Supplement to
ESTIMATION AND INFERENCE WITH WEAK, SEMI-STRONG, AND STRONG IDENTIFICATION

By

Donald W. K. Andrews and Xu Cheng

October 2010

COWLES FOUNDATION DISCUSSION PAPER NO. 1773

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

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

Supplemental Material

for

Estimation and Inference with Weak, Semi-strong, and Strong Identification

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

Xu Cheng
Department of Economics
University of Pennsylvania

August 2007
Revised: October 6, 2010

10. Outline

We let AC1 abbreviate the main paper "Estimation and Inference with Weak, Semi-strong, and Strong Identification."

This Supplement includes five appendices.

Appendix A provides some sufficient conditions for Assumptions B3, C5, C6, C1, and D1 of AC1 (in that order). Sufficient conditions for other assumptions in AC1 are given in Andrews and Cheng (2008a,b).

Appendix B gives the proofs of the results in AC1.

Appendix C verifies the assumptions of AC1 for the ARMA(1, 1) example.

Appendix D provides some additional Monte Carlo simulation results for the ARMA(1, 1) example.

Appendix E introduces the nonlinear regression example and verifies the assumptions of AC1 for it.

The notational conventions specified at the end of the Introduction to AC1 are used throughout this Supplemental Material.

11. Appendix A: Sufficient Conditions

This Appendix contains sufficient conditions for Assumptions B3, C5, C6, C1, and D1 (in that order). It also contains an initial conditions adjustment to the sufficient conditions for Assumptions C1 and D1 that is useful in some time series contexts.

11.1. Assumption B3

Assumption B3(i) can be verified using a uniform LLN, e.g., as in Andrews (1992). Assumption B3* provides sufficient conditions for Assumptions B3(ii) and B3(iii).

Assumption B3*. (i) $Q(\theta;\gamma_0)$ is continuous on Θ $\forall\gamma_0 \in \Gamma$.
(ii) For any $\pi \in \Pi$, $Q(\psi, \pi; \gamma_0)$ is uniquely minimized by ψ_0 $\forall\gamma_0 \in \Gamma$ with $\beta_0 = 0$.
(iii) $Q(\theta;\gamma_0)$ is uniquely minimized by θ_0 $\forall\gamma_0 \in \Gamma$ with $\beta_0 \neq 0$.
(iv) $\Psi(\pi)$ is compact $\forall\pi \in \Pi$, and Π and Θ are compact.
(v) $\forall\varepsilon > 0$, $\exists\delta > 0$ such that $d_H(\Psi(\pi_1), \Psi(\pi_2)) < \varepsilon$ $\forall\pi_1, \pi_2 \in \Pi$ with $\|\pi_1 - \pi_2\| < \delta$, where $d_H(\cdot)$ is the Hausdorff metric.

Assumption B3*(v) holds immediately in cases where $\Psi(\pi)$ does not depend on π. When $\Psi(\pi)$ depends on π, the boundary of $\Psi(\pi)$ is often a continuous linear function of π, as in the ARMA(1, 1) example. In such cases, it is simple to verify Assumption B3*(v).

Lemma 11.1. *Assumption* B3* *implies Assumptions* B3(ii) *and* B3(iii).

11.2. Assumption C5

The following assumption is sufficient for Assumption C5.

Assumption C5*. (i) For any $i \geq 1$, the marginal distribution of W_i has a density function $f_{W_i}(w;\gamma^*)$ wrt some σ-finite dominating measure μ that does not depend on γ^*, $\forall \gamma^* \in \Gamma$.
(ii) $f_{W_i}(w;\gamma^*)$ is partially differentiable in β^* and the partial derivative is denoted by $f_{\beta,W_i}(w;\gamma^*)$ $\forall i \geq 1$. Both $f_{W_i}(w;\gamma^*)$ and $f_{\beta,W_i}(w;\gamma^*)$ are continuous in γ^* $\forall i \geq 1$, $\forall w \in \mathcal{W}$, $\forall \gamma^* \in \Gamma$, where $\mathcal{W}$ denotes the support of μ.
(iii) For some function $f_{\beta,W}(w;\gamma^*) \in R^{d_\beta}$, $n^{-1}\sum_{i=1}^n f_{\beta,W_i}(w;\gamma^*) \to f_{\beta,W}(w;\gamma^*)$ $\forall w \in \mathcal{W}$, $\forall \gamma^* \in \Gamma$.
(iv) $m(w,\theta)$ is continuous in ψ uniformly over $\pi \in \Pi$ for $\theta \in \Theta$ with $\beta = 0$ $\forall w \in \mathcal{W}$ (i.e., $\sup_{\pi\in\Pi} |m(w,\psi,\pi) - m(w,\psi_0,\pi)| \to 0$ as $\psi \to \psi_0 = (0,\zeta_0)$ $\forall \theta_0 = (\psi_0,\pi_0) \in \Theta$).
(v) $\int_{\mathcal{W}} \sup_{\theta\in\Theta} ||m(w,\theta)|| \cdot \max_{i\leq 1}\{\sup_{\gamma\in N(\gamma^*,\delta)} ||f_{\beta,W_i}(w;\gamma)/f_{W_i}(w;\gamma)|| \cdot \sup_{\gamma\in N(\gamma^*,\delta)} |f_{W_i}(w;\gamma)|\}d\mu(w) < \infty$, where $N(\gamma^*,\delta)$ is a δ-neighborhood of γ^* for some $\delta > 0$, $\forall \gamma^* \in \Gamma$.

Assumption C5*(iii) holds automatically with identically distributed observations. Assumption C5*(v) is used for dominated convergence arguments.

Lemma 11.2. *Assumption* C5* *implies that Assumption* C5 *holds with*

$$K_n(\theta;\gamma^*) = n^{-1}\sum_{i=1}^n \int_{\mathcal{W}} m(w,\theta) f_{\beta,W_i}(w;\gamma^*)' d\mu(w) \text{ and}$$
$$K(\theta;\gamma^*) = \int_{\mathcal{W}} m(w,\theta) f_{\beta,W}(w;\gamma^*)' d\mu(w).$$

In the ARMA(1, 1) and nonlinear regression models, Assumption C5 can be verified directly without imposing Assumption C5*, see Appendices C and E.

11.3. Assumption C6

Using Assumption C1(iii), the quantities $\xi(\pi;\gamma_0,b)$ and $\eta(\pi;\gamma_0,\omega_0)$ in Assumptions C6 and C7 can be simplified, which makes the verification of Assumption C6 easier. Specifically, Assumptions C1(iii) and C2 imply that $m(W_i,\theta)$ can be partitioned as $(m_1(W_i,\theta)', m_2(W_i,\theta)')'$, where $m_2(W_i,\theta) \in R^{d_\zeta}$ does not depend on π when $\beta = 0$. In consequence, we can partition the following quantities and obtain certain sub-quantities that do not depend on π:

$$H(\pi;\gamma_0) = \begin{bmatrix} H_{11}(\pi) & H_{12}(\pi) \\ H_{21}(\pi) & H_{22} \end{bmatrix},\ G(\pi;\gamma_0) = \begin{pmatrix} G_1(\pi) \\ G_2 \end{pmatrix},\ \text{and } K(\pi;\gamma_0) = \begin{pmatrix} K_1(\pi) \\ K_2 \end{pmatrix}, \tag{11.1}$$

where H_{22}, G_2, and K_2 do not depend on π, $H_{11}(\pi) \in R^{d_\beta \times d_\beta}$, $H_{22} \in R^{d_\zeta \times d_\zeta}$, $G_1(\pi) \in R^{d_\beta}$, $G_2 \in R^{d_\zeta}$, $K_1(\pi) \in R^{d_\beta \times d_\beta}$, and $K_2 \in R^{d_\zeta \times d_\beta}$. Define

$$\begin{aligned}
G_1^*(\pi;\gamma_0) &= G_1(\pi) - H_{12}(\pi)H_{22}^{-1}G_2, \\
K_1^*(\pi;\gamma_0) &= K_1(\pi) - H_{12}(\pi)H_{22}^{-1}K_2, \\
H_{11}^*(\pi;\gamma_0) &= H_{11}(\pi) - H_{12}(\pi)H_{22}^{-1}H_{12}(\pi)', \\
\xi_1(\pi;\gamma_0,b) &= -\frac{1}{2}(G_1^*(\pi;\gamma_0) + K_1^*(\pi;\gamma_0)b)'H_{11}^*(\pi;\gamma_0)^{-1}(G_1^*(\pi;\gamma_0) + K_1^*(\pi;\gamma_0)b), \\
\xi_2(\gamma_0,b) &= -\frac{1}{2}(G_2 + K_2 b)'H_{22}^{-1}(G_2 + K_2 b), \\
\eta_1(\pi;\gamma_0,\omega_0) &= -\frac{1}{2}\omega_0' K_1^*(\pi;\gamma_0)' H_{11}^*(\pi;\gamma_0)^{-1} K_1^*(\pi;\gamma_0)\,\omega_0, \text{ and} \\
\eta_2(\gamma_0,\omega_0) &= -\frac{1}{2}\omega_0' K_2' H_{22}^{-1} K_2 \omega_0.
\end{aligned} \tag{11.2}$$

Lemma 11.3. *Suppose Assumptions* C1(iii) *and* C2-C5 *hold. Then,*
(a) $\xi(\pi;\gamma_0,b) = \xi_1(\pi;\gamma_0,b) + \xi_2(\gamma_0,b)$ *and*
(b) $\eta(\pi;\gamma_0,\omega_0) = \eta_1(\pi;\gamma_0,\omega_0) + \eta_2(\gamma_0,\omega_0)$.

Comment. By Lemma 11.3, Assumptions C6 and C7 hold if and only if they hold with $\xi_1(\pi;\gamma_0,b)$ and $\eta_1(\pi;\gamma_0,\omega_0)$ in place of $\xi(\pi;\gamma_0,b)$ and $\eta(\pi;\gamma_0,\omega_0)$, respectively, because $\xi_2(\gamma_0,b)$ and $\eta_2(\gamma_0,\omega_0)$ do not depend on π. The quantities $\xi_1(\pi;\gamma_0,b)$ and $\eta_1(\pi;\gamma_0,\omega_0)$ are simpler than $\xi(\pi;\gamma_0,b)$ and $\eta(\pi;\gamma_0,\omega_0)$, because they are based on lower dimensional vectors, i.e., the d_β-vectors $G_1^*(\pi;\gamma_0) + K_1^*(\pi;\gamma_0)b$ and $K_1^*(\pi;\gamma_0)\omega_0$.

Using Lemma 11.3 and a generalization of Lemma 2.6 of Kim and Pollard (1990) (KP) (see Lemma 12.6 below), we obtain the following sufficient condition for Assumption C6 when β is a scalar.[46]

Assumption C6*. (i) $d_\beta = 1$ (i.e., β is a scalar).
(ii) $Var(G_1^*(\pi_1;\gamma_0) - G_1^*(\pi_2;\gamma_0)) \neq 0$ and $Var(G_1^*(\pi_1;\gamma_0) + G_1^*(\pi_2;\gamma_0)) \neq 0$, $\forall \pi_1, \pi_2 \in \Pi$ with $\pi_1 \neq \pi_2$, $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$.

Lemma 11.4. *Assumption* C6* *implies Assumption* C6.

Note that the proof of Lemma 4.1, which is stated in Section 4.4, is given after the proof of Lemma 11.4 in Appendix B below.

11.4. Assumptions C1 and D1: Quadratic Expansions for Sample Average Criterion Functions

The sample criterion function for sample average extremum estimators takes the form:

$$Q_n(\theta) = n^{-1} \sum_{i=1}^{n} \rho(W_i, \theta). \tag{11.3}$$

For example, $\rho(W_i, \theta)$ is the log-likelihood function of the ith observation in the case of the ML estimator, $\rho(W_i, \theta)$ is the squared regression residual in the case of the LS estimator, and $\rho(W_i, \theta)$ is the check function in the case of the quantile regression estimator.

For $Q_n(\theta)$ as in (11.3), $Q(\theta; \gamma_0) = E_{\gamma_0} \rho(W_i, \theta)$.

11.4.1. Sufficient Conditions via Smoothness

First, we provide sufficient conditions for Assumptions C1 and D1 when $\rho(W_i, \theta)$ is twice continuously differentiable in θ on the support of W_i. Let $\rho_\psi(W_i, \theta)$ and $\rho_{\psi\psi}(W_i, \theta)$ denote the first-order and second-order partial derivatives wrt ψ and $\rho_\theta(W_i, \theta)$ and $\rho_{\theta\theta}(W_i, \theta)$ denote the first-order and second-order partial derivatives wrt θ. The support of W_i for all $\gamma \in \Gamma$ is contained in a set $\mathcal{W}$.

Assumption Q1. (i) For some function $\rho(w, \theta) \in R$, $Q_n(\theta) = n^{-1} \sum_{i=1}^n \rho(W_i, \theta)$.
(ii) $\rho(w, \theta)$ is twice continuously differentiable in θ on an open set containing Θ^* $\forall w \in \mathcal{W}$.

[46] Kim and Pollard (1990, Lem 2.6) provides conditions under which the sample paths of a Gaussian process are maximized at a unique point with probability one.

(iii) Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, for all constants $\delta_n \to 0$,
$\sup_{\psi \in \Psi(\pi): ||\psi - \psi_{0,n}|| \leq \delta_n} ||n^{-1} \sum_{i=1}^n \left(\rho_{\psi\psi}(W_i, \psi, \pi) - \rho_{\psi\psi}(W_i, \psi_{0,n}, \pi)\right)|| = o_{p\pi}(1)$.
(iv) Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, for all constants $\delta_n \to 0$,
$\sup_{\theta \in \Theta_n(\delta_n)} ||n^{-1} \sum_{i=1}^n B^{-1}(\beta_n) \left[\rho_{\theta\theta}(W_i, \theta) - \rho_{\theta\theta}(W_i, \theta_n)\right] B^{-1}(\beta_n)|| = o_p(1)$, where
$\Theta_n(\delta_n) = \{\theta \in \Theta : ||\psi - \psi_n|| \leq \delta_n ||\beta_n|| \text{ and } ||\pi - \pi_n|| \leq \delta_n\}$.

Assumption Q1(iii) can be verified by a uniform LLN, e.g., see Andrews (1992). Assumption Q1(iv) is stronger than the stochastic equicontinuity of $n^{-1} \sum_{i=1}^n \rho_{\theta\theta}(W_i, \theta)$ over $\theta \in \Theta_n(\delta_n)$ because part of the re-scaling matrix $B^{-1}(\beta_n)$ diverges to infinity as $\beta_n \to 0$. The verification of Assumption Q1(iv) relies on the fact that $n^{-1} \sum_{i=1}^n \rho_{\theta\theta}(W_i, \theta)$ is close to singularity for $\theta \in \Theta_n(\delta_n)$.

Lemma 11.5. *Suppose Assumptions* B1-B2 *hold.*
(a) *Assumption* Q1 *implies that Assumption* C1 *holds with*

$$D_\psi Q_n(\theta) = n^{-1} \sum_{i=1}^n \rho_\psi(W_i, \theta) \text{ and } D_{\psi\psi} Q_n(\theta) = n^{-1} \sum_{i=1}^n \rho_{\psi\psi}(W_i, \theta).$$

(b) *Assumption* Q1 *implies that Assumption* D1 *holds with*

$$DQ_n(\theta) = n^{-1} \sum_{i=1}^n \rho_\theta(W_i, \theta) \text{ and } D^2 Q_n(\theta) = n^{-1} \sum_{i=1}^n \rho_{\theta\theta}(W_i, \theta).$$

11.4.2. Sufficient Conditions via Stochastic Differentiability

Next, we provide sufficient conditions for Assumptions C1 and D1 that do not require point-wise smoothness of $\rho(w, \theta)$ in θ $\forall w \in \mathcal{W}$. These sufficient conditions rely on stochastic differentiability of $Q_n(\theta)$, as in Pollard (1985), van der Vaart and Wellner (1996, Theorem 3.2.16), and Andrews (2001), and on the smoothness of $E\rho(W_i, \theta)$. These sufficient conditions cover quantile regression estimators, censored and truncated regression estimators, Huber regression M-estimators, etc.

To provide sufficient conditions via stochastic differentiability, we first define the stochastic derivative vectors and the associated remainder terms. Let

$$\rho(w, \theta) = \rho(w, \theta_n) + \Delta(w, \theta_n)'(\theta - \theta_n) + r(w, \theta), \qquad (11.4)$$

where $\Delta(w, \theta_n)$ is a "stochastic derivative" wrt θ at θ_n and $r(w, \theta)$ is the remainder term.

Compared with Pollard (1985), the current definition of the remainder term does not have $||\theta - \theta_n||$ in front of $r(w,\theta)$ in order to adapt to the weak-identification situation. The conditions on $r(w,\theta)$ given in Assumption Q2 below are adjusted accordingly.

Similarly, for any $\pi \in \Pi$, let

$$\rho(w,\psi,\pi) = \rho(w,\psi_{0,n},\pi) + \Delta_\psi(w,\psi_{0,n},\pi)'(\psi - \psi_{0,n}) + r_\psi(w,\psi,\pi), \tag{11.5}$$

where $\Delta_\psi(w,\psi_{0,n},\pi)$ is a "stochastic partial derivative" wrt ψ at $\psi_{0,n}$ and $r_\psi(w,\psi,\pi)$ is the remainder term. Note that $\Delta_\psi(w,\psi_{0,n},\pi)$ is a sub-vector of $\Delta(w,\theta)$ evaluated at $\theta = (\psi_{0,n},\pi)$. (The quantities $\Delta_\psi(w,\psi_{0,n},\pi)$ and $r_\psi(w,\psi,\pi)$ in (11.5) are not derivatives of $\Delta(w,\theta_n)$ and $r(w,\theta)$ that appear in (11.4).)

For $\{\gamma_n\} \in \Gamma(\gamma_0)$, define the empirical processes $\{\nu_n r(\theta) : \theta \in \Theta\}$ by

$$\nu_n r\,(\theta) = n^{-1/2} \sum_{i=1}^{n} (r(W_i,\theta) - E_{\gamma_n} r(W_i,\theta)), \tag{11.6}$$

where $r(w,\theta)$ is defined in (11.4). Also, define the empirical process $\{\nu_n r_\psi(\theta) : \theta \in \Theta\}$, where $\nu_n r\,(\theta) = (\nu_n r_\psi\,(\theta)', \nu_n r_\pi\,(\theta)')'$ and $r_\psi\,(w,\theta)$ is defined in (11.5).

For $\{\gamma_n\} \in \Gamma(\gamma_0)$, define the non-random real-valued function

$$Q_n^*(\theta) = n^{-1} \sum_{i=1}^{n} E_{\gamma_n} \rho(W_i,\theta). \tag{11.7}$$

When $\{W_i : 1 \le i \le n\}$ are identically distributed under γ_n, $Q_n^*(\theta) = E_{\gamma_n} \rho(W_i,\theta)$.

