$. This verifies Assumption V1(i) and V1(ii) (for scalar $\beta$).

To verify Assumption V1(iii), note that

$$\Sigma(\theta; \gamma_0) = J^{-1}(\theta; \gamma_0) \text{ and } \Sigma(\pi; \gamma_0) = J^{-1}(\psi_0, \pi; \gamma_0). \tag{14.44}$$

Hence, it suffices to show that (i) $\lambda_{\min}(J(\psi_0, \pi; \gamma_0)) > 0$ and (ii) $\lambda_{\max}(J(\psi_0, \pi; \gamma_0)) < \infty$ for all $\pi \in \Pi$. Property (i) holds by essentially the same argument as in the paragraph following (14.34) with $d_i(\pi)$ in place of $d_{\psi,i}(\pi)$ using the condition in (3.12) that $E_{\gamma_0} d_i(\pi) d_i(\pi)'$ is positive definite $\forall \pi \in \Pi$. Positive definiteness of $E_{\gamma_0} d_i(\pi) d_i(\pi)'$ implies the same for $E_{\gamma_0}[L'^2(Z_i'\zeta_0)/(L(Z_i'\zeta_0)(1 - L(Z_i'\zeta_0)))] d_i(\pi) d_i(\pi)'$ because the latter is well-defined and $L'^2(Z_i'\zeta_0)/(L(Z_i'\zeta_0)(1 - L(Z_i'\zeta_0)))$ is positive a.s. Property (ii) holds by the moment conditions in (3.12). This completes the verification of Assumption V1(iii).

Assumptions V1(i) and V1(ii) hold not only under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, but also under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ in this example. This and $\widehat{\theta}_n \rightarrow_p \theta_0$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, which holds by Lemma 5.3 of AC1, imply that Assumption V2 holds. Among the assumptions employed in Lemma 5.3 of AC1, Assumptions B1, B2, and C7 are verified directly, Assumptions A, B3, and C1-C5 hold by Lemma 11.1 under Assumptions B1, B2, and S1-S4, and Assumption C6 holds by Lemma 3.1 under Assumptions S1-S3 and C6$^{\dagger}$.

## 14.11. Calculation of Partial Derivatives

Here we calculate the partial derivatives of $\rho(W_i, \theta)$ wrt $\theta$. Let $L$ abbreviate $L(g_i(\theta))$. The first-order derivative wrt $\theta$ is

$$\begin{aligned} \rho_\theta(W_i, \theta) &= -\left[\frac{Y_i}{L} - \frac{1 - Y_i}{1 - L}\right] L' \frac{\partial}{\partial \theta} g_i(\theta) \\ &= -\frac{Y_i - L}{L(1 - L)} L' \frac{\partial}{\partial \theta} g_i(\theta) = w_{1,i}(\theta)(Y_i - L) B(\beta) d_i(\pi), \text{ where} \\ w_{1,i}(\theta) &= \frac{-L'}{L(1 - L)}. \end{aligned} \quad (14.45)$$

Now we calculate the second-order derivatives. To this end, we have

$$\begin{aligned}
&\frac{\partial}{\partial\theta'}\left[\frac{Y_i}{L}-\frac{1-Y_i}{1-L}\right] = \left[\frac{-Y_i}{L^2}+\frac{-(1-Y_i)}{(1-L)^2}\right]L'\frac{\partial}{\partial\theta'}g_i(\theta)\\
=& -\left[\frac{Y_i(1-L)^2+(1-Y_i)L^2}{L^2(1-L)^2}\right]L'\frac{\partial}{\partial\theta'}g_i(\theta)\\
=& -\left[\frac{Y_i-2Y_iL+L^2}{L^2(1-L)^2}\right]L'\frac{\partial}{\partial\theta'}g_i(\theta)\\
=& -\frac{(Y_i-L)^2}{L^2(1-L)^2}L'\frac{\partial}{\partial\theta'}g_i(\theta),\\
&\frac{\partial}{\partial\theta'}L' = L''\frac{\partial}{\partial\theta'}g_i(\theta), \text{ and } \frac{\partial^2}{\partial\theta\partial\theta'}g_i(\theta) = D_i(\theta).
\end{aligned} \tag{14.46}$$

Hence,

$$\begin{aligned}
\rho_{\theta\theta}(W_i,\theta) =& \left[\frac{(Y_i-L)^2}{L^2(1-L)^2}(L')^2-\frac{Y_i-L}{L(1-L)}L''\right]\left(\frac{\partial}{\partial\theta}g_i(\theta)\frac{\partial}{\partial\theta'}g_i(\theta)\right)\\
&-\frac{Y_i-L}{L(1-L)}L'\frac{\partial^2}{\partial\theta\partial\theta'}g_i(\theta)\\
=& [w_{1,i}^2(Y_i-L)^2+w_{2,i}(Y_i-L)]B(\beta)d_i(\pi)d_i(\pi)'B(\beta)\\
&+w_{1,i}(Y_i-L)D_i(\theta), \text{ where}\\
w_{1,i}(\theta) =& \frac{-L'}{L(1-L)} \text{ and } w_{2,i}(\theta) = \frac{-L''}{L(1-L)}.
\end{aligned} \tag{14.47}$$

Lastly, we calculate the derivatives in (14.8). Let $L = L_i(\theta)$ and $L_0 = L_i(\theta_0)$. We have

$$\begin{aligned}
\text{FOC} =& -\frac{L_0}{L}+\frac{1-L_0}{1-L} = \frac{L-L_0}{L(1-L)} \text{ and}\\
\text{SOC} =& \frac{L(1-L)-(L-L_0)(1-2L)}{L^2(1-L)^2}\\
=& \frac{L-L^2-(L-L_0-2L^2+2LL_0)}{L^2(1-L)^2}\\
=& \frac{L_0+L^2-2LL_0}{L^2(1-L)^2} > \frac{(L_0-L)^2}{L^2(1-L)^2} > 0.
\end{aligned} \tag{14.48}$$

# 15. Supplemental Appendix E: STAR Example, Verification of Assumptions

## 15.1. Verification of Assumptions S1 and S2

Assumption S1 holds by Assumption STAR1(ii).