Assumption Q2. (i) For some function $\rho\,(w,\theta) \in R$, $Q_n(\theta) = n^{-1} \sum_{i=1}^{n} \rho\,(W_i,\theta)$.
(ii) $E_{\gamma^*} \rho(W_i,\theta)$ is twice continuously differentiable in θ on an open set containing Θ^* $\forall \gamma^* \in \Gamma$.
(iii) Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, for all constants $\delta_n \to 0$,

$$\sup_{\psi \in \Psi(\pi) : ||\psi - \psi_{0,n}|| \le \delta_n} \frac{a_n(\gamma_n) n^{-1/2} \left|\nu_n r_\psi(\psi,\pi)\right|}{[1 + ||a_n(\gamma_n)(\psi - \psi_{0,n})||] \cdot ||\psi - \psi_{0,n}||} = o_{p\pi}(1).$$

(iv) Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, for all constants $\delta_n \to 0$,

$$\sup_{\theta \in \Theta_n(\delta_n)} \frac{\left|\nu_n r\,(\theta)\,\right|}{[1 + n^{1/2}||B(\beta_n)(\theta - \theta_n)||] \cdot ||B(\beta_n)(\theta - \theta_n)||} = o_p\,(1)\,,$$

where $\Theta_n(\delta_n) = \{\theta \in \Theta : ||\psi - \psi_n|| \leq \delta_n ||\beta_n|| \text{ and } ||\pi - \pi_n|| \leq \delta_n\}$.
(v) Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, for all constants $\delta_n \to 0$,

$$\sup_{\psi \in \Psi(\pi): ||\psi - \psi_{0,n}|| \leq \delta_n} ||\frac{\partial^2}{\partial \psi \partial \psi'} Q_n^*(\psi, \pi) - \frac{\partial^2}{\partial \psi \partial \psi'} Q_n^*(\psi_{0,n}, \pi)|| = o_\pi(1).$$

(vi) Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, for all constants $\delta_n \to 0$,

$$\sup_{\theta \in \Theta_n(\delta_n)} ||B^{-1}(\beta_n)[\frac{\partial^2}{\partial \theta \partial \theta'} Q_n^*(\theta) - \frac{\partial^2}{\partial \theta \partial \theta'} Q_n^*(\theta_n)] B^{-1}(\beta_n)|| = o(1).$$

Because the expectation operator is a smoothing operator, $E_{\gamma^*}\rho(W_i, \theta)$ often is differentiable in θ even though $\rho(W_i, \theta)$ is not. For example, Assumption Q2(ii) holds when $\rho(W_i, \theta)$ is piece-wise differentiable in θ and is only non-smooth in θ on a negligible set of $\{W_i : 1 \leq i \leq n\}$. Such cases include quantile regression, censored and truncated regression models, etc.

Assumptions Q2(iii) and Q2(iv) are generalizations of the stochastic differentiability condition in Pollard (1985) to the case of drifting sequences of true parameters. In the special case where $\rho(W_i, \theta)$ is twice continuously differentiable, Assumptions Q2(iii) and Q2(iv) can be verified easily by omitting the "1" part in the denominators. The verification is similar to that in Lemma 11.5 above.

When $\rho(W_i, \theta)$ is not point-wise smooth, Assumptions Q2(iii) and Q2(iv) can be verified by methods provided in Pollard (1985). For example, empirical process methods can be used to show $\nu_n r_\psi(\psi, \pi)/||\psi - \psi_{0,n}|| = o_{p\pi}(1)$ uniformly for ψ in a neighborhood of $\psi_{0,n}$ to verify Assumption Q2(iii). In this case, only the "$||\psi - \psi_{0,n}||$" part of the denominator in Assumption Q2(iii) is used. Similarly, empirical process methods can be used to show $\nu_n r(\theta) / ||B(\beta_n)(\theta - \theta_n)|| = o_p(1)$ uniformly over $\Theta_n(\delta_n)$ to verify Assumption Q2(iv). Pollard (1985) provides results for empirical processes based on i.i.d. random variables. For dependent random variables, the empirical process results in Doukhan, Massart, and Rio (1995) and Arcones and Yu (1994) can be used. Hansen (1996) establishes stochastic equicontinuity of empirical process of dependent triangular arrays, which is suitable for asymptotic results under drifting sequences of true parameters. For other references, see Andrews (1994). Also, the Huber-type bracketing condition in Pollard (1985) applies with dependent random variables.

Assumption Q2(v) is not restrictive. It holds by Assumption Q2(ii) when $\{W_i : i \geq$

$1\}$ are identically distributed under $\gamma^* \in \Gamma$.

Assumption Q2(vi) is stronger than uniform continuity of $(\partial^2/\partial\theta\partial\theta')Q_n^*(\theta)$ because part of $B^{-1}(\beta_n)$ diverges when $\beta_n \to 0$. The verification of Assumption Q2(vi) relies on $(\partial^2/\partial\theta\partial\theta')Q_n^*(\theta)$ being almost singular when β is close to 0.

For $\{\gamma_n\} \in \Gamma(\gamma_0)$, define the empirical process $\{\nu_n\Delta(\theta) : \theta \in \Theta\}$ by

$$\nu_n\Delta(\theta) = n^{-1/2}\sum_{i=1}^{n}(\Delta(W_i,\theta) - E_{\gamma_n}\Delta(W_i,\theta)), \tag{11.8}$$

where $\Delta(w,\theta)$ is defined in (11.4). Also, define the empirical process $\{\nu_n\Delta_\psi(\theta) : \theta \in \Theta\}$, where $\nu_n\Delta(\theta) = (\nu_n\Delta_\psi(\theta)', \nu_n\Delta_\pi(\theta)')'$ and $\Delta_\psi(\theta)$ is as in (11.5).

Lemma 11.6. *Suppose Assumptions* B1 *and* B2 *hold.*
(a) *Assumption* Q2 *implies that Assumption* C1 *holds with*

$$D_\psi Q_n(\theta) = n^{-1/2}\nu_n\Delta_\psi(\theta) + \frac{\partial}{\partial\psi}Q_n^*(\theta) \text{ and } D_{\psi\psi}Q_n(\theta) = \frac{\partial^2}{\partial\psi\partial\psi'}Q_n^*(\theta).$$

(b) *Assumption* Q2 *implies that Assumption* D1 *holds with*

$$DQ_n(\theta) = n^{-1/2}\nu_n\Delta(\theta) + \frac{\partial}{\partial\theta}Q_n^*(\theta) \text{ and } D^2Q_n(\theta) = \frac{\partial^2}{\partial\theta\partial\theta'}Q_n^*(\theta).$$

Comments. 1. When $Q_n^*(\theta)$ is minimized at θ_n under $\{\gamma_n\} \in \Gamma(\gamma_0)$, $DQ_n(\theta)$ in Lemma 11.6(b) evaluated at $\theta = \theta_n$ simplifies to $n^{-1/2}\nu_n\Delta(\theta_n)$ because $(\partial/\partial\theta)Q_n^*(\theta_n) = 0$. With identically distributed observations, this holds under Assumption B3 because $Q_n^*(\theta) = E_{\gamma_n}\rho(W_i,\theta)$ is minimized at $\theta = \theta_n$. In Assumption C1, $D_\psi Q_n(\theta)$ is evaluated at $\theta = (\psi_{0,n}, \pi)$. The expression for $D_\psi Q_n(\theta)$ in Lemma 11.6(a) does not simplify when $\theta = (\psi_{0,n}, \pi)$ because $Q_n^*(\theta)$ is not minimized at $(\psi_{0,n}, \pi)$ under γ_n.

2. In Lemma 11.6, $D_{\psi\psi}Q_n(\theta)$ and $D^2Q_n(\theta)$ are both non-random. With identically distributed observations, $D_{\psi\psi}Q_n(\theta)$ and $D^2Q_n(\theta)$ are second-order partial derivatives of $E_{\gamma_n}\rho(W_i,\theta)$ wrt ψ and θ, respectively.

Under Assumptions B1, B2, and Q2, Assumption C2(i) holds with

$$m(W_i,\theta) = \Delta_\psi(W_i,\theta) - E_{\gamma^*}\Delta_\psi(W_i,\theta) + \frac{\partial}{\partial\psi}E_{\gamma^*}\rho(W_i,\theta). \tag{11.9}$$

Hence, $E_{\gamma^*}m(W_i,\theta) = (\partial/\partial\psi)E_{\gamma^*}\rho(W_i,\theta)$. Assumption C2(ii) holds provided $E_{\gamma^*}\rho(W_i,\theta)$

is minimized at θ^* when the true parameter is $\gamma^* \in \Gamma$, and Assumption C2(iii) holds provided $E_{\gamma^*}\rho(W_i, \theta)$ is minimized at (ψ^*, π) $\forall \pi \in \Pi$ when the true parameter is $\gamma^* \in \Gamma$ with $\beta^* = 0$. With identically distributed observations, Assumptions C2(ii) and C2(iii) are implied by Assumptions B3 and Q2(ii) with $E_{\gamma^*}\rho(W_i, \theta) = Q(\theta; \gamma^*)$.

Assumption C3 can be verified with $G_n(\pi) = \nu_n \Delta_\psi(\psi_{0,n}, \pi)$. Assumption C4(i) holds with $H(\pi; \gamma_0) = \lim_{n\to\infty}(\partial^2/\partial\psi\partial\psi')Q_n^*(\psi_0, \pi)$ provided this limit exists, which is always true for identically distributed observations. The verification of Assumption C5 requires regularity conditions on the density functions of the observations wrt some dominating measure for $\gamma \in \Gamma$. Assumption C6 can be verified using Lemma 4.1 or 11.4. Assumption C7 can be verified using the matrix Cauchy-Schwarz inequality, see Tripathi (1999). Assumption C8 is implied by Assumption C4 because $(\partial/\partial\psi')E_{\gamma_n}D_\psi Q_n(\theta) = D_{\psi\psi}Q_n(\theta)$.

Assumption D2 can be verified directly with the non-random form of $D^2Q_n(\theta_n)$ given in Lemma 11.6(b). Assumption D3 can be verified by a triangular array CLT provided $Q_n^*(\theta)$ is minimized at θ_n $\forall n \geq 1$. The latter condition yields $DQ_n(\theta_n) = n^{-1/2}\nu_n\Delta_\psi(\theta_n)$.

11.4.3. Initial Conditions Adjustment to the Sample Criterion Function

In some stationary time series models, the sample criterion function $Q_n(\theta)$ depends on initial conditions and, hence, is not an average of stationary and ergodic random variables. In such cases, Assumptions Q1 and Q2 can be adjusted to allow $Q_n(\theta)$ to equal a sample average of stationary summands, $n^{-1}\sum_{i=1}^n \rho(W_i, \theta)$, plus a term, $Q_n^{IC}(\theta)$, that is asymptotically negligible in a suitable sense. A similar adjustment was introduced in Andrews (2001).

Assumption Q3. (i) For some function $\rho(w, \theta) \in R$, $Q_n(\theta) = n^{-1}\sum_{i=1}^n \rho(W_i, \theta) + Q_n^{IC}(\theta)$.
(ii) Assumption C1(ii) holds with $R_n(\theta)$ replaced by $Q_n^{IC}(\theta) - Q_n^{IC}(\psi_{0,n}, \pi)$ and Assumption D1(ii) holds with $R_n^*(\theta)$ replaced by $Q_n^{IC}(\theta) - Q_n^{IC}(\theta_n)$.

Lemma 11.7. (a) *Lemma* 11.5 *holds with Assumption* Q1(i) *replaced by Assumption* Q3.
(b) *Lemma* 11.6 *holds with Assumption* Q2(i) *replaced by Assumption* Q3.

12. Appendix B: Proofs

This Appendix contains proofs of (i) the estimation results of AC1, (ii) the results of AC1 for t CS's and tests, and (iii) the sufficient conditions given in Appendix A.

12.1. Proofs of Estimation Results

Proof of Lemma 5.1. The first result of Lemma 5.1(a) is proved along the lines of the proof of Lemma A1 of Andrews (1993), which is a uniform consistency result under fixed true parameters. Specifically, by Assumption B3(ii), given any neighborhood Ψ_0 of ψ_0, there exists a constant $\varepsilon > 0$ such that $\forall \pi \in \Pi$, $\inf_{\psi \in \Psi(\pi)/\Psi_0} Q(\psi, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0) \geq \varepsilon$. Thus,

$$
\begin{aligned}
& P\left(\widehat{\psi}_n(\pi) \in \Psi(\pi)/\Psi_0 \text{ for some } \pi \in \Pi\right) \\
& \leq P\left(Q(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0) \geq \varepsilon \text{ for some } \pi \in \Pi\right) \to 0,
\end{aligned} \tag{12.1}
$$

where "$\to 0$" holds provided $\sup_{\pi \in \Pi} |Q(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)| \to_p 0$. The latter follows from

$$
\begin{aligned}
0 & \leq \inf_{\pi \in \Pi} \left[Q(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)\right] \\
& \leq \sup_{\pi \in \Pi} \left[Q(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)\right] \\
& \leq \sup_{\pi \in \Pi} \left[Q(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q_n(\widehat{\psi}_n(\pi), \pi; \gamma_0)\right] + \sup_{\pi \in \Pi} \left[Q_n(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)\right] \\
& \leq \sup_{\pi \in \Pi} \left[Q(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q_n(\widehat{\psi}_n(\pi), \pi; \gamma_0)\right] + \sup_{\pi \in \Pi} \left[Q_n(\psi_0, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)\right] + o(n^{-1}) \\
& \leq 2 \sup_{\psi \in \Psi(\pi), \pi \in \Pi} |Q_n(\psi, \pi; \gamma_0) - Q(\psi, \pi; \gamma_0)| + o(n^{-1}) = o_p(1),
\end{aligned} \tag{12.2}
$$

where the first inequality holds by Assumption B3(ii) and the fourth inequality holds by the definition of $\widehat{\psi}_n(\pi)$ in (5.1), and the equality holds by Assumption B3(i). This completes the proof of the first result of part (a). The second result of part (a) follows from the first result because $\widehat{\psi}_n = \widehat{\psi}_n(\widehat{\pi}_n)$ and $\widehat{\pi}_n \in \Pi$.

When $\beta_0 \neq 0$, $\widehat{\theta}_n \to_p \theta_0$ under $\{\gamma_n\}$ such that $\gamma_n \to \gamma_0$ with $\beta_0 \neq 0$ by an analogous argument to that just given for part (a), but with $\widehat{\theta}_n$, θ_0, and Θ/Θ_0, in place of $(\widehat{\psi}_n(\pi), \pi)$, (ψ_0, π), and $\Psi(\pi)/\Psi_0$, respectively, where Θ_0 is some neighborhood of θ_0, with $\inf_{\pi \in \Pi}$

and $\sup_{\pi \in \Pi}$ deleted, and with Assumption B3(iii) used in place of Assumption B3(ii). Because $\theta_n \to \theta_0$, this completes the proof of part (b). $\square$

The following two Lemmas are used in the proofs of Lemma 5.2 and Theorem 5.1.

Lemma 12.1. *Suppose Assumptions* B1, B2, C2, C3, *and* C5 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$,
(a) *when* $||b|| < \infty$, $n^{1/2} D_\psi Q_n(\psi_{0,n}, \cdot) \Rightarrow G(\cdot; \gamma_0) + K(\cdot; \gamma_0) b$, *and*
(b) *when* $||b|| = \infty$ *and* $\beta_n / ||\beta_n|| \to \omega_0$ *for any* $\omega_0 \in R^{d_\beta}$ *with* $||\omega_0|| = 1$,
$||\beta_n||^{-1} D_\psi Q_n(\psi_{0,n}, \pi) \to_p K(\pi; \gamma_0) \omega_0$ *uniformly over* $\pi \in \Pi$.

Comment. Lemma 12.1 implies that $a_n(\gamma_n) D_\psi Q_n(\psi_{0,n}, \pi) = O_{p\pi}(1)$.

Define

$$Z_n(\pi) = -a_n(\gamma_n)(D_{\psi\psi} Q_n(\psi_{0,n}, \pi))^{-1} D_\psi Q_n(\psi_{0,n}, \pi). \tag{12.3}$$

Lemma 12.2. *Suppose Assumptions* A, B1-B3, *and* C1-C5 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$,
(a) $a_n(\gamma_n)(\widehat{\psi}_n(\pi) - \psi_{0,n}) = O_{p\pi}(1)$,
(b) $a_n(\gamma_n)(\widehat{\psi}_n(\pi) - \psi_{0,n}) = Z_n(\pi) + o_{p\pi}(1)$, *and*
(c) $a_n^2(\gamma_n)\left(Q_n(\widehat{\psi}_n(\pi), \pi) - Q_n(\psi_{0,n}, \pi)\right) = -\frac{1}{2} Z_n(\pi)' D_{\psi\psi} Q_n(\psi_{0,n}, \pi) Z_n(\pi) + o_{p\pi}(1)$.

Comment. When $||b|| < \infty$, Lemma 12.2(b) is used to derive the asymptotic distribution of $\widehat{\psi}_n$. Lemma 12.2(c) is used in the proof of Lemma 5.2 below.

Proof of Lemma 12.1. First, we decompose $D_\psi Q_n(\psi_{0,n}, \pi)$ as

$$D_\psi Q_n(\psi_{0,n}, \pi) = n^{-1/2} G_n(\pi) + n^{-1} \sum_{i=1}^n E_{\gamma_n} m_i(\psi_{0,n}, \pi). \tag{12.4}$$

To analyze $n^{-1} \sum_{i=1}^n E_{\gamma_n} m_i(\psi_{0,n}, \pi)$ when β_n is close to 0, we view this average expectation as a function of β_n and we carry out element-by-element mean value expansions around $\beta_n = 0$. This gives

$$n^{-1} \sum_{i=1}^n E_{\gamma_n} m_i(\psi_{0,n}, \pi) = n^{-1} \sum_{i=1}^n E_{\gamma_{0,n}} m_i(\psi_{0,n}, \pi) + K_n(\psi_{0,n}, \pi; \widetilde{\gamma}_n)\beta_n = K_n(\psi_{0,n}, \pi; \widetilde{\gamma}_n)\beta_n, \tag{12.5}$$

where $\widetilde{\gamma}_n = (\widetilde{\beta}_n, \zeta_n, \pi_n, \phi_n)$ may differ across the rows of $K_n(\psi_{0,n}, \pi; \widetilde{\gamma}_n)$, $\widetilde{\beta}_n$ is on the line segment connecting β_n and 0, which implies that $\widetilde{\beta}_n$ converges to 0 as $\gamma_n \to \gamma_0$ with $\beta_0 = 0$, and the second equality holds by Assumption C2(iii) applied with $\gamma^* = \gamma_{0,n}$ because $\gamma_n = (\beta_n, \zeta_n, \pi_n, \phi_n) \in \Gamma$ with $||\beta_n|| < \delta$, which holds for n large, implies that $\gamma_{0,n} = (0, \zeta_n, \pi_n, \phi_n) \in \Gamma$ by Assumption B2(iii). Furthermore, $(\psi_{0,n}, \pi, \widetilde{\gamma}_n)$ is in the domain $\Theta_{00} \times \Gamma_0$ of $K_n(\cdot;\cdot)$ by Assumption B2(iii).