Assumption S2(i) holds with

$$\rho(W_t, \theta) = U_t^2(\theta)/2, \text{ where } U_t(\theta) = Y_t - X_t'\zeta - X_t'\beta \cdot m(Z_t, \pi). \tag{15.1}$$

The residual $U_t(\theta)$ is twice continuously differentiable in $\theta$ for both the logistic function and the exponential function. When $\beta = 0$, $U_t(\theta) = Y_t - X_t'\zeta$, which does not depend on $\pi$. This verifies Assumption S2(ii).

To verify Assumptions S2(iii) and S2(iv), we have

$$\begin{aligned} E_{\gamma_0}\rho(W_t, \theta) &= E_{\gamma_0}\left[Y_t - X_t'\zeta - X_t'\beta \cdot m(Z_t, \pi)\right]^2 \\ &= E_{\gamma_0}\left(U_t - X_t'(\zeta - \zeta_0) - X_t'\left[\beta m(Z_t, \pi) - \beta_0 m(Z_t, \pi_0)\right]\right)^2 \\ &= E_{\gamma_0}U_t^2 + E_{\gamma_0}[X_t'(\zeta - \zeta_0) + X_t'(\beta m(Z_t, \pi) - \beta_0 m(Z_t, \pi_0))]^2. \end{aligned} \tag{15.2}$$

To verify Assumption S2(iii), we need that when $\beta_0 = 0$,

$$E_{\gamma_0}\rho(W_t, \psi, \pi) - E_{\gamma_0}\rho(W_t, \psi_0, \pi) = E_{\gamma_0}[X_t'(\zeta - \zeta_0) + X_t'\beta m(Z_t, \pi)]^2 > 0 \tag{15.3}$$

$\forall \psi \neq \psi_0$ and $\forall \pi \in \Pi$. The inequality in (15.3) holds unless

$$P_{\gamma_0}((X_t' + X_t' m(Z_t, \pi))a = 0) = 1, \tag{15.4}$$

where $a = ((\zeta - \zeta_0)', \beta')'$. By Assumption STAR2(i), (15.4) does not hold for any $a \neq 0$. Hence, the inequality in (15.3) holds $\forall \psi \neq \psi_0$. This completes the verification of Assumption S2(iii).

To verify Assumption S2(iv), we need that when $\beta_0 \neq 0$,

$$\begin{aligned} &E_{\gamma_0}\rho(W_t, \theta) - E_{\gamma_0}\rho(W_t, \theta_0) \\ &= E_{\gamma_0}[X_t'(\zeta - \zeta_0) + X_t'\beta m(Z_t, \pi) - X_t'\beta_0 m(Z_t, \pi_0)]^2 > 0 \end{aligned} \tag{15.5}$$

$\forall\theta \neq \theta_0$. The inequality in (15.5) holds unless

$$P_{\gamma_0}\left(X_t'(\zeta - \zeta_0) + X_t'\beta m(Z_t, \pi) - X_t'\beta_0 m(Z_t, \pi_0) = 0\right) = 1 \tag{15.6}$$

for some $\theta \neq \theta_0$. Because $\beta_0 \neq 0$, Assumption STAR2(i) implies that (15.6) does not hold for any $\pi \neq \pi_0$. When $\pi = \pi_0$, (15.6) becomes

$$P_{\gamma_0}\left(X_t'(\zeta - \zeta_0) + X_t'(\beta - \beta_0) m(Z_t, \pi_0) = 0\right) = 1. \tag{15.7}$$

Because (15.4) does not hold for any $a \neq 0$ for any $\pi \in \Pi$, (15.7) cannot hold for $(\beta, \zeta) \neq (\beta_0, \zeta_0)$. This completes the verification of Assumption S2(iv).

Assumption S2(v) holds by Assumption STAR5(ii). Assumption S2(vi) holds because $\Psi$ does not depend on $\pi$.

## 15.2. Verification of Assumption S3(i)

Now we verify Assumption S3 (vector $\beta$). In the STAR model, $Z_t$ is an element of $X_t$ and the function $\rho(\omega, \theta)$ takes the form in (3.4) with

$$\begin{aligned} a(X_t, \beta) &= X_t'\beta \in R,\ h(X_t, \pi) = m\left(Z_t, \pi\right) \in R, \text{ and} \\ \rho^*(W_t, a(X_t, \beta) h(X_t, \pi), \zeta) &= [Y_t - X_t'\zeta - a(X_t, \beta) h(X_t, \pi)]^2/2. \end{aligned} \tag{15.8}$$

By Lemma 9.1, we verify Assumption S3(i) by showing that Assumption S3$^*$ holds.

To verify Assumption S3$^*$(i), we have

$$\rho'(W_t, a(X_t, \beta_0) h(X_t, \pi_0), \zeta_0) = -[Y_t - X_t'\zeta_0 - a(X_t, \beta_0) h(X_t, \pi_0)] = -U_t. \tag{15.9}$$

Note that $\rho'(\cdot)$ and $\rho''(\cdot)$ in Assumption S3$^*$ are partial derivatives of $\rho^*(\cdot)$ wrt $a(X_t, \beta)$ $h(X_t, \pi)$. Assumption S3$^*$(i) holds immediately by Assumption STAR1(i).

To verify Assumption S3$^*$(ii), we first derive the terms that appear in it. By (15.8),

$$\begin{aligned} &\rho''(W_t, a(X_t, \beta) h(X_t, \pi), \zeta) = 1, \\ &h(X_t, \pi) = m(Z_t, \pi), h_\pi(X_t, \pi) = m_\pi(Z_t, \pi),\ h_{\pi\pi}(X_t, \pi) = m_{\pi\pi}(Z_t, \pi), \\ &a_\beta(X_t, \beta) = X_t,\ a_{\beta\beta}(X_t, \beta) = 0. \end{aligned} \tag{15.10}$$

Assumption S3$^*$(ii) holds because $E_{\gamma_0} \sup_{\pi \in \Pi}(|m(Z_t, \pi)| + ||m_\pi(Z_t, \pi)||) \cdot (|m(Z_t, \pi)| +$

$||m_\pi(Z_t,$
$\pi)|| + ||m_{\pi\pi}(Z_t,\pi)||) \cdot ||X_t||^2 \leq C$ for some $C < \infty$ by Assumption STAR2(iii) and the Cauchy-Schwarz inequality.

This completes the verification of Assumption S3(i).

## 15.3. Verification of Assumption S3(ii)

Next, we verify Assumption S3(ii). We first show some generic results that are used in the calculation below. Let $A = aa'$, where $a = [a_1', ..., a_m']' \in R^{d_a}$ and $a_1, ..., a_m$ are sub-vectors of $a$. Similarly, $A^* = a^* a^{*\prime}$ and $a_1^*, ..., a_m^*$ are sub-vectors of $a^*$. Then,

$$\begin{aligned} ||A - A^*|| = ||aa' - a^* a^{*\prime}|| &\leq \sum_{i=1}^m \sum_{j=1}^m ||a_i a_j' - a_i^* a_j^{*\prime}|| \\ &\leq \sum_{i=1}^m \sum_{j=1}^m \left(||a_i a_j' - a_i a_j^{*\prime}|| + ||a_i a_j^{*\prime} - a_i^* a_j^{*\prime}||\right) \\ &\leq \sum_{i=1}^m (||a_i|| + ||a_i^*||) \sum_{j=1}^m ||a_i - a_i^*||, \end{aligned} \tag{15.11}$$

where the first inequality holds by the inequality $(x^2 + y^2)^{1/2} \leq x + y$ for non-negative scalars $x$ and $y$, the second inequality holds by the triangle inequality, and the third inequality holds by the inequality $||AB|| \leq ||A|| \cdot ||B||$ for matrices $A$ and $B$.

By (9.3),

$$\begin{aligned} &||\rho_{\psi\psi}(W_t, \theta_1) - \rho_{\psi\psi}(W_t, \theta_2)|| \leq ||d_{\psi,t}(\pi_1) d_{\psi,t}(\pi_1)' - d_{\psi,t}(\pi_2) d_{\psi,t}(\pi_2)'|| \\ &\leq 4||X_t|| \cdot ||X_t' m(Z_t, \pi_1) - X_t' m(Z_t, \pi_2)|| \\ &\leq 4||X_t||^2 \cdot \sup_{\pi \in \Pi} ||m_\pi(Z_t, \pi)|| \cdot ||\pi_1 - \pi_2||, \end{aligned} \tag{15.12}$$

where the first inequality holds by applying the inequality in (15.11) to $a = d_{\psi,t}(\pi_1) = (X_t' m(Z_t, \pi_1), X_t')'$ and $a^* = d_{\psi,t}(\pi_2) = (X_t' m(Z_t, \pi_2), X_t')'$ and the second inequality holds by a mean-value expansion of $m(Z_t, \pi)$ wrt $\pi$.

Applying the arguments in (15.12) to $\rho_{\theta\theta}^\dagger(W_t, \theta^+)$ with $a = (X_t' m(Z_t, \pi_1), X_t', \omega_1' X_t m_\pi(Z_t,$

$\pi_1)')'$ and $a^* = (X_t' m(Z_t, \pi_2), X_t', \omega_2' X_t m_\pi(Z_t, \pi_2)')'$ yields

$$||\rho^\dagger_{\theta\theta}(W_t, \theta_1^+) - \rho^\dagger_{\theta\theta}(W_t, \theta_2^+)|| \leq 2||X_t||^2(2 + \sup_{\pi\in\Pi} ||m_\pi(Z_t, \pi)||) \times \left(\sup_{\pi\in\Pi} (||m_\pi(Z_t, \pi)|| + ||m_{\pi\pi}(Z_t, \pi)||) \cdot ||\pi_1 - \pi_2|| + \sup_{\pi\in\Pi} ||m_\pi(Z_t, \pi)|| \cdot ||\omega_1 - \omega_2||\right). \quad (15.13)$$

Therefore, the function $M_1(W_t)$ in Assumption S3(ii) takes the form

$$M_1(W_t) = 4||X_t||^2 \cdot \sup_{\pi\in\Pi} ||m_\pi(Z_t, \pi)|| + 2||X_t||^2(2 + \sup_{\pi\in\Pi} ||m_\pi(Z_t, \pi)||) \cdot \sup_{\pi\in\Pi} (2||m_\pi(Z_t, \pi)|| + ||m_{\pi\pi}(Z_t, \pi)||). \quad (15.14)$$

The form of $M_1(W_t)$ is used in the verification of Assumption S3(iii) below.