By Assumption C5,

$$K_n(\psi_{0,n}, \pi; \widetilde{\gamma}_n) \to_p K(\pi; \gamma_0) \tag{12.6}$$

uniformly over $\pi \in \Pi$. From (12.4)-(12.6), we obtain

$$D_\psi Q_n(\psi_{0,n}, \pi) = n^{-1/2} G_n(\pi) + K(\pi; \gamma_0)\beta_n + o_{p\pi}(||\beta_n||). \tag{12.7}$$

In part (a), in which case $n^{1/2}\beta_n \to b$ with $||b|| < \infty$, (12.7) leads to

$$n^{1/2} D_\psi Q_n(\psi_{0,n}, \cdot) = G_n(\cdot) + K(\cdot; \gamma_0) n^{1/2}\beta_n + o_{p\pi}(1) \Rightarrow G(\cdot; \gamma_0) + K(\cdot; \gamma_0) b, \tag{12.8}$$

where the weak-convergence result holds by Assumption C3.

In part (b), in which case $n^{1/2}||\beta_n|| \to \infty$ and $\beta_n/||\beta_n|| \to \omega_0$, (12.7) leads to

$$||\beta_n||^{-1} D_\psi Q_n(\psi_{0,n}, \pi) = (n^{1/2}||\beta_n||)^{-1} G_n(\pi) + K(\pi; \gamma_0)\beta_n/||\beta_n|| + o_{p\pi}(1) \to_p K(\pi; \gamma_0)\omega_0 \tag{12.9}$$

uniformly over $\pi \in \Pi$ using Assumption C3. $\square$

Proof of Lemma 12.2. The proof of part (a) is analogous to the proof of Theorem 1 of Andrews (1999), which in turn uses the method in Chernoff (1954, Lemma 1). For notational simplicity, $D_{\psi\psi} Q_n(\psi_{0,n}, \pi)$ is abbreviated as $D_{\psi\psi,n}(\pi)$. Let $\kappa_{n,\pi} = D^{1/2}_{\psi\psi,n}(\pi) a_n(\gamma_n)(\widehat{\psi}_n(\pi) - \psi_{0,n})$. We have

$$\begin{aligned} o_{p\pi}(1) &\geq a_n^2(\gamma_n)\left(Q_n(\widehat{\psi}_n(\pi), \pi) - Q_n(\psi_{0,n}, \pi)\right) \\ &= a_n(\gamma_n) D_\psi Q_n(\psi_{0,n}, \pi)' D^{-1/2}_{\psi\psi,n}(\pi)\kappa_{n,\pi} + \frac{1}{2}\left\|\kappa_{n,\pi}\right\|^2 + a_n^2(\gamma_n) R_n(\widehat{\psi}_n(\pi), \pi) \\ &= O_{p\pi}(\|\kappa_{n,\pi}\|) + \frac{1}{2}\left\|\kappa_{n,\pi}\right\|^2 + \left(1 + \left\|D^{-1/2}_{\psi\psi,n}(\pi)\kappa_{n,\pi}\right\|\right)^2 o_{p\pi}(1) \\ &= O_{p\pi}(\|\kappa_{n,\pi}\|) + \frac{1}{2}\left\|\kappa_{n,\pi}\right\|^2 + o_{p\pi}(\|\kappa_{n,\pi}\|) + o_{p\pi}\left(\|\kappa_{n,\pi}\|^2\right) + o_{p\pi}(1), \end{aligned} \tag{12.10}$$

where the inequality holds $\forall \pi \in \Pi$ for n large by (5.1) and the fact that $\psi_{0,n} \in \Psi(\pi)$

$\forall\pi\in\Pi$ for n large, which holds because this condition is equivalent to $(\psi_{0,n},\pi)\in\Theta$ $\forall\pi\in\Pi$ for n large and the latter holds because (i) $(\psi_{0,n},\pi)=(0,\zeta_n,\pi)\in\{\beta\in R^{d_\beta}:||\beta||<\delta\}\times\mathcal{Z}^0\times\Pi\subset\Theta$ $\forall\pi\in\Pi$ by Assumption B1(ii) provided $\zeta_n\in\mathcal{Z}^0$, and (ii) $\zeta_n\in\mathcal{Z}^0$ for n large by Assumption B1(ii) because $\theta_n=(\beta_n,\zeta_n,\pi_n)\to\theta_0=(0,\zeta_0,\pi_0)$ implies that $||\beta_n||<\delta$, and $\theta_n\in\Theta_\delta^*\subset\{\beta\in R^{d_\beta}:||\beta||<\delta\}\times\mathcal{Z}^0\times\Pi$ for n large. The first equality in (12.10) holds by Assumption C1(i) with $\psi=\widehat{\psi}_n(\pi)$, and the second equality holds by Lemma 5.1(a), Assumptions C1(ii) and C4, and the implication of Lemma 12.1 that $a_n(\gamma_n)D_\psi Q_n(\psi_{0,n},\pi)=O_{p\pi}(1)$. Rearranging (12.10) gives $\|\kappa_{n,\pi}\|^2\leq 2||\kappa_{n,\pi}||O_{p\pi}(1)+o_{p\pi}(1)$. Let $\xi_{n,\pi}$ denote the $O_{p\pi}(1)$ term. Then, we have

$$(||\kappa_{n,\pi}||-\xi_{n,\pi})^2\leq\xi_{n,\pi}^2+o_{p\pi}(1). \tag{12.11}$$

Taking square roots gives $||\kappa_{n,\pi}||=O_{p\pi}(1)$, which together with Assumption C4 completes the proof of part (a).

Now, we prove part (b). Define

$$\Delta_n(\pi)=a_n(\gamma_n)(\widehat{\psi}_n(\pi)-\psi_{0,n})\text{ and }\psi_n^\dagger(\pi)=\psi_{0,n}+a_n^{-1}(\gamma_n)Z_n(\pi). \tag{12.12}$$

First, we apply the quadratic approximation in Assumption C1(i) with $\psi=\psi_n^\dagger(\pi)$. Re-scaling both sides by $a_n^2(\gamma_n)$, we get

$$a_n^2(\gamma_n)\left(Q_n(\psi_n^\dagger(\pi),\pi)-Q_n(\psi_{0,n},\pi)\right)=-\frac{1}{2}Z_n(\pi)'D_{\psi\psi,n}(\pi)Z_n(\pi)+o_{p\pi}(1), \tag{12.13}$$

where the $o_{p\pi}(1)$ term is obtained from Assumption C1(ii), Lemma 12.1, and $\psi_{0,n}-\psi_n\to 0$.

Next, we apply the quadratic approximation in Assumption C1(i) with $\psi=\widehat{\psi}_n(\pi)$ to obtain

$$\begin{aligned}&a_n^2(\gamma_n)\left(Q_n(\widehat{\psi}_n(\pi),\pi)-Q_n(\psi_{0,n},\pi)\right)\\&=-Z_n(\pi)'D_{\psi\psi,n}(\pi)\Delta_n(\pi)+\frac{1}{2}\Delta_n(\pi)'D_{\psi\psi,n}(\pi)\Delta_n(\pi)+o_{p\pi}(1)\\&=\frac{1}{2}(\Delta_n(\pi)-Z_n(\pi))'D_{\psi\psi,n}(\pi)(\Delta_n(\pi)-Z_n(\pi))\\&\quad-\frac{1}{2}Z_n(\pi)'D_{\psi\psi,n}(\pi)Z_n(\pi)+o_{p\pi}(1),\end{aligned} \tag{12.14}$$

where the $o_{p\pi}(1)$ term in the first equality is obtained from Assumption C1(ii) and

Lemma 12.2(a).

We can write $a_n^{-1}(\gamma_n)Z_n(\pi) = (\beta_n^\dagger(\pi), \zeta_n^{\dagger\dagger}(\pi))$, where $\beta_n^\dagger(\pi) = o_{p\pi}(1)$ and $\zeta_n^{\dagger\dagger}(\pi) = o_{p\pi}(1)$ using Assumptions C3 and C4 and $a_n^{-1}(\gamma_n) \leq n^{-1/2} \to 0$. This and Assumption B1(ii) lead to

$$\psi_n^\dagger(\pi) = (0, \zeta_n) + (\beta_n^\dagger(\pi), \zeta_n^{\dagger\dagger}(\pi)) \in \Psi(\pi) \tag{12.15}$$

$\forall \pi \in \Pi$, where "$\in$" holds with probability that goes to one as $n \to \infty$. Specifically, (12.15) holds because (i) $\gamma_n \to \gamma_0$ with $\beta_0 = 0$, (ii) for n large, $(\beta_n, \zeta_n, \pi_n, \phi_n) \in \Gamma$ satisfies $||\beta_n|| < \delta/2$ and $||\zeta_n - \zeta_0|| < \delta_{\zeta_0}/2$ for some $\delta > 0$ and $\delta_{\zeta_0} > 0$ chosen such that the ball centered at ζ_0 with radius δ_{ζ_0} is in $\mathcal{Z}^0$, (iii) the latter, $\beta_n^\dagger(\pi) = o_{p\pi}(1)$, and $\zeta_n^{\dagger\dagger}(\pi) = o_{p\pi}(1)$ imply that $||\beta_n^\dagger(\pi)|| < \delta$, $||\zeta_n + \zeta_n^{\dagger\dagger}(\pi) - \zeta_0|| < \delta_{\zeta_0}$, $\zeta_n + \zeta_n^{\dagger\dagger}(\pi) \in \mathcal{Z}^0$, and $\psi_n^\dagger(\pi) \in \{\beta \in R^{d_\beta} : ||\beta|| < \delta\} \times \mathcal{Z}^0$ $\forall \pi \in \Pi$ with probability that goes to one, and (iv) $\{\beta \in R^{d_\beta} : ||\beta|| < \delta\} \times \mathcal{Z}^0 \subset \Psi(\pi) \cap \{\psi = (\beta, \zeta) \in R^{d_\psi} : ||\beta|| < \delta\}$ by Assumption B1(ii). Results (iii) and (iv) combine to establish (12.15).

Using (12.15) and (5.1), we have

$$Q_n(\widehat{\psi}_n(\pi), \pi) \leq Q_n(\psi_n^\dagger(\pi), \pi) + o_{p\pi}(n^{-1}) \tag{12.16}$$

$\forall \pi \in \Pi$. This, (12.13), and (12.14) give

$$\frac{1}{2}(\Delta_n(\pi) - Z_n(\pi))' D_{\psi\psi,n}(\pi)(\Delta_n(\pi) - Z_n(\pi)) \leq o_{p\pi}(1). \tag{12.17}$$

Assumption C4 and (12.17) imply that $\Delta_n(\pi) = Z_n(\pi) + o_{p\pi}(1)$, which is the result of part (b).

Part (c) holds because the first summand on the right-hand side (rhs) of (12.14) is $o_{p\pi}(1)$ by Lemma 12.2(b) and Assumption C4. $\square$

Proof of Lemma 5.2. Lemma 12.1(a) and Assumption C4 yield

$$Z_n(\cdot) \Rightarrow -H^{-1}(\cdot; \gamma_0)(G(\cdot; \gamma_0) + K(\cdot; \gamma_0)b) \tag{12.18}$$

under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ when $||b|| < \infty$. Lemma 12.1(b) and Assumption C4 yield

$$Z_n(\pi) \to_p -H^{-1}(\pi; \gamma_0)K(\pi; \gamma_0)\omega_0 \tag{12.19}$$

uniformly over $\pi \in \Pi$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ when $||b|| = \infty$ and $\beta_n/||\beta_n|| \to \omega_0$.

The result of part (a) holds by Lemma 12.2(c), (12.18), Assumption C4, and the CMT. Replacing (12.18) with (12.19) gives the result of part (b). □

Proof of Theorem 5.1. First we prove part (a). We have $\widehat{\pi}_n \rightarrow_d \pi^*(\gamma_0, b)$ by (5.2), Lemma 5.2(a), Assumptions A, B1(iii), C3, C4(i), C5(iii), and C6, and the CMT. For details, see the proof of the argmax/min Theorem 3.2.2 in van der Vaart and Wellner (1996, p. 286). Note that Assumptions C3, C4, and C5(iii) are used to guarantee that $\xi(\pi; \gamma_0, b)$ is continuous on Π a.s. and Assumption B1(iii) guarantees that the sequence of distributions of $\{\widehat{\pi}_n\}$ is tight.

Define $\tau_n(\pi) = n^{1/2}(\widehat{\psi}_n(\pi) - \psi_n)$. We have

$$\begin{aligned} \tau_n(\cdot) &= n^{1/2}(\widehat{\psi}_n(\cdot) - \psi_{0,n}) - n^{1/2}(\psi_n - \psi_{0,n}) \\ &= Z_n(\cdot) - (n^{1/2}\beta_n, 0_{d_\zeta}) + o_{p\pi}(1) \\ &\Rightarrow -H^{-1}(\cdot; \gamma_0)\,(G(\cdot; \gamma_0) + K(\cdot; \gamma_0)\, b) - (b, 0_{d_\zeta}), \end{aligned} \tag{12.20}$$

where the second equality holds by Lemma 12.2(b) and the definition of $\psi_{0,n}$ and the weak-convergence result holds by Lemma 12.1(a) and Assumption C4. Furthermore, joint convergence $(\tau_n(\cdot), \widehat{\pi}_n) \Rightarrow (\tau(\cdot; \gamma_0, b), \pi^*(\gamma_0, b))$ holds because $\tau_n(\cdot)$ and $\widehat{\pi}_n$ are continuous functions of $Z_n(\cdot)$ and $D_{\psi\psi}Q_n(\psi_{0,n}, \cdot)$, which converge jointly since the limit of the latter, $H(\cdot; \gamma_0)$, is non-random.

To prove part (b), we write

$$Q_n(\widehat{\theta}_n) = Q_n(\widehat{\psi}_n(\widehat{\pi}_n), \widehat{\pi}_n) = Q_n^c(\widehat{\pi}_n) = \inf_{\pi \in \Pi} Q_n^c(\pi) + o(n^{-1}), \tag{12.21}$$

where the first equality holds by assumption (see the paragraph following (5.2)), the second equality holds by the definition of $Q_n^c(\pi)$ given just above (5.2), and the third equality holds by (5.2). Part (b) follows from Lemma 5.2(a), (12.21), and the CMT. □

Proof of Lemma 5.3. When $\beta_0 = 0$, $\widehat{\pi}_n \rightarrow_p \pi_0$ by a standard consistency argument, such as a simplification of the argument given in the proof of Lemma 5.1(a) with $\widehat{\pi}_n$, π_0, Π/Π_0, $||\beta_n||^{-2}(Q_n^c(\pi) - Q_{0,n})$, and $\eta(\pi; \gamma_0, \omega_0)$ in place of $(\widehat{\psi}_n(\pi), \pi)$, (ψ_0, π), $\Psi(\pi)/\Psi_0$, $Q_n(\psi, \pi; \gamma_0)$, and $Q(\psi, \pi; \gamma_0)$, respectively, where Π_0 is some neighborhood of π_0, and with $\inf_{\pi \in \Pi}$ and $\sup_{\pi \in \Pi}$ deleted. The argument uses Lemma 5.2(b) (which applies because the set of sequences $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_0 = 0$ is the same as the set of sequences $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| = \infty$ and $\beta_n/||\beta_n|| \rightarrow \omega_0$) in place of Assumption B3(i). In place of Assumption B3(ii), the argument uses the fact that $\eta(\pi;\gamma_0,\omega_0)$ is continuous on Π by Assumptions C4 and C5(iii) and is uniquely minimized at π_0 by Assumption C7, and Π is compact by Assumption B1(iii). Because $\pi_n \to \pi_0$, this completes the proof that $\widehat{\pi}_n - \pi_n \to_p 0$.

When $\beta_0 = 0$, $\widehat{\psi}_n - \psi_n \to_p 0$ because $||\widehat{\psi}_n - \psi_n|| = ||\widehat{\psi}_n(\widehat{\pi}_n) - \psi_n|| \leq \sup_{\pi\in\Pi} ||\widehat{\psi}_n(\pi) - \psi_n|| = o_p(1)$ by Lemma 5.1(a).

When $\beta_0 \neq 0$, the desired results are given in Lemma 5.1(b). $\square$

The following two Lemmas are used in the proof of Theorem 5.2. Let $S_\beta = [I_{d_\beta} : 0_{d_\beta \times d_\zeta}]$ denote the $d_\beta \times d_\psi$ selector matrix that selects β out of ψ.

Lemma 12.3. *Suppose Assumptions* C2, C4, C5, *and* C8 *hold. Then,* $K(\pi_0;\gamma_0) = -H(\pi_0;\gamma_0)S'_\beta$.

Lemma 12.4. *Suppose Assumptions* A, B1-B3, C1-C5, C7, *and* C8 *hold. Then,* $\|\beta_n\|^{-1}(\widehat{\psi}_n - \psi_n) = o_p(1)$ *under* $\{\gamma_n\} \in \Gamma(\gamma_0,\infty,\omega_0)$ *with* $\beta_0 = 0$.

Proof of Lemma 12.3. For notational simplicity, define a function

$$h^n(\gamma^*,\psi) = n^{-1}\sum_{i=1}^n E_{\gamma^*} m(W_i,\psi,\pi^*). \tag{12.22}$$

Let $h^n_{\psi^*}(\gamma^*,\psi)$ denote the partial derivative of $h^n(\gamma^*,\psi)$ wrt ψ^*, which is a sub-vector of γ^*, and let $h^n_{\psi}(\gamma^*,\psi)$ denote its partial derivative wrt ψ. By Assumption C2(ii),

$$h^n(\gamma^*,\psi^*) = 0 \ \forall \gamma^* \in \Gamma. \tag{12.23}$$

In (12.23), ψ^* enters $h^n(\gamma^*,\psi^*)$ through both γ^* and the second argument of $h^n(\cdot,\cdot)$. Taking the derivative of $h^n(\gamma^*,\psi^*)$ wrt ψ^* gives

$$h^n_{\psi^*}(\gamma^*,\psi^*) + h^n_{\psi}(\gamma^*,\psi^*) = 0 \ \forall \gamma^* \in \Gamma. \tag{12.24}$$

The definition of $h^n(\cdot,\cdot)$ in (12.22) and the equality in (12.24) yield

$$n^{-1}\sum_{i=1}^n \frac{\partial}{\partial\psi^{*\prime}} E_{\gamma^*} m_i(\psi^*,\pi^*) = h^n_{\psi^*}(\gamma^*,\psi^*) = -h^n_{\psi}(\gamma^*,\psi^*) = -n^{-1}\sum_{i=1}^n \frac{\partial}{\partial\psi'} E_{\gamma^*} m_i(\psi^*,\pi^*). \tag{12.25}$$

Post-multiplying both sides of (12.25) by S_β', which selects the first d_β columns, yields

$$n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\beta^{*\prime}}E_{\gamma^*}m_i(\psi^*,\pi^*)=\left(-n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\psi'}E_{\gamma^*}m_i(\psi^*,\pi^*)\right)S_\beta'. \tag{12.26}$$

The partial derivative $(\partial/\partial\beta^{*\prime})E_{\gamma^*}m_i(\psi^*,\pi^*)$ on the left-hand side (lhs) of (12.26) denotes the partial derivative of $E_{\gamma^*}m_i(\psi^*,\pi^*)$ wrt β^*, which is a sub-vector of the true value γ^*, whereas $(\partial/\partial\psi')E_{\gamma^*}m_i(\psi^*,\pi^*)$ on the rhs of (12.26) denotes the partial derivative wrt ψ, which is an argument of the function $m_i(\psi,\pi)$.