Next, we show the form of $M_2(W_t)$ in Assumption S3(ii) (vector $\beta$). By (9.3),

$$||\rho_\psi(W_t, \theta_1) - \rho_\psi(W_t, \theta_2)|| = ||U_t(\theta_1) d_{\psi,t}(\pi_1) - U_t(\theta_2) d_{\psi,t}(\pi_2)|| \leq |U_t(\theta_1) - U_t(\theta_2)| \cdot ||d_{\psi,t}(\pi_2)|| + |U_t(\theta_1)| \cdot ||d_{\psi,t}(\pi_1) - d_{\psi,t}(\pi_2)||, \quad (15.15)$$

where the inequality holds by the triangle inequality and $||aB|| = |a| \cdot ||B||$ when $a$ is a scalar.

Let $\overline{\beta} = \sup_{\theta\in\Theta} ||\beta||$ and $\overline{\zeta} = \sup_{\theta\in\Theta} ||\zeta||$.

Note that in (15.15), the terms concerning $U_t(\theta)$ satisfy

$$\begin{aligned} |U_t(\theta_1) - U_t(\theta_2)| &\leq \sup_{\theta\in\Theta} ||\frac{\partial}{\partial\theta'} U_t(\theta)|| \cdot ||\theta_1 - \theta_2|| \\ &\leq (2||X_t|| + ||X_t|| \cdot \overline{\beta} \cdot \sup_{\pi\in\Pi} ||m_\pi(Z_t, \pi)||) \cdot ||\theta_1 - \theta_2||, \\ |U_t(\theta_1)| &\leq ||Y_t|| + ||X_t||\overline{\zeta} + ||X_t||\overline{\beta}. \end{aligned} \quad (15.16)$$

The terms concerning $d_{\psi,t}(\pi)$ satisfy

$$\begin{aligned} ||d_{\psi,t}(\pi_2)|| &\leq 2||X_t|| \text{ and} \\ ||d_{\psi,t}(\pi_1) - d_{\psi,t}(\pi_2)|| &\leq ||X_t|| \cdot \sup_{\pi\in\Pi} ||m_\pi(Z_t, \pi)|| \cdot ||\pi_1 - \pi_2||. \end{aligned} \quad (15.17)$$

The inequalities in (15.15)-(15.17) imply that

$$
\begin{aligned}
&||\rho_\psi(W_t,\theta_1)-\rho_\psi(W_t,\theta_2)|| \leq M_\psi(W_t)\cdot||\theta_1-\theta_2||, \text{ where}\\
&M_\psi(W_t)=2||X_t||^2(2+\overline{\beta}\cdot\sup_{\pi\in\Pi}||m_\pi(Z_t,\pi)||)\\
&+(|Y_t|+||X_t||\overline{\zeta}+||X_t||\overline{\beta})\cdot||X_t||\cdot\sup_{\pi\in\Pi}||m_\pi(Z_t,\pi)||. \qquad (15.18)
\end{aligned}
$$

Similarly, (9.5) gives

$$
\begin{aligned}
&||\rho_\theta^\dagger(W_t,\theta_1^+)-\rho_\theta^\dagger(W_t,\theta_2^+)|| = ||U_t(\theta_1^+)d_t(\pi_1,\omega_1)-U_t(\theta_2^+)d_t(\pi_2,\omega_2)|| \qquad (15.19)\\
&\leq |U_t(\theta_1^+)-U_t(\theta_2^+)|\cdot||d_t(\pi_2,\omega_2)||+|U_t(\theta_1^+)|\cdot||d_t(\pi_1,\omega_1)-d_t(\pi_2,\omega_2)||.
\end{aligned}
$$

In (15.19), the terms concerning $U_t(\theta^+)$ satisfy that

$$
\begin{aligned}
&U_t(\theta^+)=Y_t-X_t'\zeta-||\beta||\omega'X_t\cdot m(Z_t,\pi),\\
&|U_t(\theta_1^+)|\leq||Y_t||+||X_t||\overline{\zeta}+||X_t||\overline{\beta},\\
&\frac{\partial}{\partial\theta^{+\prime}}U(\theta^+)=-(\omega'X_tm(Z_t,\pi),||\beta||X_t'm(Z_t,\pi),X_t',||\beta||\omega'X_tm_\pi(Z_t,\pi)'),\\
&|U_t(\theta_1^+)-U_t(\theta_2^+)|\leq\sup_{\theta^+\in\Theta^+}||\frac{\partial}{\partial\theta^+}U(\theta^+)||\cdot||\theta_1^+-\theta_2^+||\\
&\leq\left(2+\overline{\beta}\cdot\left(\sup_{\pi\in\Pi}||m_\pi(Z_t,\pi)||+1\right)\right)\cdot||X_t||\cdot||\theta_1^+-\theta_2^+||. \qquad (15.20)
\end{aligned}
$$

In (15.19), the terms concerning $d_t(\pi,\omega)$ satisfy

$$
\begin{aligned}
&||d_t(\pi,\omega)||\leq||X_t||(2+\sup_{\pi\in\Pi}||m_\pi(Z_t,\pi)||) \text{ and}\\
&||d_t(\pi_1,\omega_1)-d_t(\pi_2,\omega_2)||\leq||X_t||\cdot\left(\sup_{\pi\in\Pi}||m_\pi(Z_t,\pi)||+\sup_{\pi\in\Pi}||m_{\pi\pi}(Z_t,\pi)||\right)\cdot||\pi_1-\pi_2||\\
&+||X_t||\cdot\sup_{\pi\in\Pi}||m_\pi(Z_t,\pi)||\cdot||\omega_1-\omega_2||. \qquad (15.21)
\end{aligned}
$$