Under $\{\gamma_n\}\in\Gamma(\gamma_0,\infty,\omega_0)$, (12.26) with γ^* replaced by γ_n becomes

$$n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\beta^{*\prime}}E_{\gamma_n}m_i(\psi_n,\pi_n)=\left(-n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\psi'}E_{\gamma_n}m_i(\psi_n,\pi_n)\right)S_\beta'. \tag{12.27}$$

Under $\{\gamma_n\}\in\Gamma(\gamma_0,\infty,\omega_0)$ with $\beta_0=0$, the lhs of (12.27) satisfies

$$n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\beta^{*\prime}}E_{\gamma_n}m_i(\psi_n,\pi_n)=K_n(\psi_n,\pi_n;\gamma_n)\to K(\pi_0;\gamma_0), \tag{12.28}$$

where the equality holds by definition and the convergence follows from Assumption C5.

Under $\{\gamma_n\}\in\Gamma(\gamma_0,\infty,\omega_0)$ with $\beta_0=0$, the rhs of (12.27) satisfies

$$n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\psi'}E_{\gamma_n}m_i(\psi_n,\pi_n)=\frac{\partial}{\partial\psi'}E_{\gamma_n}D_\psi Q_n(\psi_n,\pi_n)\to H(\pi_0;\gamma_0), \tag{12.29}$$

where the equality holds by Assumption C2(i) and the convergence follows from Assumption C8.

Equations (12.27)-(12.29) yield the desired result. □

Proof of Lemma 12.4. From Lemma 12.2(b), we have

$$\begin{aligned}\|\beta_n\|^{-1}\left(\widehat{\psi}_n-\psi_{0,n}\right) &= \|\beta_n\|^{-1}\left(\widehat{\psi}_n(\widehat{\pi}_n)-\psi_{0,n}\right)\\ &= -\left(D_{\psi\psi}Q_n\left(\psi_{0,n},\widehat{\pi}_n\right)\right)^{-1}||\beta_n||^{-1}D_\psi Q_n\left(\psi_{0,n},\widehat{\pi}_n\right)+o_p(1)\\ &\to_p -H^{-1}(\pi_0;\gamma_0)K(\pi_0;\gamma_0)\,\omega_0=S_\beta'\omega_0, \end{aligned} \tag{12.30}$$

where the convergence in probability holds by Lemma 12.1(b), Assumption C4, $\widehat{\pi}_n-\pi_n=o_p(1)$ (which holds by Lemma 5.3), and $\pi_n=\pi_0+o(1)$, and the last equality holds by

Lemma 12.3.

Note that

$$\psi_n = \psi_{0,n} + S'_\beta \beta_n \tag{12.31}$$

by the definition of $\psi_{0,n}$. Hence,

$$\begin{aligned} & \|\beta_n\|^{-1}\left(\widehat{\psi}_n - \psi_n\right) = \|\beta_n\|^{-1}\left(\widehat{\psi}_n - \psi_{0,n}\right) - \|\beta_n\|^{-1}\left(\psi_n - \psi_{0,n}\right) \\ = & \left(S'_\beta \omega_0 + o_p(1)\right) - \|\beta_n\|^{-1} S'_\beta \beta_n = o_p(1), \end{aligned} \tag{12.32}$$

where the first equality is straightforward, the second equality uses (12.30) and (12.31), and the last equality holds because $\|\beta_n\|^{-1}\beta_n \to \omega_0$. $\square$

Proof of Theorem 5.2. We show $n^{1/2}B(\beta_n)(\widehat{\theta}_n - \theta_n) = O_p(1)$ before proving parts (a) and (b). The proof is similar to the proof of Lemma 12.2. Let $\kappa_n = J_n^{1/2} n^{1/2} B(\beta_n)(\widehat{\theta}_n - \theta_n)$. We have

$$\begin{aligned} o_p(1) & \geq n\left(Q_n(\widehat{\theta}_n) - Q_n(\theta_n)\right) \\ & = n^{1/2}\left(B^{-1}(\beta_n) D Q_n(\theta_n)\right)' J_n^{-1/2}\kappa_n + \frac{1}{2}\|\kappa_n\|^2 + nR_n^*(\widehat{\theta}_n) \\ & = O_p(\|\kappa_n\|) + \frac{1}{2}\|\kappa_n\|^2 + \left(1 + \left\|J_n^{-1/2}\kappa_n\right\|\right)^2 o_p(1) \\ & = O_p(\|\kappa_n\|) + \frac{1}{2}\|\kappa_n\|^2 + o_p(\|\kappa_n\|) + o_p\left(\|\kappa_n\|^2\right) + o_p(1), \end{aligned} \tag{12.33}$$

where the inequality holds by (3.5), the first equality holds by Assumption D1(i) with $\theta = \widehat{\theta}_n$, and the second equality holds by Assumptions D2 and D3, and the fact that $\widehat{\theta}_n \in \Theta_n(\delta_n)$ for some $\delta_n \to 0$ with probability that goes to one as $n \to \infty$. To see the latter, note that $\widehat{\pi}_n - \pi_n = o_p(1)$ and $\widehat{\psi}_n - \psi_n = o_p(1)$ by Lemma 5.3 and $||\beta_n||^{-1}(\widehat{\psi}_n - \psi_n) = o_p(1)$ by Lemma 12.4 when $\beta_n \to 0$. Rearranging (12.33) gives $\|\kappa_n\|^2 \leq 2\|\kappa_n\| O_p(1) + o_p(1)$. Let ξ_n^* denote the $O_p(1)$ term. Then, we have

$$(\|\kappa_n\| - \xi_n^*)^2 \leq (\xi_n^*)^2 + o_p(1). \tag{12.34}$$

Taking square roots gives $\|\kappa_n\| = O_p(1)$, which together with Assumption D2 gives $n^{1/2}B(\beta_n)(\widehat{\theta}_n - \theta_n) = O_p(1)$.

Now, we prove parts (a) and (b) of the Theorem at the same time. Define

$$Z_n^* = -n^{1/2} J_n^{-1} B^{-1}(\beta_n) D Q_n(\theta_n), \ \Delta_n^* = n^{1/2} B(\beta_n)(\widehat{\theta}_n - \theta_n), \text{ and}$$
$$\theta_n^\dagger = \theta_n + n^{-1/2} B^{-1}(\beta_n) Z_n^*. \tag{12.35}$$

First, we apply the quadratic approximation in Assumption D1(i) with $\theta = \theta_n^\dagger$. Rescaling both sides by n, we get

$$n\left(Q_n(\theta_n^\dagger) - Q_n(\theta_n)\right) = -\frac{1}{2} Z_n^{*\prime} J_n Z_n^* + o_p(1), \tag{12.36}$$

where the $o_p(1)$ term is obtained from Assumption D1(ii) and the fact that $\theta_n^\dagger \in \Theta_n(\delta_n)$ with probability that goes to one as $n \to \infty$ for some $\delta_n \to 0$. To see the latter, let $\theta_n^\dagger = (\psi_n^\dagger, \pi_n^\dagger)$, then (12.35), the structure of $B(\beta_n)$, $Z_n^* = O_p(1)$, and $n^{1/2}||\beta_n|| \to \infty$, yield

$$\psi_n^\dagger - \psi_n = n^{-1/2} O_p(1) = o_p(||\beta_n||) \text{ and } \pi_n^\dagger - \pi_n = n^{-1/2} \|\beta_n\|^{-1} O_p(1) = o_p(1) \tag{12.37}$$

under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, \omega_0)$.

Next, we apply the quadratic approximation in Assumption D1(i) with $\theta = \widehat{\theta}_n$ to obtain

$$\begin{aligned} n\left(Q_n(\widehat{\theta}_n) - Q_n(\theta_n)\right) &= -Z_n^{*\prime} J_n \Delta_n^* + \frac{1}{2} \Delta_n^{*\prime} J_n \Delta_n^* + o_p(1) \\ &= \frac{1}{2}(\Delta_n^* - Z_n^*)' J_n (\Delta_n^* - Z_n^*) - \frac{1}{2} Z_n^{*\prime} J_n Z_n^* + o_p(1), \end{aligned} \tag{12.38}$$

where the $o_p(1)$ term in the first equality is obtained from Assumption D1(ii) and $\widehat{\theta}_n \in \Theta_n(\delta_n)$ with probability that goes to one for some $\delta_n \to 0$ as shown above.

We have $\theta_n^\dagger \in \Theta$ with probability that goes to 1 as $n \to \infty$ by (12.37), $\theta_n \in \Theta^*$, and Assumption B1(i). In consequence,

$$Q_n(\widehat{\theta}_n) \le Q_n(\theta_n^\dagger) + o_p(1) \tag{12.39}$$

using (3.5). This, (12.36) and (12.38), give

$$\frac{1}{2}(\Delta_n^* - Z_n^*)' J_n (\Delta_n^* - Z_n^*) \le o_p(1). \tag{12.40}$$

Assumption D2, (12.38), and (12.40) imply

$$\Delta_n^* = Z_n^* + o_p(1) \text{ and } n\left(Q_n(\widehat{\theta}_n) - Q_n(\theta_n)\right) = -\frac{1}{2} Z_n^{*\prime} J_n Z_n^* + o_p(1). \tag{12.41}$$

This, combined with Assumptions D2 and D3, gives the desired results. □

12.2. Proofs for t Tests

12.2.1. Proofs of Asymptotic Distributions

The proof of Theorem 6.1 given below uses the following Lemma. Define $\widehat{\omega}_n = \widehat{\beta}_n / ||\widehat{\beta}_n||$.

Lemma 12.5. *Suppose Assumptions* A, B1-B3, C1-C8, *and* V1 *hold.*
(a) *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ *with* $||b|| < \infty$, $\widehat{\omega}_n \to_d \omega^*(\pi^*(\gamma_0, b); \gamma_0, b)$.
(b) *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $\widehat{\omega}_n \to_p \omega_0$.

Proof of Lemma 12.5. To prove Lemma 12.5(a), we have

$$\widehat{\omega}_n = n^{1/2}\widehat{\beta}_n / ||n^{1/2}\widehat{\beta}_n|| \to_d \frac{\tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)}{||\tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)||} = \omega^*(\pi^*(\gamma_0, b); \gamma_0, b) \tag{12.42}$$

by the CMT, because $n^{1/2}\widehat{\beta}_n \to_d \tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)$ by Theorem 5.1(a) and Comment 1 to Theorem 5.1 and $P(\tau_\beta(\pi^*; \gamma_0, b) = 0) = 0$ by Assumption V1(iv) (vector β).

Next, we prove that Lemma 12.5(b) holds when $\beta_0 = 0$. By Lemma 12.4, $||\beta_n||^{-1}(\widehat{\beta}_n - \beta_n) = o_p(1)$. This implies that $\widehat{\beta}_n = \beta_n + ||\beta_n|| o_p(1)$ and $||\widehat{\beta}_n|| / ||\beta_n|| = 1 + o_p(1)$. Hence,

$$\widehat{\omega}_n = \frac{\widehat{\beta}_n}{||\widehat{\beta}_n||} = \frac{\widehat{\beta}_n - \beta_n}{||\beta_n||} \frac{||\beta_n||}{||\widehat{\beta}_n||} + \frac{\beta_n}{||\beta_n||} \frac{||\beta_n||}{||\widehat{\beta}_n||} \to_p \omega_0. \tag{12.43}$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_0 \neq 0$, $\widehat{\omega}_n \to \omega_0$ by the CMT given that $\widehat{\beta}_n \to_p \beta_0$ by Lemma 5.3. □

Proof of Theorem 6.1. Under the null hypothesis $H_0 : r(\theta_n) = v_n$, the t statistic defined in (6.2) with $v = v_n$ becomes

$$T_n = \frac{n^{1/2}(r(\widehat{\theta}_n) - r(\theta_n))}{(r_\theta(\widehat{\theta}_n) B^{-1}(\widehat{\beta}_n) \widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n) r_\theta(\widehat{\theta}_n)')^{1/2}}. \tag{12.44}$$

First, we prove Theorem 6.1(a). We start with the case in which β is a scalar. Because $d_r = 1$, $d_\pi^* = 0$ implies that $r_\pi(\theta) = 0\ \forall \theta \in \Theta_\delta$ for some $\delta > 0$ by Assumption R1(iii). In consequence, $r_\theta(\theta) = [r_\psi(\theta) : 0]$ and the denominator of the t statistic in (12.44) becomes

$$\left(r_\theta(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)r_\theta(\widehat{\theta}_n)'\right)^{1/2} = \left(r_\psi(\widehat{\theta}_n)\widehat{\Sigma}_{\psi\psi,n}r_\psi(\widehat{\theta}_n)'\right)^{1/2} \tag{12.45}$$

with probability that goes to one as $n \to \infty$ (wp$\to$ 1), where $\widehat{\Sigma}_{\psi\psi,n}$ is the upper left $\psi \times \psi$ sub-matrix of $\widehat{\Sigma}_n$. We have: $r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n) = 0$ wp$\to$ 1 by (i) a mean-value expansion wrt π, (ii) Assumptions R1(i) and R1(iii), (iii) $r_\pi(\theta) = 0\ \forall \theta \in \Theta_\delta$, and (iv) $\beta_n \to 0$. Hence, we have

$$r(\widehat{\theta}_n) - r(\theta_n) = r(\widehat{\psi}_n, \widehat{\pi}_n) - r(\psi_n, \widehat{\pi}_n) + r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n) = r_\psi(\widetilde{\psi}_n, \widehat{\pi}_n)(\widehat{\psi}_n - \psi_n) \tag{12.46}$$

wp$\to$ 1, where the first equality is immediate, the second equality uses $r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n) = 0$ and a mean-value expansion of $r(\widehat{\psi}_n, \widehat{\pi}_n)$ wrt ψ around ψ_n with $\widetilde{\psi}_n$ between $\widehat{\psi}_n$ and ψ_n.

Under the conditions of Theorem 6.1(a),

$$\begin{aligned} T_n &= \frac{r_\psi(\widetilde{\psi}_n, \widehat{\pi}_n)n^{1/2}(\widehat{\psi}_n - \psi_n)}{(r_\psi(\widehat{\theta}_n)\widehat{\Sigma}_{\psi\psi,n}r_\psi(\widehat{\theta}_n)')^{1/2}} \\ &= \frac{r_\psi(\psi_0, \widehat{\pi}_n)n^{1/2}(\widehat{\psi}_n - \psi_n)}{(r_\psi(\psi_0, \widehat{\pi}_n)\widehat{\Sigma}_{\psi\psi,n}r_\psi(\psi_0, \widehat{\pi}_n)')^{1/2}} + o_p(1) \\ &= T_{\psi,n}(\widehat{\pi}_n) + o_p(1) \to_d\ T_\psi(\pi^*(b, \gamma_0); b, \gamma_0), \end{aligned} \tag{12.47}$$

where the first equality follows from (12.44)-(12.46), the second equality holds by the consistency of $\widehat{\psi}_n(\pi)$ uniformly over $\pi \in \Pi$ and the continuity of $r_\psi(\theta)$, the third equality defines $T_{\psi,n}(\pi)$ implicitly, and the convergence follows from the joint convergence $(T_{\psi,n}(\cdot), \widehat{\pi}_n) \Rightarrow (T_\psi(\cdot; \gamma_0, b), \pi^*(\gamma_0, b))$ and the CMT. The latter joint convergence holds by $\tau_n(\pi) = n^{1/2}(\widehat{\psi}_n(\pi) - \psi_n) \Rightarrow \tau(\pi; \gamma_0, b)$ (which is established in (12.20)), Assumptions V1 (scalar β) and R1, Theorem 5.1(a), the uniform consistency of $\widehat{\psi}_n(\pi)$ over $\pi \in \Pi$, and the fact that $\tau_n(\cdot)$ and $\widehat{\pi}_n$ can be written as continuous functions of the empirical process $G_n(\cdot)$ plus $o_p(1)$ terms.

In the case of a vector β, (12.47) holds with $\widehat{\Sigma}_{\psi\psi,n}$ being the $d_\psi \times d_\psi$ upper left sub-matrix of $\widehat{\Sigma}_n = \widehat{\Sigma}_n(\widehat{\theta}_n^+) = \widehat{J}_n^{-1}(\widehat{\theta}_n^+)\widehat{V}_n(\widehat{\theta}_n^+)\widehat{J}_n^{-1}(\widehat{\theta}_n^+)$ using Assumption V1 (vector β)

and with $T_{\psi,n}(\widehat{\pi}_n)$ replaced by $T_{\psi,n}(\widehat{\pi}_n, \widehat{\omega}_n)$, which is defined implicitly. In this case, the convergence in (12.47) follows from the joint convergence $(T_{\psi,n}(\cdot), \widehat{\pi}_n, \widehat{\omega}_n) \Rightarrow (T_\psi(\cdot;\gamma_0, b)$, $\pi^*(\gamma_0, b)$, $\omega^*(\pi^*(\gamma_0, b); \gamma_0, b)$, which holds by the same argument as above plus Lemma 12.5(a) and Assumption V1 (vector β). This completes the proof of part (a).

Next, we prove Theorem 6.1(b). Note that

$$\begin{aligned} r_\theta(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n) &= [r_\psi(\widehat{\theta}_n) : r_\pi(\widehat{\theta}_n)\iota^{-1}(\widehat{\beta}_n)] \\ &= \iota^{-1}(\widehat{\beta}_n)[r_\psi(\widehat{\theta}_n)\iota(\widehat{\beta}_n) : r_\pi(\widehat{\theta}_n)] \\ &= \iota^{-1}(\widehat{\beta}_n)\Big([0 : r_\pi(\widehat{\theta}_n)] + o_p(1)\Big), \end{aligned} \tag{12.48}$$

where the first equality follows from the definition of $B^{-1}(\widehat{\beta}_n)$, the second equality is straightforward, and the third equality follows from $\widehat{\beta}_n \to 0$ by Lemma 5.1(a).

When β is a scalar, in Theorem 6.1(b), the t statistic becomes

$$T_n = \frac{n^{1/2}|\iota(\widehat{\beta}_n)|\Big(r(\widehat{\theta}_n) - r(\theta_n)\Big)}{(r_\pi(\widehat{\theta}_n)\widehat{\Sigma}_{\pi\pi,n}r_\pi(\widehat{\theta}_n)')^{1/2} + o_p(1)} \to_d T_\pi(\pi^*; b, \gamma_0), \tag{12.49}$$

where the equality follows from (12.44) and (12.48) and Assumption V1 (scalar β) and the convergence holds by arguments analogous to those used to establish the convergence in (12.47).

In the case of a vector β, (12.49) holds with $\widehat{\Sigma}_{\pi\pi,n}$ being the $d_\pi \times d_\pi$ lower right sub-matrix of $\widehat{\Sigma}_n = \widehat{\Sigma}_n(\widehat{\theta}_n^+) = \widehat{J}_n^{-1}(\widehat{\theta}_n^+)\widehat{V}_n(\widehat{\theta}_n^+)\widehat{J}_n^{-1}(\widehat{\theta}_n^+)$ using Assumption V1 (vector β) and with $T_{\pi,n}(\widehat{\pi}_n)$ replaced by $T_{\pi,n}(\widehat{\pi}_n, \widehat{\omega}_n)$, which is defined implicitly. In this case, the convergence in (12.49) follows from the joint convergence $(T_{\pi,n}(\cdot), \widehat{\pi}_n, \widehat{\omega}_n) \Rightarrow (T_\pi(\cdot;\gamma_0, b)$, $\pi^*(\gamma_0, b)$, $\omega^*(\pi^*(\gamma_0, b); \gamma_0, b)$, which holds by the same argument as used to establish the convergence in (12.47) plus Lemma 12.5(a) and Assumption V1 (vector β). This completes the proof of Theorem 6.1(b).

Next, we prove Theorem 6.1(c). The proof is the same for the scalar and vector β cases because it relies on Assumption V2 which applies in both cases. First we prove the result when $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ and $\beta_n \to 0$. When $d_\pi^* = 0$, the first equality in (12.47) holds by the same arguments as above. This equality, Assumptions V2 and R1, the consistency of $\widehat{\theta}_n$ established in Lemma 5.3, Theorem 5.2(a), and the delta method together imply that $T_n \to_d N(0, 1)$.

When $d_\pi^* = 1$ and $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_n \to 0$, (12.48) still holds using $\widehat{\beta}_n \to 0$ by Lemma 5.3(b). Hence, the equality in (12.49) also holds. In this case, the t statistic becomes

$$\begin{aligned} T_n &= \frac{n^{1/2}|\iota(\widehat{\beta}_n)| \left(r_\psi(\widetilde{\theta}_n)(\widehat{\psi}_n - \psi_n) + r_\pi(\widetilde{\theta}_n)(\widehat{\pi}_n - \pi_n) \right)}{(r_\pi(\widehat{\theta}_n)\widehat{\Sigma}_{\pi\pi,n} r_\pi(\widehat{\theta}_n)')^{1/2} + o_p(1)} \\ &= \frac{n^{1/2}|\iota(\widehat{\beta}_n)| r_\pi(\widetilde{\theta}_n)(\widehat{\pi}_n - \pi_n)}{(r_\pi(\widehat{\theta}_n)\widehat{\Sigma}_{\pi\pi,n} r_\pi(\widehat{\theta}_n)')^{1/2} + o_p(1)} + o_p(1) \\ &\to_d N(0,1), \end{aligned} \quad (12.50)$$

where the first equality follows from (12.44), (12.48), and a mean-value expansion of $r(\widehat{\theta}_n)$ wrt θ around θ_n with $\widetilde{\theta}_n$ between $\widehat{\theta}_n$ and θ_n, the second equality holds because (i) $n^{1/2}(\widehat{\psi}_n - \psi_n) = O_p(1)$ by Theorem 5.2(a), (ii) $\beta_n \to 0$ and the consistency of $\widehat{\theta}_n$ in Lemma 5.3, (iii) the continuity of $r_\theta(\theta)$ in Assumption R1, and (iv) Assumption V2, and the convergence in distribution holds by (i) the consistency of $\widehat{\theta}_n$, (ii) the continuity of $r_\theta(\theta)$, (iii) $n^{1/2}\iota(\widehat{\beta}_n)(\widehat{\pi}_n - \pi_n) \to_d N(0, \Sigma_{\pi\pi}(\gamma_0))$ by Theorem 5.2(a), where $\Sigma_{\pi\pi}(\gamma_0)$ is the lower right $d_\pi \times d_\pi$ sub-matrix of $\Sigma(\gamma_0) = J^{-1}(\gamma_0)V(\gamma_0)J^{-1}(\gamma_0)$, (iv) Assumption V2, and (v) the delta method.

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ and $\beta_n \to \beta_0 \neq 0$,

$$n^{1/2}(r(\widehat{\theta}_n) - r(\theta_n)) \to_d N(0, r_\theta(\theta_0)B^{-1}(\beta_0)\Sigma(\gamma_0)B^{-1}(\beta_0)r_\theta(\theta_0)') \quad (12.51)$$

by Theorem 5.2(a) and the delta method. By Assumptions R1(i) and V2 and the consistency of $\widehat{\theta}_n$ established in Lemma 5.3,

$$r_\theta(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)r_\theta(\widehat{\theta}_n)' \to_p r_\theta(\theta_0)B^{-1}(\beta_0)\Sigma\left(\gamma_0\right)B^{-1}(\beta_0)r_\theta(\theta_0)'. \quad (12.52)$$

The desired result follows from (12.44), (12.51), and (12.52). □

12.2.2. Proofs of Asymptotic Size Results

Proof of Theorem 6.2. We only prove the asymptotic size result of Theorem 6.2 for the symmetric two-sided CI, which is based on $|T_n|$. The proofs for the one-sided CI's, which are based on T_n and $-T_n$, are analogous. We prove the result of Theorem 6.2 for $|T_n|$ by applying Corollary 1.1(b) of ACG. To this end, we verify the high-level conditions in Assumptions B1, B2*, and C of ACG.

To verify Assumptions B1 and B2* of ACG, we start with the specification of λ and $h_n(\lambda)$ in this application. Let

$$\lambda = (||\beta||, \beta/||\beta||, \zeta, \pi, \phi) \text{ and } h_n(\lambda) = (n^{1/2}||\beta||, ||\beta||, \beta/||\beta||, \zeta, \pi, \phi), \tag{12.53}$$

where $\gamma = (\beta, \zeta, \pi, \phi) \in \Gamma$, $\theta = (\beta, \zeta, \pi) \in R^{d_\beta + d_\zeta + d_\pi}$, $\phi \in \Phi^*(\theta) \subset \Phi^*$ for some compact metric space Φ^*, and by definition $\beta/||\beta|| = 1_{d_\beta}/||1_{d_\beta}||$ with $1_{d_\beta} = (1, ..., 1) \in R^{d_\beta}$ if $\beta = 0$. The parameter space of λ is $\Lambda = \{\lambda : \gamma = (\beta, \zeta, \pi, \phi) \in \Gamma\}$. Note that λ is an equivalent reparameterization of $\gamma = (\beta, \zeta, \pi, \phi)$. Hence, λ also indexes the true distribution and the parameter of interest $r(\theta)$ can be written as a function of λ, as desired in the definition of CS_n and $CP_n(\lambda)$ in ACG.

First, we verify Assumption B2* of ACG. With the specification of λ and $h_n(\lambda)$ in (12.53), Assumptions B2*(i) and B2*(ii) of ACG hold with $\lambda_1 = ||\beta||$, $(\lambda_2, ..., \lambda_q) = (\beta/||\beta||, \zeta, \pi)$, $\lambda_{q+1} = \phi$, $r = 1$, $d_{n,1} = n^{1/2}$, $h_{n,1}(\lambda) = n^{1/2}||\beta||$, $(h_{n,2}(\lambda), ..., h_J(\lambda)) = (m_2(\lambda), ..., m_J(\lambda)) = (||\beta||, \beta/||\beta||, \zeta, \pi) \in R^{d_\theta + 1}$, and $h_{J+1}(\lambda) = m_{J+1}(\lambda) = \phi \in \Phi^*(\theta) \subset \Phi^*$, where Φ^* is compact. Assumption B2*(iii) of ACG holds because $m_2(\lambda) = \lambda_1$ is continuous in λ_1 (uniformly over $(\lambda_{r+1}, ..., \lambda_{q+1})$ because it does not depend on $(\lambda_{r+1}, ..., \lambda_{q+1})$) and $m_j(\lambda)$ does not depend on λ_1 $\forall j \geq 3$. Note that by the reparameterization, $||\beta||$ and $\beta/||\beta||$ are treated as two parameters that can take values independently. Assumption B2*(iv) of ACG holds by Assumption B2(iii) of this paper.

We use H^* and h^* to denote H and h in ACG to distinguish them from H and h in the current paper as defined in (6.14).

Next, we verify Assumption B1 of ACG. For every $\{\lambda_n \in \Lambda : n \geq 1\}$, there is an equivalent reparameterization $\{\gamma_n \in \Gamma : n \geq 1\}$. Moreover, $h_n(\lambda_n) \to h^* \in H^*$ implies that $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$ or $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$. To see this, let $h_n(\lambda_n) \to h^* = (h_1^*, h_2^*, h_3^*, h_4^*)$, where $n^{1/2}||\beta_n|| \to h_1^*$, $||\beta_n|| \to h_2^*$, $\beta_n/||\beta_n|| \to h_3^*$, and $(\zeta_n, \pi_n, \phi_n) \to h_4^*$. Note that $||h_3^*|| = 1$ by construction. Then, under $\{\lambda_n : n \geq 1\}$ such that $h_n(\lambda_n) \to h^*$, we have $\gamma_n \to \gamma_0 = (h_2^* h_3^*, h_4^*)$ and $n^{1/2}\beta_n \to b = h_1^* h_3^*$. Hence, (i) if $h_1^* \in R$, we have $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$ and $|T_n| \to_d |T(h)|$ with $h = (b, \gamma_0) = (h_1^* h_3^*, h_2^* h_3^*, h_4^*)$ by Theorem 6.1(a) and 6.1(b) and (6.14) and (ii) if $|h_1^*| = \infty$, we have $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\omega_0 = h_3^*$ and $|T_n| \to_d |Z|$, where $Z \sim N(0, 1)$, by Theorem 6.1(c).

By definition, $CP_n(\lambda_n) = P_{\lambda_n}(|T_n| \leq z_{1-\alpha/2})$. By Theorem 6.1 and Assumption V3, $CP_n(\lambda_n) \to P(|T(h)| \leq z_{1-\alpha/2})$ under $\{\lambda_n : n \geq 1\}$ for which $h_n(\lambda_n) \to h^* \in H^*$ with

$h_1^* \in R$. Similarly, we have $CP_n(\lambda_n) \rightarrow P(|Z| \leq z_{1-\alpha/2}) = 1 - \alpha$ under $\{\lambda_n : n \geq 1\}$ for which $h_n(\lambda_n) \rightarrow h^* \in H^*$ with $||h_1^*|| = \infty$. Therefore, Assumption B1 holds with $CP^-(h^*) = CP^+(h^*) = P(|T(h)| \leq z_{1-\alpha/2})$ if $h_1^* \in R$ and it holds with $CP^-(h^*) = CP^+(h^*) = 1 - \alpha$ if $||h_1^*|| = \infty$, where h is a function of h^* as defined above.

Assumption C of ACG holds automatically because $CP^-(h^*) = CP(h^*)\ \forall h^* \in H^*$.

By Lemma 1.1(b) of ACG, the nominal $1 - \alpha$ symmetric two-sided t CS has

$$AsySz = \min\{\inf_{h^* \in H^* : h_1^* \in R} P(|T(h)| \leq z_{1-\alpha/2}), 1 - \alpha\}. \tag{12.54}$$

It remains to show that the set $H' = \{(b, \gamma_0) : h^* \in H^*$ with $h_1^* \in R$, $b = h_1^* h_3^*$, $\gamma_0 = (h_2^* h_3^*, h_4^*)\}$ is equivalent to the set $H = \{(b, \gamma_0) : ||b|| < \infty$ and $\gamma_0 \in \Gamma$ with $\beta_0 = 0\}$. This holds because (i) $h^* \in H^*$ implies $\gamma_0 \in \Gamma$, because γ_0 is the limit of a convergent sequence in Γ and Γ is compact, and $h_1^* \in R$ implies that $b = h_1^* h_3^* \in R^{d_\beta}$ and $\beta_0 = 0$ and (ii) for any $h = (b, \gamma_0) \in H$, there exists $h^* \in H^*$ with $h_1^* \in R$ such that $b = h_1^* h_3^*$ and $\gamma_0 = (h_2^* h_3^*, h_4^*)$, i.e., $h_1^* = ||b||$, $h_2^* = 0$, $h_3^* = b/||b||$, and $h_4^* = (\zeta_0, \pi_0, \phi_0)$. (If $b = 0$, let $h_3^* = (1, ..., 1) \in R^{d_\beta}$). This completes the proof. $\square$

Proof of Theorem 7.1. The proof of Theorem 7.1(a) for the LF critical value is the same as that of Theorem 6.2 with $c_{|t|,1-\alpha}^{LF}$ $(= \max\{\sup_{h \in H} c_{|t|,1-\alpha}(h), z_{1-\alpha/2}\})$ in place of $z_{1-\alpha/2}$, $z_{1-\alpha}$, and $z_{1-\alpha}$ for $\mathcal{T}_n = |T_n|$, T_n, and $-T_n$, respectively, using Assumption LF(i) in place of Assumption V3. For the case of $\mathcal{T}_n = |T_n|$, this proof delivers

$$AsySz = \min\{\inf_{h \in H} P(|T(h)| \leq c_{|t|,1-\alpha}^{LF}), P(|Z| \leq c_{|t|,1-\alpha}^{LF})\}, \tag{12.55}$$

where $Z \sim N(0, 1)$. The rhs of (12.55) is greater than or equal to $1 - \alpha$ because (i) $P(|T(h)| \leq c_{|t|,1-\alpha}^{LF}) \geq P(|T(h)| \leq c_{|t|,1-\alpha}(h)) \geq 1 - \alpha\ \forall h \in H$, where the second inequality holds by the definition of the quantile $c_{|t|,1-\alpha}(h)$, and (ii) $P(|Z| \leq c_{|t|,1-\alpha}^{LF}) \geq P(|Z| \leq z_{1-\alpha/2}) = 1 - \alpha$. The rhs of (12.55) is less than or equal to $1 - \alpha$ because if $c_{|t|,1-\alpha}^{LF} = z_{1-\alpha/2}$, then $P(|Z| \leq c_{|t|,1-\alpha}^{LF}) = 1 - \alpha$ and if $c_{|t|,1-\alpha}^{LF} > z_{1-\alpha/2}$, then $P(|T(h_{\max})| \leq c_{|t|,1-\alpha}^{LF}) = P(|T(h_{\max})| \leq c_{|t|,1-\alpha}(h_{\max})) = 1 - \alpha$, where both equalities hold using Assumption LF. Hence, $AsySz = 1 - \alpha$. The proofs for $\mathcal{T}_n = T_n$ and $-T_n$ are analogous.

The proof of Theorem 7.1(b) for the NI-LF critical value is the same as that just given for the LF critical value except that H, $c_{|t|,1-\alpha}^{LF}$, $h_{\max}$, and Assumption LF are replaced by $H(v)$, $c_{|t|,1-\alpha}^{LF}(v)$ $(= \max\{\sup_{h \in H(v)} c_{|t|,1-\alpha}(h), z_{1-\alpha/2}\})$, $h_{\max}(v)$, and Assumption NI-LF, respectively, for $v \in V_r$ and the rhs of (12.55) has $\inf_{v \in V_r}$ added.

Theorem 7.1(c) is proved along the lines of the proof of Theorem 6.2 by verifying Assumptions B1, B2*, and C of ACG. Assumption B2* of ACG already has been verified in the proof of Theorem 6.2. Now we verify Assumptions B1 and C of ACG for the robust t CI's. The notation used here is the same as in the proof of Theorem 6.2.

We first show $\widetilde{c}_{|t|,1-\alpha,n} \rightarrow_p c^{LF}_{|t|,1-\alpha}$ under $\{\lambda_n : n \geq 1\}$ for which $h_n(\lambda_n) \rightarrow h^* \in H^*$ with $h_1^* \in R$, i.e., $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$. By the construction of $\widetilde{c}_{|t|,1-\alpha,n}$, it suffices to show that $P_{\gamma_n}(A_n \leq \kappa_n) \rightarrow 1$. This holds if $A_n = O_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$, because $\kappa_n \rightarrow \infty$ by Assumption K(i).

When β is a scalar, we have

$$A_n = \left(n^{1/2}\widehat{\beta}_n' \widehat{\Sigma}^{-1}_{\beta\beta,n} n^{1/2}\widehat{\beta}_n\right)^{1/2} \rightarrow_d \left(\tau_\beta(\pi^*)' \Sigma^{-1}_{\beta\beta}(\pi^*; \gamma_0)\tau_\beta(\pi^*)\right)^{1/2}, \tag{12.56}$$

where π^* and $\tau_\beta(\cdot)$ abbreviate $\pi^*(\gamma_0, b)$ and $\tau(\cdot; \gamma_0, b)$, respectively, and the convergence in distribution holds by Theorem 5.1(a) and Assumption V1. By Assumptions B1(iii), V1(ii), and V1(iii), $\inf_{\pi \in \Pi} \Sigma_{\beta\beta}(\pi; \gamma_0) > 0$. Hence, $A_n = O_p(1)$ as desired.

When β is a vector, (12.56) holds with $\Sigma_{\beta\beta}(\pi^*; \gamma_0)$ replaced by $\Sigma_{\beta\beta}(\pi^*, \omega^*(\pi^*); \gamma_0, \omega_0)$ by Theorem 5.1(a), Assumption V1, and the joint convergence $(n^{1/2}\widehat{\beta}_n, \widehat{\pi}_n, \widehat{\omega}_n) \rightarrow_d (\tau_\beta(\pi^*), \pi^*, \omega^*(\pi^*))$. By Assumptions B1(iii), V1(ii), and V1(iii), $\inf_{\pi \in \Pi, ||\omega||=1} \lambda_{\min}(\Sigma_{\beta\beta}(\pi, \omega; \gamma_0, \omega_0)) > 0$. Hence, $A_n = O_p(1)$ as desired.