By (15.19)-(15.21),

$$
\begin{aligned}
&||\rho_\theta^\dagger(W_t,\theta_1^+)-\rho_\theta^\dagger(W_t,\theta_2^+)||\leq M_\rho(W_t)\cdot||\theta_1^+-\theta_2^+||, \text{ where}\\
&M_\rho(W_t)=[2+\overline{\beta}\cdot(\sup_{\pi\in\Pi}||m_\pi(Z_t,\pi)||+1)]\cdot||X_t||^2\cdot(2+\sup_{\pi\in\Pi}||m_\pi(Z_t,\pi)||)\\
&+\left(||Y_t||+||X_t||\overline{\zeta}+||X_t||\overline{\beta}\right)\cdot||X_t||\cdot\sup_{\pi\in\Pi}(2||m_\pi(Z_t,\pi)||+||m_{\pi\pi}(Z_t,\pi)||) (15.22)
\end{aligned}
$$

Another term in Assumption S3(ii) is $||\varepsilon(W_t, \theta_1^+) - \varepsilon(W_t, \theta_2^+)||$, which satisfies

$$\begin{aligned}
&||\varepsilon(W_t, \theta_1^+) - \varepsilon(W_t, \theta_2^+)|| \\
&\leq |U_t(\theta_1^+) - U_t(\theta_2^+)| \cdot ||X_t|| \cdot \sup_{\pi \in \Pi}(2||m_\pi(Z_t, \pi) + ||m_{\pi\pi}(Z_t, \pi)||) \\
&\quad + |U_t(\theta_1^+)| \cdot ||X_t|| \cdot \sup_{\pi \in \Pi}(||3m_{\pi\pi}(Z_t, \pi) + M_{\pi\pi}(Z_t)) \cdot ||\theta_1^+ - \theta_2^+||, \qquad (15.23)
\end{aligned}$$

where $M_{\pi\pi}(Z_t)$ is as in Assumption STAR2. This and the inequalities in (15.20) imply that

$$\begin{aligned}
&||\varepsilon(W_t, \theta_1^+) - \varepsilon(W_t, \theta_2^+)|| \leq M_\varepsilon(W_t) \cdot ||\theta_1^+ - \theta_2^+||, \text{ where } M_\varepsilon(W_t) = \\
&\left(2 + \overline{\beta} \cdot \left(\sup_{\pi \in \Pi} ||m_\pi(Z_t, \pi)|| + 1\right)\right) \cdot ||X_t||^2 \cdot \sup_{\pi \in \Pi}(2||m_\pi(Z_t, \pi)|| + ||m_{\pi\pi}(Z_t, \pi)||) \\
&+ \left(|Y_t| + ||X_t||\overline{\zeta} + ||X_t||\overline{\beta}\right) \cdot ||X_t|| \cdot (\sup_{\pi \in \Pi} ||3m_{\pi\pi}(Z_t, \pi)|| + M_{\pi\pi}(Z_t)). \qquad (15.24)
\end{aligned}$$

Equations (15.18), (15.22), and (15.24) yield that Assumption S3(ii) holds with

$$M_2(W_t) = M_\psi(W_t) + M_\rho(W_t) + M_\varepsilon(W_t). \qquad (15.25)$$

## 15.4. Verification of Assumption S3(iii)

In the verification of Assumption S3(iii) below, we use

$$\begin{aligned}
E_{\gamma_0} \sup_{\theta \in \Theta} |U_t(\theta)|^{2q} &= E_{\gamma_0} \sup_{\theta \in \Theta} |Y_t - X_t'\zeta - X_t'\beta \cdot m(Z_t, \pi)|^{2q} \\
&\leq C_1 E_{\gamma_0}(|Y_t| + ||X_t||)^{2q} \leq C_2 \qquad (15.26)
\end{aligned}$$

for some $C_1, C_2 < \infty$, where the first inequality holds because the parameter spaces of $\zeta$ and $\beta$ are bounded and $|m(Z_t, \pi)| \in [0, 1]$ and the second inequality holds by Holder's inequality and Assumptions STAR1(ii) and STAR2(iii). Because the value of $U_t(\theta)$ does not change when $\theta$ is reparameterized as $\theta^+$, (15.26) is equivalent to

$$E_{\gamma_0} \sup_{\theta^+ \in \Theta^+} |U_t(\theta^+)|^{2q} \leq C \qquad (15.27)$$

for some $C < \infty$.

By (15.1),

$$E_{\gamma_0} \sup_{\theta\in\Theta} |\rho(W_t,\theta)|^{1+\delta} = \frac{1}{2^{1+\delta}} E_{\gamma_0} \sup_{\theta\in\Theta} |U_t(\theta)|^{2(1+\delta)} \le C \tag{15.28}$$

for some $C < \infty$ by (15.26).

By (9.5),

$$\begin{aligned} E_{\gamma_0} \sup_{\theta^+\in\Theta^+} ||\rho_\theta^\dagger(W_t,\theta^+)||^q &\le E_{\gamma_0} \sup_{\theta^+\in\Theta^+} |U_i(\theta^+)|^{2q} E_{\gamma_0} \sup_{\theta^+\in\Theta^+} ||d_t(\pi,\omega)||^{2q} \\ &\le C_1 E_{\gamma_0} \sup_{\pi\in\Pi} (2||X_t|| + ||X_t|| \cdot ||m_\pi(Z_t,\pi)||)^{2q} \le C_2 \end{aligned} \tag{15.29}$$

for some $C_1, C_2 < \infty$, where the first inequality holds by the Cauchy-Schwarz inequality, the second inequality holds by (15.27) and $||AB|| \le ||A|| \cdot ||B||$, and the third inequality holds by Holder's inequality and Assumptions STAR1(ii) and STAR2(iii).

In the calculation of $E_{\gamma_0} \sup_{\theta\in\Theta} ||\rho_{\psi\psi}(W_i,\theta)||^{1+\delta}$ and $E_{\gamma_0} \sup_{\theta\in\Theta} ||\rho_{\theta\theta}^\dagger(W_i,\theta)||^{1+\delta}$ below, we use the following inequality. Let $A = aa'$, where $a = [a_1', ..., a_m'] \in R^{d_a}$ and $a_1, ..., a_m$ are sub-vectors of $a$. Then,

$$||A|| \le \sum_{i=1}^m \sum_{j=1}^m ||a_i a_j'|| \le \left( \sum_{i=1}^m ||a_i|| \right)^2, \tag{15.30}$$

by arguments analogous to those in (15.11).

By (9.3),

$$E_{\gamma_0} \sup_{\theta\in\Theta} ||\rho_{\psi\psi}(W_i,\theta)||^{1+\delta} = E_{\gamma_0} \sup_{\theta\in\Theta} ||d_{\psi,t}(\pi) d_{\psi,t}(\pi)'||^{1+\delta} \le E_{\gamma_0} (2||X_t||)^{2(1+\delta)} \le C \tag{15.31}$$

for some $C < \infty$, where the first inequality holds by (15.30) with $a = (X_t' m(Z_t,\pi), X_t')'$ and the second inequality holds by Assumptions STAR1(ii) and STAR2(iii).

Similarly, by (9.5),

$$\begin{aligned} &E_{\gamma_0} \sup_{\theta^+\in\Theta^+} ||\rho_{\theta\theta}^\dagger(W_i,\theta^+)||^{1+\delta} = E_{\gamma_0} \sup_{\theta^+\in\Theta^+} ||d_t(\pi,\omega) d_t(\pi,\omega)'||^{1+\delta} \\ &\le E_{\gamma_0} \sup_{\pi\in\Pi} (2||X_t|| + ||X_t|| \cdot ||m_\pi(Z_t,\pi)||)^{2(1+\delta)} \le C \end{aligned} \tag{15.32}$$

for some $C < \infty$, where the first inequality holds by (15.30) with $a = (X_t' m(Z_t,\pi), X_t', \omega' X_t m_\pi(Z_t,\pi))'$ and the second inequality holds by Assumption STAR2.

By (9.5),

$$E_{\gamma_0} \sup_{\theta^+ \in \Theta^+} ||\varepsilon(W_i, \theta^+)||^q \leq E_{\gamma_0} \sup_{\theta^+ \in \Theta^+} ||U_t(\theta^+)||^{2q}$$
$$\times E_{\gamma_0} \sup_{\pi \in \Pi} (2||X_t|| \cdot ||m_\pi(Z_t, \pi)|| + ||X_t|| \cdot ||m_{\pi\pi}(Z_t, \pi)||)^{2q} \leq C \qquad (15.33)$$

for some $C < \infty$, where the first inequality holds by the Cauchy-Schwarz inequality and the inequality $||A|| \leq \sum_{i,j} ||A_{i,j}||$ for any matrix $A$, where $A_{i,j}$ denotes an element of $A$, and the second inequality follows from (15.27), Holder's inequality, and Assumptions STAR1(ii) and STAR2(iii).

Finally, $E_{\gamma_0}(M_1(W_t) + M_2(W_t)^q) \leq C$ for some $C < \infty$ by Holder's inequality, (15.14), (15.18), (15.22), (15.24), (15.25), and Assumptions STAR(ii) and STAR2(iii).

This completes the verification of Assumption S3(iii) (vector $\beta$).

## 15.5. Verification of Assumptions S3(iv) and S3(v)

To verify Assumption S3(iv), note that

$$\begin{aligned} E_{\gamma_0} \rho_{\psi\psi}(W_t, \psi_0, \pi) &= E_{\gamma_0} d_{\psi,t}(\pi) d_{\psi,t}(\pi)' \text{ and} \\ E_{\gamma_0} \rho^\dagger_{\theta\theta}(W_t, \theta_0) &= E_{\gamma_0} d_t(\pi_0, \omega_0) d_t(\pi_0, \omega_0)'. \end{aligned} \qquad (15.34)$$

For any $\lambda = (\lambda_1, \lambda_2) \neq 0$, $\lambda_1, \lambda_2 \in R^{d_\beta}$, and $\forall \pi \in \Pi$,

$$\lambda' E_{\gamma_0} d_{\psi,t}(\pi) d_{\psi,t}(\pi)' \lambda = E_{\gamma_0} \left( \lambda_1' X_t m(Z_t, \pi) + \lambda_2' X_t \right)^2 > 0, \qquad (15.35)$$

where the inequality holds by Assumption STAR2(i). This implies that $E_{\gamma_0} d_{\psi,t}(\pi) d_{\psi,t}(\pi)'$ is positive definite $\forall \pi \in \Pi$.

For any $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) \neq 0$, $\lambda_1, \lambda_2 \in R^{d_\beta}$, $\lambda_3, \lambda_4 \in R$, $\forall \omega$ with $||\omega|| = 1$ and $\forall \pi \in \Pi$,

$$\begin{aligned} &\lambda' E_{\gamma_0} d_t(\pi, \omega) d_t(\pi, \omega)' \lambda \qquad (15.36) \\ &= E_{\gamma_0} (\lambda_1' X_t m(Z_t, \pi) + \lambda_2' X_t + \lambda_3 \omega' X_t m_{\pi,1}(Z_t, \pi) + \lambda_4 \omega' X_t m_{\pi,2}(Z_t, \pi))^2 > 0, \end{aligned}$$

where the inequality holds by Assumption STAR2(ii) with $a = (\lambda_1, \lambda_2, \lambda_3 \omega, \lambda_4 \omega)$. Note that $\lambda \neq 0$ implies that $a \neq 0$. The inequality in (15.36) implies that $E_{\gamma_0} \rho^\dagger_{\theta\theta}(W_t, \theta_0)$ is positive definite $\forall \gamma_0 \in \Gamma$.