Using Theorem 6.1(a) and 6.1(b), $\widetilde{c}_{|t|,1-\alpha,n} \rightarrow_p c^{LF}_{|t|,1-\alpha}$, and Assumption V3, we obtain $CP_n(\lambda_n) = P_{\lambda_n}(|T_n| \leq \widetilde{c}_{|t|,1-\alpha,n}) \rightarrow P(|T(h)| \leq c^{LF}_{|t|,1-\alpha})$ under $\{\lambda_n : n \geq 1\}$ for which $h_n(\lambda_n) \rightarrow h^* \in H^*$ with $h_1^* \in R$. Hence, when $h_1^* \in R$, Assumption B1 of ACG holds with

$$CP^-(h^*) = CP^+(h^*) = P(|T(h)| \leq c^{LF}_{|t|,1-\alpha}) \geq 1 - \alpha. \tag{12.57}$$

By the construction of $\widetilde{c}_{|t|,1-\alpha,n}$, we have $z_{1-\alpha/2} \leq \widetilde{c}_{|t|,1-\alpha,n} \leq c^{LF}_{|t|,1-\alpha}$. Hence,

$$P_{\lambda_n}(|T_n| \leq z_{1-\alpha/2}) \leq P_{\lambda_n}(|T_n| \leq \widetilde{c}_{|t|,1-\alpha,n}) \leq P_{\lambda_n}(|T_n| \leq c^{LF}_{|t|,1-\alpha}). \tag{12.58}$$

Under $\{\lambda_n : n \geq 1\}$ for which $h_n(\lambda_n) \rightarrow h^* \in H^*$ with $||h_1^*|| = \infty$, i.e., $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,

$$\begin{aligned} &P_{\lambda_n}(|T_n| \leq z_{1-\alpha/2}) \rightarrow P(|Z| \leq z_{1-\alpha/2}) = 1 - \alpha \text{ and} \\ &P_{\lambda_n}(|T_n| \leq c^{LF}_{|t|,1-\alpha}) \rightarrow P(|Z| \leq c^{LF}_{|t|,1-\alpha}) \geq 1 - \alpha. \end{aligned} \tag{12.59}$$

By (12.58) and (12.59), when $||h_1^*|| = \infty$, Assumption B1 of ACG holds with

$$CP^-(h^*) = 1 - \alpha \text{ and } CP^+(h^*) = P(|Z| \le c^{LF}_{|t|,1-\alpha}). \tag{12.60}$$

This completes the verification of Assumption B1 of ACG.

Next, we verify Assumption C of ACG. By the verification of Assumption B1 of ACG, we have $\inf_{h^* \in H^*} CP^-(h^*) = 1 - \alpha$. It remains to show that for some $h^* \in H$, $CP^-(h^*) = CP^+(h^*) = 1 - \alpha$. To this end, we consider $h^{**} \in H$ with $||h_1^{**}|| = \infty$ and $h_2^{**} \ne 0$, where $h^{**} = (h_1^{**}, ..., h_4^{**})$. Let $\{\lambda_n^{**} : n \ge 1\}$ be any sequence of true parameters under which $h_n(\lambda_n^{**}) \to h^{**}$. The equivalent reparameterization $\{\gamma_n : n \ge 1\}$ satisfies $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_0 = h_2^{**} h_3^{**} \ne 0$.

Now we show $\widetilde{c}_{|t|,1-\alpha,n} \to_p z_{1-\alpha/2}$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_0 \ne 0$. It suffices to show that $P_{\gamma_n}(A_n > \kappa_n) \to 1$. We have

$$\kappa_n^{-1} A_n = \left(n^{1/2}\kappa_n^{-1}\right) \left(\widehat{\beta}_n' \widehat{\Sigma}^{-1}_{\beta\beta,n} \widehat{\beta}_n\right)^{1/2} \to_p \infty, \tag{12.61}$$

where the divergence to infinity holds because $n^{1/2}\kappa_n^{-1} \to \infty$ by Assumption K(ii), $\widehat{\beta}_n \to_p \beta_0 \ne 0$ by Lemma 5.1(b), $\widehat{\Sigma}_{\beta\beta,n} \to_p \Sigma_{\beta\beta}(\gamma_0)$ by Assumption V2, where $\Sigma_{\beta\beta}(\gamma_0)$ denote the upper left $d_\beta \times d_\beta$ sub-matrix of $\Sigma(\gamma_0) = J^{-1}(\gamma_0)V(\gamma_0)J^{-1}(\gamma_0)$, and $\Sigma_{\beta\beta}(\gamma_0)$ is nonsingular by Assumptions D2 and D3. Hence, $P_{\gamma_n}(A_n > \kappa_n) \to 1$.

Using $|T_n| \to_d |Z|$ by Theorem 6.1(c), $\widetilde{c}_{|t|,1-\alpha,n} \to_p z_{1-\alpha/2}$, and the continuity of the df of Z, we obtain $CP_n(\lambda_n) = P_{\lambda_n}(|T_n| \le \widetilde{c}_{|t|,1-\alpha,n}) \to 1 - \alpha$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_0 \ne 0$. This implies $CP^-(h^{**}) = CP^+(h^{**}) = 1 - \alpha$ for $h^{**} \in H$ with $||h_1^{**}|| = \infty$ and $h_2^{**} \ne 0$ and completes the verification of Assumption C of ACG.

Applying Lemma 1.1 of ACG, we conclude that the nominal $1 - \alpha$ type 1 robust two-sided t CI has $AsySz = 1 - \alpha$. The proofs for one-sided t CI's are analogous.

The proof of Theorem 7.1(d) for the NI type 1 robust critical value is analogous to that just given for the type 1 robust critical value except that H, $c^{LF}_{|t|,1-\alpha}$, and $\widetilde{c}_{|t|,1-\alpha,n}$ are replaced by $H(v)$, $c^{LF}_{|t|,1-\alpha}(v)$, and $\widetilde{c}_{|t|,1-\alpha,n}(v)$, respectively, for $v \in V_r$.

The proof of Theorem 7.1(e) for the type 2 robust critical value uses the following results. First, under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$,

$$(|T_n|, \widehat{c}_{|t|,1-\alpha,n}) \to_d (|T(h)|, \widehat{c}_{|t|,1-\alpha}(h)), \tag{12.62}$$

because (i) $T_n \to_d T(h)$ by Theorem 6.1, (ii) $A_n \to_d A(h)$ by (12.56), (iii) $\widehat{c}_{|t|,1-\alpha,n} \to_d$

$\widehat{c}_{|t|,1-\alpha}(h)$ by the continuous mapping theorem using result (ii), (7.6), (7.10), and the continuity of $s(x)$ for $x \in [0, \infty)$ (which implies that $\widehat{c}_{|t|,1-\alpha}(h)$ is a continuous function of $A(h)$), and (iv) the convergence is joint because $|T_n|$ and $\widehat{c}_{|t|,1-\alpha,n}$ are functions of the same underlying statistics.

Equation (12.62) and Assumption Rob2(i) imply: Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$,

$$P(|T_n| \leq \widehat{c}_{|t|,1-\alpha,n}) \rightarrow_d P(|T(h)| \leq \widehat{c}_{|t|,1-\alpha}(h)) \; \forall h = (b, \gamma_0) \in H. \tag{12.63}$$

Second, under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, we have: (i) $A_n \rightarrow_p \infty$ by Theorem 6.1(c) with $r(\theta) = \beta$ plus the fact that the estimator $\widehat{\beta}_n$ in A_n is centered at 0, rather than at β_n, which causes the divergence in probability to ∞, (ii) $s(A_n - \kappa) \rightarrow_p 0$ by results (i) and (ii) and the assumption that $s(x) \rightarrow 0$ as $x \rightarrow \infty$, and (iii) $\widehat{c}_{|t|,1-\alpha,n} \rightarrow_p c_{|t|,1-\alpha}(\infty) + \Delta_2 = z_{1-\alpha/2} + \Delta_2$ using result (ii) and (7.6). Result (iii) and $|T_n| \rightarrow_d |Z|$ for $Z \sim N(0, 1)$, which holds by Theorem 6.1(c), yield: Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,

$$P(|T_n| \leq \widehat{c}_{|t|,1-\alpha,n}) \rightarrow_d P(|Z| \leq z_{1-\alpha/2} + \Delta_2). \tag{12.64}$$

Given (12.63) and (12.64), the proof of Theorem 7.1(e) for the case of $\mathcal{T}_n = |T_n|$ is analogous to that of Theorem 6.2 with $\widehat{c}_{|t|,1-\alpha,n}$ in place of $z_{1-\alpha/2}$. This proof delivers

$$AsySz = \min\{\inf_{h \in H} P(|T(h)| \leq \widehat{c}_{|t|,1-\alpha}(h)), P(|Z| \leq z_{1-\alpha/2} + \Delta_2)\}. \tag{12.65}$$

Using this, we obtain

$$AsySz = \min\{\inf_{h \in H}(1 - NRP(\Delta_1, \Delta_2; h)), P(|Z| \leq z_{1-\alpha/2} + \Delta_2)\} \geq 1 - \alpha, \tag{12.66}$$

where $NRP(\Delta_1, \Delta_2; h)$ is defined in (7.8) with $\mathcal{T}(h) = |T(h)|$, the equality holds by (7.8) and (7.10) with $\mathcal{T}(h) = |T(h)|$ and (12.65), and the inequality holds by the definitions of Δ_1 and Δ_2 in (7.9), $P(|Z| \leq z_{1-\alpha/2}) = 1 - \alpha$, and $\Delta_2 \geq 0$.

If $\Delta_2 = 0$, then $P(|Z| \leq z_{1-\alpha/2} + \Delta_2) = 1 - \alpha$ and $AsySz \leq 1 - \alpha$ by (12.66). Alternatively, if $\Delta_2 > 0$, we have

$$AsySz \leq 1 - NRP(\Delta_1, \Delta_2; h^*) = 1 - \alpha, \tag{12.67}$$

where the inequality holds using the equality in (12.66) and the equality holds by Assumption Rob2(ii). This completes the proof of Theorem 7.1(e) for the case $\mathcal{T}_n = |T_n|$. The proofs of Theorem 7.1(e) for the cases $\mathcal{T}_n = T_n$ and $-T_n$ are analogous.

The proof of Theorem 7.1(f) is analogous to that of Theorem 7.1(e) using Assumption NI-Rob2 in place of Assumption Rob2. □

12.3. Proofs of Sufficient Conditions

12.3.1. Assumption B3

Proof of Lemma 11.1. Assumptions B3*(i) and B3*(iii) and the compactness of Θ lead to Assumption B3(iii) by a standard argument. For any $\pi \in \Pi$, we have $q(\pi) = \inf_{\psi \in \Psi(\pi)/\Psi_0} Q(\psi, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0) > 0$, where Ψ_0 is defined in Assumption B3(ii), by the same argument using Assumption B3*(ii) in place of Assumption B3*(iii). To show $\inf_{\pi \in \Pi} q(\pi) > 0$, as is required by Assumption B3(ii), it suffices to show $q(\pi)$ is continuous on the compact set Π. For any $\pi \in \Pi$, $\Psi(\pi)/\Psi_0$ is compact and $\inf_{\psi \in \Psi(\pi)/\Psi_0} Q(\psi, \pi; \gamma_0) = Q(\psi^*(\pi), \pi; \gamma_0)$ for some $\psi^*(\pi) \in \Psi(\pi)$ by Assumptions B3*(i) and B3*(iv). To show $q(\pi)$ is continuous on Π, it is equivalent to show $Q(\psi^*(\pi), \pi; \gamma_0)$ is continuous on Π.

For any $\varepsilon > 0$, there exists $\delta_1 > 0$ such that $\|\psi_1 - \psi^*(\pi_2)\| < \delta_1$ and $\|\pi_1 - \pi_2\| < \delta_1$ implies that $|Q(\psi_1, \pi_1; \gamma_0) - Q(\psi^*(\pi_2), \pi_2; \gamma_0)| < \varepsilon$ by the continuity of $Q(\theta; \gamma_0)$. By Assumption B3*(v), for any $\delta_1 > 0$, there exists a $\delta_2 > 0$ such that $\|\pi_1 - \pi_2\| < \delta_2$ implies that $d_H(\Psi(\pi_1), \Psi(\pi_2)) < \delta_1$. The condition $d_H(\Psi(\pi_1), \Psi(\pi_2)) < \delta_1$ implies that $\inf_{\psi \in \Psi(\pi_1)} \|\psi - \psi^*(\pi_2)\| < \delta_1$. Because $\Psi(\pi_1)$ is compact, there exists $\psi^{**}(\pi_1) \in \Psi(\pi_1)$ such that $\|\psi^{**}(\pi_1) - \psi^*(\pi_2)\| = \inf_{\psi \in \Psi(\pi_1)} \|\psi - \psi^*(\pi_2)\|$. Hence, $||\psi^{**}(\pi_1) - \psi^*(\pi_2)|| < \delta_1$ if $\|\pi_1 - \pi_2\| < \delta_2$. Take $\delta = \min\{\delta_1, \delta_2\}$, then

$$|Q(\psi^{**}(\pi_1), \pi_1; \gamma_0) - Q(\psi^*(\pi_2), \pi_2; \gamma_0)| < \varepsilon \tag{12.68}$$

for any $\|\pi_1 - \pi_2\| < \delta$. Hence,

$$Q(\psi^*(\pi_1), \pi_1; \gamma_0) \leq Q(\psi^{**}(\pi_1), \pi_1; \gamma_0) < Q(\psi^*(\pi_2), \pi_2; \gamma_0) + \varepsilon \tag{12.69}$$

for any $\|\pi_1 - \pi_2\| < \delta$, where the first inequality is implied by the definition of $\psi^*(\pi_1)$ and the second inequality holds by (12.68).

Similarly, we can show $Q(\psi^*(\pi_2), \pi_2; \gamma_0) < Q(\psi^*(\pi_1), \pi_1; \gamma_0) + \varepsilon$ for any $\|\pi_1 - \pi_2\| < \delta$. Hence, for any $\varepsilon > 0$, there exists $\delta > 0$ such that $|Q(\psi^*(\pi_1), \pi_1; \gamma_0) - Q(\psi^*(\pi_2), \pi_2; \gamma_0)| < \varepsilon$ for any $\|\pi_1 - \pi\| < \delta$. This completes the proof. □

12.3.2. Assumption C5

Proof of Lemma 11.2. We now verify Assumption C5. Without loss of generality, suppose $\beta \in R$. Let $\{\beta_k^* : k \geq 1\}$ be a sequence that converges to β^* and suppose γ_k^* only differs from γ^* by replacing β^* with β_k^*. The partial derivative of $E_{\gamma^*} m(W_i, \theta)$ wrt β^* is

$$\lim_{k\to\infty} \frac{E_{\gamma_k^*} m(W_i,\theta) - E_{\gamma^*} m(W_i,\theta)}{\beta_k^* - \beta^*} = \lim_{k\to\infty} \int_{\mathcal{W}} m(w,\theta) \frac{f_{W_i}(w;\gamma_k^*) - f_{W_i}(w;\gamma^*)}{\beta_k^* - \beta^*} d\mu(w)$$
$$= \int_{\mathcal{W}} m(w,\theta) \left(\lim_{k\to\infty} \frac{f_{W_i}(w;\gamma_k^*) - f_{W_i}(w;\gamma^*)}{\beta_k^* - \beta^*} \right) d\mu(w) = \int_{\mathcal{W}} m(w,\theta) f_{\beta,W_i}(w;\gamma^*) d\mu(w), \tag{12.70}$$

where the first equality holds by Assumption C5*(i), the second equality holds by the dominated convergence theorem (DCT), and the last equality holds by the differentiability of $f_{W_i}(w;\gamma^*)$ wrt β^*. The DCT holds in the second equality using

$$\frac{f_{W_i}(w;\gamma_k^*) - f_{W_i}(w;\gamma^*)}{\beta_k^* - \beta^*} = f_{\beta,W_i}(w;\widetilde{\gamma}_k(w)) \text{ and}$$
$$\int_{\mathcal{W}} \sup_{\theta\in\Theta} ||m(w,\theta)|| \cdot \sup_{\gamma\in N(\gamma^*,\varepsilon)} |f_{\beta,W_i}(w;\gamma)| d\mu(w) < \infty, \tag{12.71}$$

where the equality holds by the mean-value expansion with $\widetilde{\gamma}_k(w)$ between γ_k^* and γ^* and the inequality holds by Assumption C5*(v), Hence, Assumption C5(i) holds with $K_n(\theta;\gamma^*) = n^{-1} \sum_{i=1}^n \int_{\mathcal{W}} m(w,\theta) f_{\beta,W_i}(w;\gamma^*) d\mu(w)$.

We now show Assumption C5(ii) holds with $K(\psi_0,\pi;\gamma_0) = \int_{\mathcal{W}} m(w,\psi_0,\pi) f_{\beta,W}(w;\gamma_0) d\mu(w)$. To show Assumption C5(ii), we have

$$\sup_{\pi\in\Pi} |K_n(\psi_n,\pi;\widetilde{\gamma}_n) - K(\psi_0,\pi;\gamma_0)|$$
$$\leq \int \sup_{\pi\in\Pi} \left| m(w,\psi_n,\pi) \left(n^{-1} \sum_{i=1}^n f_{\beta,W_i}(w;\widetilde{\gamma}_n) \right) - m(w,\psi_0,\pi) f_{\beta,W}(w;\gamma_0) \right| d\mu(w)$$
$$\leq \int \sup_{\theta\in\Theta} |m(w,\theta)| \cdot \left| (n^{-1} \sum_{i=1}^n f_{\beta,W_i}(w;\widetilde{\gamma}_n) - f_{\beta,W}(w;\gamma_0) \right| d\mu(w) +$$
$$\int \sup_{\pi\in\Pi} |m(w,\psi_n,\pi) - m(w,\psi_0,\pi)| \, f_{\beta,W}(w;\gamma_0) d\mu(w), \tag{12.72}$$

where the first inequality is obvious, and the second inequality holds by the triangle inequality. The third line of (12.72) converges to 0 by the DCT under Assumptions C5*(ii), C5*(iii), and C5*(v) using $\widetilde{\gamma}_n \to \gamma_0$. The fourth line of (12.72) converges to 0 by Assumptions C5*(iv) and C5*(v). This yields Assumption C5(ii).