To verify Assumption S3(v), note that $\forall m \neq 0$,

$$Cov_{\phi}(\rho_{\theta}^{\dagger}(W_t, \theta_0), \rho_{\theta}^{\dagger}(W_{t+m}, \theta_0)) = E_{\gamma_0} U_t U_{t+m} d_t(\pi_0, \omega_0) d_{t+m}(\omega_0, \pi_0)' = 0 \tag{15.37}$$

by Assumption STAR1(i). This yields that

$$\begin{aligned} V^{\dagger}(\theta_0, \theta_0; \gamma_0) &= Cov_{\phi_0} U_t^2 d_t(\pi_0, \omega_0) d_t(\pi_0, \omega_0)' \\ &= E_{\gamma_0} U_t^2 d_t(\pi_0, \omega_0) d_t(\pi_0, \omega_0)', \end{aligned} \tag{15.38}$$

where the second equality uses $E_{\gamma_0} U_t d_t(\pi_0, \omega_0) = 0$ by Assumption STAR1(i). The matrix $E_{\gamma_0} U_t^2 d_t(\pi_0, \omega_0) d_t(\pi_0, \omega_0)'$ is positive definite by the argument in (15.36) with $d_{\psi,t}(\pi)$ replaced by $U_t d_t(\pi, \omega)$ and using $E_{\gamma_0}(U_t^2 | \mathcal{F}_{t-1}) = \sigma^2 > 0$.

## 15.6. Verification of Assumption S4

To verify Assumption S4, we have

$$\begin{aligned} E_{\gamma_0} \rho_{\psi}(W_t, \theta) &= -E_{\gamma_0} U_t(\theta) d_{\psi,t}(\pi) \\ &= -E_{\gamma_0}(U_t + X_t'(\zeta_0 - \zeta) + X_t'[\beta_0 m(Z_t, \pi_0) - \beta m(Z_t, \pi)]) d_{\psi,t}(\pi) \text{ and} \\ K(\theta; \gamma_0) &= -E_{\gamma_0} d_{\psi,t}(\pi) X_t' m(Z_t, \pi_0) \\ &= -E_{\gamma_0} d_{\psi,t}(\pi) d_{\psi,t}(\pi_0)' \cdot S_{\beta}'. \end{aligned} \tag{15.39}$$

where $S_{\beta} = [I_{d_{\beta}} : 0] \in R^{d_{\beta} \times (2 d_{\beta})}$.

Assumption S4(i) holds with $K(\theta; \gamma_0)$ in (15.39) by the moment conditions in Assumption STAR2(iii). To verify Assumption S4(ii), we need to show that $E_{\gamma_0} d_{\psi,t}(\pi) d_{\psi,t}(\pi_0)'$ is continuous in $\pi, \pi_0$, and $\phi$. Continuity in $\pi$ and $\pi_0$ follows from the the continuity of $m(Z_t, \pi)$ in $\pi$ and the moment conditions in Assumption STAR2(iii). Continuity in $\phi$ holds because $\phi_n \to \phi_0$ under $d_{\Phi}$ implies weak convergence of $(Y_t, Y_{t+m})$ for all $t, m \geq 1$, which in turn implies the convergence of $E_{\gamma_n} d_{\psi,t}(\pi) d_{\psi,t}(\pi_0)'$ to $E_{\gamma_0} d_{\psi,t}(\pi) d_{\psi,t}(\pi_0)'$ by the moment conditions in Assumption STAR2(iii).

The continuity in $\pi, \pi_0$, and $\phi$ holds uniformly over $\pi \in \Pi$ by Lemma 11.2 using (i) the pointwise convergence above, (ii) the fact that $E_{\gamma_0} d_{\psi,t}(\pi) d_{\psi,t}(\pi_0)'$ is differentiable in $\pi$ and the partial derivative is bounded over $\pi \in \Pi$, and (iii) the compactness of $\Pi$. This completes the verification of Assumption S4.

## 15.7. Verification of Assumptions B1 and B2

Now we verify Assumptions B1 and B2. Assumptions B1(i) and B1(iii) hold by Assumptions STAR5(i) and STAR5(ii) immediately. Assumption B1(ii) holds with $\mathcal{Z}^0 = int(\mathcal{Z}_0)$ by Assumptions STAR4(iv) and STAR5(iii). Assumption B2(i) holds immediately because the true parameter space $\Gamma$ is of the form in (2.6) and $\Gamma$ is assumed to be compact. Assumption B2(ii) holds by Assumption STAR4(ii). Assumption B2(iii) holds by Assumption STAR4(iv) and the form of the true parameter space in (7.8).

## 15.8. Verification of Assumptions C6 and C7

Assumption C6 is implied by Assumption STAR3(i).

Now we verify Assumption C7 with $H(\pi;\gamma_0)$ and $K(\pi;\gamma_0)$ given in (7.7). By the matrix Cauchy-Schwarz inequality in Tripathi (1999),

$$K(\pi;\gamma_0)'H^{-1}(\pi;\gamma_0)K(\pi;\gamma_0) \leq E_{\gamma_0}X_tX_t'm^2(Z_t,\pi_0). \tag{15.40}$$

The matrix "$\leq$" holds as an equality if and only if $X_tm(Z_t,\pi_0)a+(X_t', X_t'm(Z_t,\pi))c = 0$ with probability 1 for some $a \in R^{d_\beta}$ and $c \in R^{2d_\beta}$ with $(a',c')' \neq 0$. The "$\leq$" holds as an equality uniquely at $\pi = \pi_0$ by Assumption STAR2(i).