Assumption C5(iii) holds by the DCT using Assumptions C5*(iv) and C5*(v). □

12.3.3. Assumption C6

Proof of Lemma 11.3. We block diagonalize $H(\pi;\gamma_0)$ using the $d_\psi \times d_\psi$ matrix $A(\pi)$ defined by

$$A(\pi) = \begin{bmatrix} I_{d_\beta} & -H_{12}(\pi)H_{22}^{-1} \\ 0_{d_\zeta \times d_\beta} & I_{d_\zeta} \end{bmatrix}. \tag{12.73}$$

Simple calculations yield

$$\begin{aligned}
A(\pi)H(\pi;\gamma_0)A(\pi)' &= \begin{bmatrix} H_{11}^*(\pi) & 0_{d_\beta \times d_\zeta} \\ 0_{d_\zeta \times d_\beta} & H_{22} \end{bmatrix}, \\
A(\pi)[G(\pi;\gamma_0) + K(\pi;\gamma_0)b] &= \begin{bmatrix} G_1^*(\pi;\gamma_0) + K_1^*(\pi;\gamma_0)b \\ G_2 + K_2 b \end{bmatrix}, \text{ and} \\
A(\pi)K(\pi;\gamma_0)\omega_0 &= K_1^*(\pi;\gamma_0)\omega_0.
\end{aligned} \tag{12.74}$$

In consequence, we have

$$\begin{aligned}
&\xi(\pi;\gamma_0,b) \\
&= -\frac{1}{2}\left(G(\pi;\gamma_0) + K(\pi;\gamma_0)\,b\right)' A(\pi)'[A(\pi)H(\pi;\gamma_0)A(\pi)']^{-1}A(\pi)\left(G(\pi;\gamma_0) + K(\pi;\gamma_0)b\right) \\
&= \xi_1(\pi;\gamma_0,b) + \xi_2(\gamma_0,b).
\end{aligned} \tag{12.75}$$

Similarly, we have

$$\begin{aligned}
\eta(\pi;\gamma_0,\omega_0) &= -\frac{1}{2}\omega_0' K\left(\pi;\gamma_0\right)' A(\pi)'[A(\pi)H\left(\pi;\gamma_0\right)A(\pi)']^{-1}A(\pi)K\left(\pi;\gamma_0\right)\omega_0 \\
&= \eta_1(\pi;\gamma_0,\omega_0) + \eta_2(\gamma_0,\omega_0),
\end{aligned} \tag{12.76}$$

which completes the proof. □

Lemma 11.4 follows immediately from the following Lemma, which is an extension of Lemma 2.6 of Kim and Pollard (1990).

Lemma 12.6. *Let $\{Z(t) : t \in T\}$ be a univariate Gaussian process with continuous* *sample paths, indexed by a σ-compact metric space T. If $Var(Z(s) - Z(t)) \neq 0$ and $Var(Z(s) + Z(t)) \neq 0$, $\forall s, t \in T$ with $s \neq t$, then, with probability one, no sample path of $Z^2(\cdot)$ can achieve its supremum at two distinct points of T.*

Proof of Lemma 12.6. A sample path of Z^2 achieves its supremum only where Z achieves its supremum or infimum. By Lemma 2.6 of KP, if $Var(Z(s) - Z(t)) \neq 0$, $\forall s \neq t$, no sample path of Z achieves its supremum at two distinct points of T with probability one. By the same argument, no sample path of Z achieves its infimum at two distinct points in T with probability one.

It only remains to show that with probability one, no sample path of Z has its supremum equal to minus its infimum at two distinct points. To show this, we use the condition

$$Var(Z(s) + Z(t)) \neq 0, \ \forall s \neq t. \tag{12.77}$$

The argument is analogous to that in KP. For each pair of distinct points t_0 and t_1, instead of taking the supremum of $Z(t)$ over neighborhoods N_0 of t_0 and N_1 of t_1 as in KP, take the supremum of $Z(t)$ over N_0 and the supremum of $-Z(t)$ over N_1. Using the notation in KP, $Cov(Z(t_0), -Z(t_1)) = -H(t_0, t_1)$. By (12.77), $-H(t_0, t_1)$ cannot equal both $H(t_0, t_0)$ and $H(t_1, t_1)$. Suppose $H(t_0, t_0) > -H(t_0, t_1)$ (the other cases are handled similarly), then $h(t_0) = 1 > -h(t_1)$, where $h(t) = H(t_1, t_0)/H(t_0, t_0)$ as in KP. The rest of the proof is the same as in KP, except that $\beta_1 = \sup_{t \in N_1}(h(t))$ and $\Gamma_1(z) = \sup_{t \in N_1}(Y(t) + h(t)z)$ are changed to $\beta_1 = \sup_{t \in N_1}(-h(t))$ and $\Gamma_1(z) = \sup_{t \in N_1}(-Y(t) - h(t)z)$, respectively. This leads to the desired result $P\{\sup_{t \in N_0} Z(t) = \sup_{t \in N_1}(-Z(t))\} = 0$. □

Proof of Lemma 4.1. For any $\pi_1, \pi_2 \in \Pi$,

$$\begin{aligned} & Var(G_1^*(\pi_1; \gamma_0) - G_2^*(\pi_2; \gamma_0)) \\ = & Var(G_1(\pi_1) - G_2(\pi_2) - (H_{12}(\pi_1) - H_{12}(\pi_2))H_{22}^{-1}G_2) \\ = & a'\Omega_G(\pi_1, \pi_2; \gamma_0)a > 0, \end{aligned} \tag{12.78}$$

where $a = (1, -1, -(H_{12}(\pi_1) - H_{12}(\pi_2))H_{22}^{-1})'$ and the inequality holds by Assumption C6**(ii). Similarly, we can show that $Var(G_1^*(\pi_1; \gamma_0) + G_1^*(\pi_2; \gamma_0)) \neq 0 \ \forall \pi_1, \pi_2 \in \Pi$ with $\pi_1 \neq \pi_2$. Hence, Assumption C6* holds. By Lemma 11.4, Assumption C6 holds as well. □

12.3.4. Quadratic Expansions: Assumptions C1 and D1

Proof of Lemma 11.5. We first prove part (a). Let δ_n be any sequence of constants such that $\delta_n \to 0$ as $n \to \infty$. By a second-order Taylor expansion of $Q_n(\psi, \pi)$ about $\psi_{0,n}$, for $\psi \in \Psi(\pi)$ with $||\psi - \psi_{0,n}|| \leq \delta_n$ and $\pi \in \Pi$, we have

$$
\begin{aligned}
&|R_n(\psi, \pi)| \\
&= \left| \frac{1}{2}(\psi - \psi_{0,n})' \left(n^{-1} \sum_{i=1}^{n} \left(\rho_{\psi\psi}(W_i, \psi_{0,n}^{\dagger}(\pi), \pi) - \rho_{\psi\psi}(W_i, \psi_{0,n}, \pi) \right) \right) (\psi - \psi_{0,n}) \right| \\
&\leq \left\| (\psi - \psi_{0,n}) \right\|^2 \left\| n^{-1} \sum_{i=1}^{n} \left(\rho_{\psi\psi}(W_i, \psi_{0,n}^{\dagger}(\pi), \pi) - \rho_{\psi\psi}(W_i, \psi_{0,n}, \pi) \right) \right\| \\
&= o_{p\pi}(\left\| \psi - \psi_{0,n} \right\|^2), \qquad (12.79)
\end{aligned}
$$

where $\psi_{0,n}^{\dagger}(\pi)$ lies between ψ and $\psi_{0,n}$ and the $o_{p\pi}(\left\| \psi - \psi_{0,n} \right\|^2)$ term follows from Assumption Q1(iii). This immediately implies Assumption C1 using the "$||a_n(\gamma_n)(\psi - \psi_{0,n})||$" part of the denominator in Assumption C1(ii).

Next, we show part (b). By a second-order Taylor expansion of $Q_n(\theta)$ wrt θ,

$$
\begin{aligned}
|R_n^*(\theta)| &= \left| \frac{1}{2}(\theta - \theta_n)' \left(n^{-1} \sum_{i=1}^{n} \left(\rho_{\theta\theta}(W_i, \theta_n^{\dagger}) - \rho_{\theta\theta}(W_i, \theta_n) \right) \right) (\theta - \theta_n) \right| \\
&= \left| \frac{1}{2}(B(\beta_n)(\theta - \theta_n))'[B^{-1}(\beta_n) n^{-1} \sum_{i=1}^{n} (\rho_{\theta\theta}(W_i, \theta_n^{\dagger}) - \rho_{\theta\theta}(W_i, \theta_n)) B^{-1}(\beta_n)] \times \right. \\
&\qquad \left. B(\beta_n)(\theta - \theta_n) \right| \\
&\leq ||B(\beta_n)(\theta - \theta_n)||^2 ||B^{-1}(\beta_n) n^{-1} \sum_{i=1}^{n} (\rho_{\theta\theta}(W_i, \theta_n^{\dagger}) - \rho_{\theta\theta}(W_i, \theta_n)) B^{-1}(\beta_n)|| \\
&= o_p(||B(\beta_n)(\theta - \theta_n)||^2), \qquad (12.80)
\end{aligned}
$$

where $\theta_n^{\dagger}$ is between θ and θ_n and the $o_p(||B(\beta_n)(\theta - \theta_n)||^2)$ term follows from Assumption Q1(iv). This immediately implies Assumption D1 using the "$||n^{1/2} B(\beta_n)(\theta - \theta_n)||$" part of the denominator in Assumption D1(ii). □

Proof of Lemma 11.6. We first prove part (a). For any function $f(w, \theta)$, define the empirical process $\{\nu_n f(\theta) : \theta \in \Theta\}$ by $\nu_n f(\theta) = n^{-1/2} \sum_{i=1}^{n} \left(f(W_i, \theta) - E_{\gamma_n} f(W_i, \theta) \right)$.

Note that

$$Q_n(\theta) - Q_n\left(\psi_{0,n}, \pi\right) = n^{-1/2}\left(\nu_n \rho(\theta) - \nu_n \rho(\psi_{0,n}, \pi)\right) + Q_n^*(\theta) - Q_n^*(\psi_{0,n}, \pi). \quad (12.81)$$

The expansion in (11.5) implies that

$$\nu_n \rho(\theta) - \nu_n \rho\left(\psi_{0,n}, \pi\right) = \nu_n \Delta_\psi(\psi_{0,n}, \pi)'(\psi - \psi_{0,n}) + \nu_n r_\psi(\theta). \quad (12.82)$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, a second-order Taylor expansion of $Q_n^*(\theta)$ wrt ψ gives

$$\begin{aligned} Q_n^*(\theta) - Q_n^*(\psi_{0,n}, \pi) &= \frac{\partial}{\partial \psi} Q_n^*(\psi_{0,n}, \pi)'(\psi - \psi_{0,n}) + \\ \frac{1}{2}(\psi - \psi_{0,n})' &\left(\frac{\partial^2}{\partial \psi \partial \psi'} Q_n^*(\psi_{0,n}, \pi)\right)(\psi - \psi_{0,n}) + o_\pi(||\psi - \psi_{0,n}||^2) \end{aligned} \quad (12.83)$$

using Assumption Q2(v) (where $o_\pi(\cdot)$ denotes $o(\cdot)$ uniformly over $\pi \in \Pi$). From (12.81)-(12.83), we have

$$\begin{aligned} Q_n(\theta) - Q_n\left(\psi_{0,n}, \pi\right) &= \left(n^{-1/2}\nu_n \Delta_\psi(\psi_{0,n}, \pi) + \frac{\partial}{\partial \psi} Q_n^*(\psi_{0,n}, \pi)\right)'(\psi - \psi_{0,n}) + \\ \frac{1}{2}(\psi - \psi_{0,n})' \frac{\partial^2}{\partial \psi \partial \psi'} Q_n^*(\psi_{0,n}, \pi)(\psi - \psi_{0,n}) &+ n^{-1/2}\nu_n r_\psi(\theta) + o_\pi(\left\|\psi - \psi_{0,n}\right\|^2). \end{aligned} \quad (12.84)$$

When $D_\psi Q_n(\theta)$ and $D_{\psi\psi} Q_n(\theta)$ take the form as in Lemma 11.6(a), the quadratic approximation in Assumption C1(i) holds with

$$R_n(\psi, \pi) = n^{-1/2}\nu_n r_\psi(\theta) + o_\pi(\left\|\psi - \psi_{0,n}\right\|^2). \quad (12.85)$$

To verify Assumption C1(ii), we have

$$\begin{aligned} &\sup_{\psi \in \Psi(\pi): ||\psi - \psi_{0,n}|| \leq \delta_n} \frac{|a_n^2(\gamma_n) R_n(\psi, \pi)|}{(1 + \left\|a_n(\gamma_n)\left(\psi - \psi_{0,n}\right)\right\|)^2} \\ \leq &\sup_{\psi \in \Psi(\pi): ||\psi - \psi_{0,n}|| \leq \delta_n} \frac{|a_n^2(\gamma_n) n^{-1/2}\nu_n r_\psi(\theta)|}{(1 + \left\|a_n(\gamma_n)\left(\psi - \psi_{0,n}\right)\right\|)^2} + o_\pi(1) = o_{p\pi}(1), \end{aligned} \quad (12.86)$$

where the inequality follows from (12.85) and the triangle inequality and the equality is implied by Assumption Q2(iii) by using $[1 + ||a_n(\gamma_n)(\psi - \psi_{0,n})||] \cdot ||a_n(\gamma_n)(\psi - \psi_{0,n})||$ in the denominator.

Next, we prove part (b). The sample criterion function satisfies

$$Q_n(\theta) - Q_n(\theta_n) = n^{-1/2}\left(\nu_n\rho\left(\theta\right) - \nu_n\rho(\theta_n)\right) + Q_n^*(\theta) - Q_n^*(\theta_n). \tag{12.87}$$

The expansion in (11.4) gives

$$\nu_n\rho\left(\theta\right) - \nu_n\rho(\theta_n) = \nu_n\Delta(\theta_n)'(\theta - \theta_n) + \nu_n r(\theta). \tag{12.88}$$

A second-order Taylor expansion of $Q_n^*(\theta)$ about θ_n gives

$$Q_n^*(\theta) - Q_n^*(\theta_n) = \frac{\partial}{\partial\theta}Q_n^*(\theta_n)'(\theta - \theta_n) + \frac{1}{2}(\theta - \theta_n)'\frac{\partial^2}{\partial\theta\partial\theta'}Q_n^*(\theta_n^\dagger)(\theta - \theta_n), \tag{12.89}$$

where $\theta_n^\dagger$ is between θ and θ_n. By Assumption Q2(vi),

$$B^{-1}(\beta_n)\frac{\partial^2}{\partial\theta\partial\theta'}Q_n^*(\theta_n^\dagger)B^{-1}(\beta_n) = B^{-1}(\beta_n)\frac{\partial^2}{\partial\theta\partial\theta'}Q_n^*(\theta_n)B^{-1}(\beta_n) + o\left(1\right), \tag{12.90}$$

where the $o\left(1\right)$ term holds uniformly over $\theta \in \Theta_n(\delta_n)$.

Equations (12.87)-(12.90) yield

$$\begin{aligned} Q_n(\theta) - Q_n(\theta_n) = & \left(n^{-1/2}\nu_n\Delta(\theta_n) + \frac{\partial}{\partial\theta}Q_n^*(\theta_n)\right)'(\theta - \theta_n) + \\ & \frac{1}{2}(\theta - \theta_n)'\frac{\partial^2}{\partial\theta\partial\theta'}Q_n^*(\theta_n)(\theta - \theta_n) + n^{-1/2}\nu_n r(\theta) + o(\|B(\beta_n)(\theta - \theta_n)\|^2). \end{aligned} \tag{12.91}$$

When $DQ_n(\theta)$ and $D^2Q_n(\theta)$ take the form in Lemma 11.6(b), the quadratic approximation in Assumption D1 holds with

$$R_n^*(\theta) = n^{-1/2}\nu_n r(\theta) + o(\|B(\beta_n)(\theta - \theta_n)\|^2). \tag{12.92}$$

To verify Assumption D1(ii), we have

$$\begin{aligned} & \sup_{\theta\in\Theta_n(\delta_n)} \frac{|nR_n^*(\theta)|}{(1 + n^{1/2}||B(\beta_n)(\theta - \theta_n)||)^2} \\ \leq & \sup_{\theta\in\Theta_n(\delta_n)} \frac{|n^{1/2}\nu_n r(\theta)|}{(1 + n^{1/2}||B(\beta_n)(\theta - \theta_n)||)^2} + o(1) = o_p\left(1\right), \end{aligned} \tag{12.93}$$

where the inequality holds by (12.92) and the triangle inequality and the equality is implied by Assumption Q2(iv) by using $[1+n^{1/2}||B(\beta_n)(\theta-\theta_n)||]\cdot n^{1/2}||B(\beta_n)(\theta-\theta_n)||$ in the denominator. □

Proof of Lemma 11.7. Lemma 11.7(a) is proved using the proof of Lemma 11.5 with (12.79) and (12.80) changed to

$$\begin{aligned}|R_n(\psi,\pi)| &\le o_{p\pi}(||\psi-\psi_{0,n}||^2)+|Q_n^{IC}(\psi,\pi)-Q_n^{IC}(\psi_{0,n},\pi)| \text{ and}\\ |R_n^*(\theta)| &\le o_p(||B(\beta_n)(\theta-\theta_n)||^2)+|Q_n^{IC}(\theta)-Q_n^{IC}(\theta_n)|,\end{aligned} \tag{12.94}$$

respectively. By Assumption Q3(ii), Assumptions C1 and D1 follow from the same arguments as those in the proof of Lemma 11.5.

Lemma 11.7(b) is proved using the proof of Lemma 11.6 with (12.85) and (12.92) changed to

$$\begin{aligned}R_n(\psi,\pi) &= n^{-1/2}\nu_n r_\psi(\theta)+o_\pi(||\psi-\psi_{0,n}||^2)+Q_n^{IC}(\psi,\pi)-Q_n^{IC}(\psi_{0,n},\pi) \text{ and}\\ R_n^*(\theta) &= n^{-1/2}\nu_n r(\theta)+o(\|B(\beta_n)(\theta-\theta_n)\|^2)+Q_n^{IC}(\theta)-Q_n^{IC}(\theta_n),\end{aligned} \tag{12.95}$$

respectively. By Assumption Q3(ii), Assumptions C1 and D1 follow from the same arguments as those in the proof of Lemma 11.6. □

13. Appendix C: Verification of Assumptions for the ARMA(1, 1) Example

This Appendix verifies the assumptions of AC1 for the ARMA(1, 1) example of Section 9.