**Proof of Lemma 7.1.** We prove Lemma 7.1 by verifying Assumption C6$^\dagger$ and using Lemma 3.1. Note that

$$\begin{aligned}
\rho_\beta(W_t,\psi_0,\pi) &= U_tX_tm(Z_t,\pi), \\
\rho_\zeta(W_t,\psi_0,\pi) &= U_tX_t, \\
\rho_\psi^*(W_i,\psi_0,\pi_1,\pi_2) &= U_td_\psi^*(\pi_1,\pi_2), \text{ where} \\
d_\psi^*(\pi_1,\pi_2) &= (X_t'm(Z_t,\pi_1), X_t'm(Z_t,\pi_2), X_t')'.
\end{aligned} \tag{15.41}$$

The matrix $\Omega_G(\pi_1,\pi_2;\gamma_0)$ that appears in Assumption C6$^\dagger$ takes the form

$$\Omega_G(\pi_1,\pi_2;\gamma_0) = E_{\gamma_0}U_t^2d_\psi^*(\pi_1,\pi_2)d_\psi^*(\pi_1,\pi_2)' \tag{15.42}$$

by Assumption STAR1(i). Assumption C6$^\dagger$(ii) holds by Assumption STAR2(i) and $E_{\gamma_0}(U_t^2|\mathcal{F}_{t-1}) = \sigma^2 > 0$ using arguments analogous to those in (15.36). $\square$

## 15.9. Verification of Assumptions V1 (vector $\beta$) and V2

Here we verify Assumptions V1 (vector $\beta$) and V2, which are stated in Supplemental Appendix B above.

In the STAR model, Assumption V1(i) holds with

$$\begin{aligned} &J(\theta^+;\gamma_0) = E_{\gamma_0} d_t(\pi,\omega) d_t(\pi,\omega)' \text{ and} \\ &V(\theta^+;\gamma_0) = E_{\gamma_0} U_t^2 d_t(\pi,\omega) d_t(\pi,\omega)' \\ &+E_{\gamma_0}[X_t'(\zeta_0-\zeta)+X_t'(||\beta_0||\omega_0 m(Z_t,\pi_0)-||\beta||\omega m(Z_t,\pi))]^2 d_t(\pi,\omega)d_t(\pi,\omega)', \end{aligned} \tag{15.43}$$

by the uniform law of large numbers in Lemma 11.3.

Assumption V1(ii) holds by the continuity of $m(z,\pi)$ and $m_\pi(z,\pi)$ in $\pi$ and Assumption STAR2(iii).

To verify Assumption V1(iii), note that $\Sigma(\pi,\omega;\gamma_0)$ takes the form

$$\begin{aligned} &\Sigma(\pi,\omega;\gamma_0) \\ &= \left(E_{\gamma_0} d_t(\pi,\omega) d_t(\pi,\omega)'\right)^{-1} E_{\gamma_0} U_t^2 d_t(\pi,\omega) d_t(\pi,\omega)' \left(E_{\gamma_0} d_t(\pi,\omega) d_t(\pi,\omega)'\right)^{-1}. \end{aligned} \tag{15.44}$$

Given that $E_{\gamma_0} d_t(\pi,\omega) d_t(\pi,\omega)'$ and $E_{\gamma_0} U_t^2 d_t(\pi,\omega) d_t(\pi,\omega)'$ are both positive definite, $\Sigma(\pi,\omega;\gamma_0)$ is positive definite $\forall \pi \in \Pi$ and $\forall \omega$ with $||\omega|| = 1$.

Because the determinant of $E_{\gamma_0} d_t(\pi,\omega) d_t(\pi,\omega)'$ is bounded away from 0 as a function of $(\pi,\omega)$ $\forall \gamma_0 \in \Gamma$ and $||E_{\gamma_0} d_t(\pi,\omega) d_t(\pi,\omega)'|| \leq C_1$ for some $C_1 < \infty$ $\forall \gamma_0 \in \Gamma$ by Assumption STAR2(iii), we have $||\left(E_{\gamma_0} d_t(\pi,\omega) d_t(\pi,\omega)'\right)^{-1}|| \leq C_2$ for some $C_2 < \infty$. Hence, $||\Sigma(\pi,\omega;\gamma_0)|| \leq C$ $\forall \pi \in \Pi$ and $\forall \omega$ with $||\omega|| = 1$. This completes the verification of Assumption V1(iii).

Assumption V1(iv) holds by Assumption STAR3(ii).

Assumptions V1(i) and V1(ii) hold not only under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, but also under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ in this example. This and $\widehat{\theta}_n \to_p \theta_0$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, which holds by Lemma 5.3 of AC1, imply that Assumption V2 holds. Regarding the assumptions employed in Lemma 5.3 of AC1, Assumptions B1, B2, C6, and C7 are verified above and Assumptions A, B3, and C1-C4 hold by Lemma 11.1 under Assumptions B1, B2, and S1-S4. $\square$

# References


Andrews, D. W. K., 1992. Generic uniform convergence. Econometric Theory 8, 241-257.

Andrews, D. W. K., 1994. Empirical process methods in econometrics. In: Engle, R. F., McFadden, D. (Eds), Handbook of Econometrics, Vol. IV. North-Holland, Amsterdam.

Andrews, D. W. K., Cheng, X., 2012. Estimation and inference with weak, semi-strong, and strong identification. Econometrica 80, 2153-2211.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, Oxford.

Tripathi, G. T., 1999. A matrix extension of the Cauchy-Schwarz inequality. Economics Letters 63, 1-3.

van der Vaart, A. W., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge, UK.