First, we give some details concerning the form of the criterion function $Q_n(\theta)$ for this example. To specify the quasi-log likelihood function, it is useful to write the innovations as a function of the observations and the unknown parameters. By repeated substitution for $\varepsilon_{t-1},...,\varepsilon_1$ in (3.2), we have

$$\varepsilon_t=\sum_{j=0}^{t-1}\pi_0^j(Y_{t-j}-\rho_0Y_{t-j-1})+\pi_0^t\varepsilon_0. \tag{13.1}$$

The Gaussian quasi-log likelihood function for $\theta=(\beta,\zeta,\pi)$ conditional on Y_0 and ε_0

is a constant plus

$$-\frac{n}{2}\log\zeta-\frac{1}{2\zeta}\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1}\pi^{j}[Y_{t-j}-(\pi+\beta)Y_{t-j-1}]+\pi^{t}\varepsilon_{0}\right)^{2}. \tag{13.2}$$

The conditioning value ε_0 is asymptotically negligible, so for simplicity (and wlog for the asymptotic results) we set $\varepsilon_0 = Y_0$ in the log likelihood. Thus, the (conditional) QML criterion function for $\theta = (\beta, \zeta, \pi)'$ (multiplied by $-n^{-1}$ and ignoring a constant) is

$$Q_n(\theta)=\frac{1}{2}\log\zeta+\frac{1}{2\zeta}n^{-1}\sum_{t=1}^{n}\left(Y_t-\beta\sum_{j=0}^{t-1}\pi^{j}Y_{t-j-1}\right)^{2}. \tag{13.3}$$

13.1. ARMA Example: Initial Conditions Adjustment

We use the initial conditions adjustment of the criterion function given in Lemma 11.7(a) of Section 11.4.3. This Lemma implies that it suffices to establish Assumptions C1-C8 and D1-D3 with $Q_n(\theta)$ replaced by an approximation $Q_n^{\infty}(\theta)$. Lemma 11.7(a) relies on Assumption Q3. We verify Assumption Q3 with

$$\begin{aligned}
Q_n^{\infty}(\theta) &= n^{-1}\sum_{t=1}^{n}\rho_t(\theta), \text{ where}\\
\rho_t(\theta) &= \frac{1}{2}\log\zeta+\frac{1}{2\zeta}\left(Y_t-\beta\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)^{2}, \text{ and}\\
Q_n^{IC}(\theta) &= Q_n(\theta)-Q_n^{\infty}(\theta) \qquad (13.4)\\
&= -\frac{\beta^2}{2\zeta}n^{-1}\sum_{t=1}^{n}\left(\sum_{j=t}^{\infty}\pi^{j}Y_{t-j-1}\right)^{2}+\frac{\beta}{\zeta}n^{-1}\sum_{t=1}^{n}\left(Y_t-\beta\sum_{j=0}^{t-1}\pi^{j}Y_{t-j-1}\right)\sum_{j=t}^{\infty}\pi^{j}Y_{t-j-1}.
\end{aligned}$$

Note that the difference between $Q_n^{\infty}(\theta)$ and $Q_n(\theta)$ is that the sum over j goes to ∞ in the former and to $t-1$ in the latter. In (13.4), $W_t = (Y_t, Y_{t-1})'$ and $\rho_t(\theta)$ depends not only on W_t but also on $W_{t-1}, ..., W_1$. This does not affect the results in Lemma 11.7(a).

Lemma 13.1. *For the* ARMA(1, 1) *model,* $\{Q_n^{IC}(\theta) : n \geq 1\}$ *satisfies*

(a) *under* $\{\gamma_n\} \in \Gamma(\gamma_0)$, $\sup_{\theta\in\Theta}|Q_n^{IC}(\theta)| \to_p 0$,

(b) *under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$,

$$\sup_{\psi \in \Psi(\pi): ||\psi - \psi_{0,n}|| \leq \delta_n} \frac{|a_n^2(\gamma_n)(Q_n^{IC}(\psi, \pi) - Q_n^{IC}(\psi_{0,n}, \pi))|}{(1 + a_n(\gamma_n)||\psi - \psi_{0,n}||)^2} = o_{p\pi}(1)$$

for all constants $\delta_n \to 0$, *and*

(c) *under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,

$$\sup_{\theta \in \Theta_n(\delta_n)} \frac{|n(Q_n^{IC}(\theta) - Q_n^{IC}(\theta_n))|}{(1 + ||n^{1/2} B(\beta_n)(\theta - \theta_n)||)^2} = o_p(1)$$

for all $\delta_n \to 0$, *where* $\Theta_n(\delta_n) = \{\theta \in \Theta : \|\psi - \psi_n\| \leq \delta_n |\beta_n|$ *and* $|\pi - \pi_n| \leq \delta_n\}$.

Comments. 1. Lemma 13.1(a) implies that it suffices to establish Assumption B3 with $Q_n^\infty(\theta)$ in place of $Q_n(\theta)$.

2. Assumption Q3 holds by Lemma 13.1(b) and 13.1(c).

The proof of Lemma 13.1 is given in Section 13.4 below.

13.2. ARMA Example: Derivation of Formulae for Key Quantities

The quantities that appear in Assumptions B1-B3, C1-C8, and D1-D3, viz., $Q(\theta; \gamma_0)$, $D_\psi Q_n(\theta)$, $\Omega(\pi_1, \pi_2; \gamma_0)$, $D_{\psi\psi} Q_n(\theta)$, $H(\pi; \gamma_0)$, $K(\pi; \gamma_0)$, $\Omega_G(\pi_1, \pi_2; \gamma_0)$, $DQ_n(\theta)$, $D^2 Q_n(\theta)$, $J(\gamma_0)$, and $V(\gamma_0)$, are specified in Section 4 of AC1. In this section, we derive the formulae for these quantities based on the criterion function $Q_n^\infty(\theta) = n^{-1} \sum_{t=1}^n \rho_t(\theta)$. (For convenience, the formula for $K(\pi; \gamma_0)$ is derived in Section 13.3.4 below.)

The expressions for $D_\psi Q_n(\theta)$ and $D_{\psi\psi} Q_n(\theta)$ are the ordinary first and second partial derivatives of $n^{-1} \sum_{t=1}^n \rho_t(\theta)$ wrt ψ for $\rho_t(\theta)$ defined in (13.4). Analogously, $DQ_n(\theta)$ and $D^2 Q_n(\theta)$ are the ordinary first and second partial derivatives of $n^{-1} \sum_{t=1}^n \rho_t(\theta)$ wrt θ.

Now, we derive the formula for $\Omega(\pi_1, \pi_2; \gamma_0)$. For any sequence $\{\gamma_n\} \in \Gamma(\gamma_0)$ with

$\beta_0 = 0$, we have

$$\begin{aligned}\Omega(\pi_1, \pi_2; \gamma_0) &= \lim_{n\to\infty} Cov_{\gamma_n}\left(n^{-1/2}\sum_{t=1}^{n}\rho_{\psi,t}(\psi_{0,n}, \pi_1),\ n^{-1/2}\sum_{t=1}^{n}\rho_{\psi,t}(\psi_{0,n}, \pi_2)\right) \\ &= \sum_{m=-\infty}^{\infty} Cov_{\gamma_0}(\rho_{\psi,t}(\psi_0, \pi_1), \rho_{\psi,t+m}(\psi_0, \pi_2)) \\ &= Cov_{\gamma_0}(\rho_{\psi,t}(\psi_0, \pi_1), \rho_{\psi,t}(\psi_0, \pi_2)) \\ &= \begin{bmatrix} (1-\pi_1\pi_2)^{-1} & 0 \\ 0 & (1/4)\zeta_0^{-4}E_{\gamma_0}\left(\varepsilon_t^2 - \zeta_0\right)^2 \end{bmatrix}, \end{aligned} \tag{13.5}$$

where the first equality holds by the definition of $G_n(\pi)$ in Assumption C3 with $\psi_{0,n} = (0, \zeta_n)$, the second equality holds by strict stationarity for given γ_n and $\gamma_n \to \gamma_0$, and the third and fourth equalities hold because $\{\varepsilon_t : t \geq 1\}$ are independent and have mean zero plus

$$\begin{aligned}\rho_{\beta,t}(\psi_0, \pi) &= -\zeta_0^{-1}\varepsilon_t\sum_{j=0}^{\infty}\pi^j\varepsilon_{t-j-1} \text{ and} \\ \rho_{\zeta,t}(\psi_0, \pi) &= -(1/2)\zeta_0^{-2}\left(\varepsilon_t^2 - \zeta_0\right)\end{aligned} \tag{13.6}$$

when the true parameter is γ_0 with $\beta_0 = 0$, using the definitions of $\rho_{\beta,t}(\theta)$ and $\rho_{\zeta,t}(\theta)$ in (4.8). The off-diagonal elements in (13.5) are zero because $E_{\gamma_0}\varepsilon_t(\varepsilon_t^2 - \zeta_0)\varepsilon_{t-j-1} = E_{\gamma_0}\varepsilon_t(\varepsilon_t^2 - \zeta_0)E_{\gamma_0}\varepsilon_{t-j-1} = 0\ \forall j \geq 0$.

Next, we derive the formula for $H(\pi; \gamma_0)$, which is shown in Section 13.3.3 to equal $E_{\gamma_0}\rho_{\psi\psi,t}(\psi_0, \pi)$. Using the definitions of $\rho_{\psi\psi,t}(\theta), ..., \rho_{\zeta\zeta,t}(\theta)$ in (4.15), when the true parameter is γ_0 with $\beta_0 = 0$, we have

$$\begin{aligned}\rho_{\beta\beta,t}(\psi_0, \pi) &= \zeta_0^{-1}\left(\sum_{j=0}^{\infty}\pi^j\varepsilon_{t-j-1}\right)^2,\ \rho_{\beta\zeta,t}(\psi_0, \pi) = \zeta_0^{-2}\varepsilon_t\sum_{j=0}^{\infty}\pi^j\varepsilon_{t-j-1},\ \text{and} \\ \rho_{\zeta\zeta,t}(\psi_0, \pi) &= -(1/2)\zeta_0^{-2} + \zeta_0^{-3}\varepsilon_t^2.\end{aligned} \tag{13.7}$$

Using these expressions, we obtain

$$H(\pi;\gamma_0) = E_{\gamma_0}\rho_{\psi\psi,t}(\psi_0,\pi) = \begin{bmatrix} \zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi^j\varepsilon_{t-j-1}\right)^2 & 0 \\ 0 & (2\zeta_0^2)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=0}^{\infty}\pi^{2j} & 0 \\ 0 & (2\zeta_0^2)^{-1} \end{bmatrix} = \begin{bmatrix} (1-\pi^2)^{-1} & 0 \\ 0 & (2\zeta_0^2)^{-1} \end{bmatrix}. \quad (13.8)$$

Now, we calculate the matrix $\Omega_G(\pi_1,\pi_2;\gamma_0)$. For $\beta_0 = 0$, we define

$$\rho^*_{\psi,t}(\psi_0,\pi_1,\pi_2) = (\rho_{\beta,t}(\psi_0,\pi_1), \rho_{\beta,t}(\psi_0,\pi_2), \rho_{\zeta,t}(\psi_0,\pi)')', \text{ where}$$
$$\rho_{\beta,t}(\psi_0,\pi) = -\zeta_0^{-1}\varepsilon_t\sum_{k=0}^{\infty}\pi^k Y_{t-k-1} = -\zeta_0^{-1}\varepsilon_t\sum_{k=0}^{\infty}\pi^k\varepsilon_{t-k-1}, \text{ and}$$
$$\rho_{\zeta,t}(\psi_0,\pi) = -(1/2)\zeta_0^{-2}(\varepsilon_t^2 - \zeta_0). \quad (13.9)$$

Using these definitions, for $\beta_0 = 0$, we have

$$\Omega_G(\pi_1,\pi_2;\gamma_0) = \sum_{m=-\infty}^{\infty} Cov_{\gamma_0}(\rho^*_{\psi,t}(\psi_0,\pi_1,\pi_2), \rho^*_{\psi,t+m}(\psi_0,\pi_1,\pi_2))$$
$$= Var_{\gamma_0}(\rho^*_{\psi,t}(\psi_0,\pi_1,\pi_2))$$
$$= \begin{bmatrix} \zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_1^j\varepsilon_{t-j-1}\right)^2 & \zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_1^j\varepsilon_{t-j-1}\right)\left(\sum_{j=0}^{\infty}\pi_2^j\varepsilon_{t-j-1}\right) & \cdots \\ \zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_1^j\varepsilon_{t-j-1}\right)\left(\sum_{j=0}^{\infty}\pi_2^j\varepsilon_{t-j-1}\right) & \zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_2^j\varepsilon_{t-j-1}\right)^2 & \vdots \\ 0 & 0 & \cdots \end{bmatrix}$$
$$\begin{bmatrix} \cdots & 0 \\ \vdots & 0 \\ \cdots & (1/4)\zeta_0^{-4}E_{\gamma_0}(\varepsilon_t^2-\zeta_0)^2 \end{bmatrix}$$
$$= \begin{bmatrix} (1-\pi_1^2)^{-1} & (1-\pi_1\pi_2)^{-1} & 0 \\ (1-\pi_1\pi_2)^{-1} & (1-\pi_2^2)^{-1} & 0 \\ 0 & 0 & (1/4)\zeta_0^{-4}E_{\gamma_0}(\varepsilon_t^2-\zeta_0)^2 \end{bmatrix}. \quad (13.10)$$

The second and third equalities of (13.10) hold using (13.9) and $E_{\gamma_0}\varepsilon_t(\varepsilon_t^2-\zeta_0)\varepsilon_{t-j-1} = E_{\gamma_0}\varepsilon_t(\varepsilon_t^2-\zeta_0)E_{\gamma_0}\varepsilon_{t-j-1} = 0 \ \forall j \geq 0$.

To determine $J(\gamma_0)$, we first provide the (generalized) second derivative matrix

$D^2Q_n(\theta)$:

$$D^2Q_n(\theta) = n^{-1}\sum_{t=1}^{n}\rho_{\theta\theta,t}(\theta) = n^{-1}\sum_{t=1}^{n}\begin{bmatrix}\rho_{\beta\beta,t}(\theta) & \rho_{\beta\zeta,t}(\theta) & \rho_{\beta\pi,t}(\theta)\\ \rho_{\beta\zeta,t}(\theta) & \rho_{\zeta\zeta,t}(\theta) & \rho_{\zeta\pi,t}(\theta)\\ \rho_{\beta\pi,t}(\theta) & \rho_{\zeta\pi,t}(\theta) & \rho_{\pi\pi,t}(\theta)\end{bmatrix}, \tag{13.11}$$

where

$$\begin{aligned}
\rho_{\beta\beta,t}(\theta) &= \zeta^{-1}\left(\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)^2,\\
\rho_{\beta\zeta,t}(\theta) &= \zeta^{-2}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)\sum_{k=0}^{\infty}\pi^k Y_{t-k-1},\\
\rho_{\beta\pi,t}(\theta) &= \zeta^{-1}\left(\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)\beta\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1}\\
&\quad -\zeta^{-1}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1},
\end{aligned} \tag{13.12}$$

and

$$\begin{aligned}
\rho_{\zeta\zeta,t}(\theta) &= -(1/2)\zeta^{-2} + \zeta^{-3}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)^2,\\
\rho_{\zeta\pi,t}(\theta) &= \zeta^{-2}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)\beta\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1},\\
\rho_{\pi\pi,t}(\theta) &= \zeta^{-1}\left(\beta\sum_{j=0}^{\infty}j\pi^{j-1}Y_{t-j-1}\right)\beta\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1}\\
&\quad -\zeta^{-1}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)\beta\sum_{k=0}^{\infty}k(k-1)\pi^{k-2}Y_{t-k-1}.
\end{aligned} \tag{13.13}$$

To determine $J(\gamma_0)$ via the expression $J(\gamma_0) = E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0)$ given in (13.51) below (in the verification of Assumption D2), we define $\rho^{\dagger}_{\theta\theta,t}(\theta)$ and $\chi_t(\theta)$ via

$$B^{-1}(\beta)\rho_{\theta\theta,t}(\theta)B^{-1}(\beta) = \rho^{\dagger}_{\theta\theta,t}(\theta) + \beta^{-1}\chi_t(\theta), \tag{13.14}$$

where $\rho_{\theta\theta,t}(\theta)$ is defined in (13.11)-(13.13) and $\rho^{\dagger}_{\theta\theta,t}(\theta)$ is defined by

$$\rho^{\dagger}_{\theta\theta,t}(\theta) = \begin{bmatrix} \rho_{\beta\beta,t}(\theta) & \rho_{\beta\zeta,t}(\theta) & \rho^{\dagger}_{\beta\pi,t}(\theta) \\ \rho_{\beta\zeta,t}(\theta) & \rho_{\zeta\zeta,t}(\theta) & \rho^{\dagger}_{\zeta\pi,t}(\theta) \\ \rho^{\dagger}_{\beta\pi,t}(\theta) & \rho^{\dagger}_{\zeta\pi,t}(\theta) & \rho^{\dagger}_{\pi\pi,t}(\theta) \end{bmatrix},$$

$$\rho^{\dagger}_{\beta\pi,t}(\theta) = \zeta^{-1}\left(\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1},$$

$$\rho^{\dagger}_{\zeta\pi,t}(\theta) = \beta^{-1}\rho_{\zeta\pi,t}(\theta) = \zeta^{-2}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1}, \text{ and}$$

$$\rho^{\dagger}_{\pi\pi,t}(\theta) = \zeta^{-1}\left(\sum_{j=0}^{\infty}j\pi^{j-1}Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1}. \tag{13.15}$$

The matrix $\chi_t(\theta)$ is defined by

$$\chi_t(\theta) = \begin{bmatrix} 0 & 0 & \chi_{\beta\pi,t}(\theta) \\ 0 & 0 & 0 \\ \chi_{\beta\pi,t}(\theta) & 0 & \chi_{\pi\pi,t}(\theta) \end{bmatrix}, \text{ where}$$

$$\chi_{\beta\pi,t}(\theta) = -\zeta^{-1}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1} \text{ and}$$

$$\chi_{\pi\pi,t}(\theta) = -\zeta^{-1}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)\sum_{k=0}^{\infty}k(k-1)\pi^{k-2}Y_{t-k-1}. \tag{13.16}$$

Now, using $J(\gamma_0) = E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0)$ and (13.12), (13.13), and (13.15), we have

$$\begin{aligned} J(\gamma_0) &= E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0) \\ &= Diag\left\{\zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_0^{j}Y_{t-j-1}\right)^2,\ \frac{1}{2\zeta_0^2},\ \zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}j\pi_0^{j-1}Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi_0^{k-1}Y_{t-k-1}\right\} \\ &\quad + \left(\zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_0^{j}Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi_0^{k-1}Y_{t-k-1}\right) \times \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{13.17}$$

As shown in Section 13.3.7 below, the matrix $n^{-1}\sum_{t=1}^{n}\beta^{-1}\chi_t(\theta)$ evaluated at $\theta = \theta_n$ $(\rightarrow \theta_0)$ does not contribute to $J(\gamma_0)$ because its probability limit is zero.

To derive the formulae for $V(\gamma_0)$, we define

$$
\begin{aligned}
\rho^{\dagger}_{\theta,t}(\theta) &= B^{-1}(\beta)\rho_{\theta,t}(\theta) = (\rho_{\beta,t}(\theta), \rho_{\zeta,t}(\theta), \beta^{-1}\rho_{\pi,t}(\theta))' \text{ and} \\
V^{\dagger}(\theta_1,\theta_2;\gamma_0) &= \sum_{m=-\infty}^{\infty} Cov_{\gamma_0}(\rho^{\dagger}_{\theta,t}(\theta_1), \rho^{\dagger}_{\theta,t+m}(\theta_2)).
\end{aligned} \tag{13.18}
$$

For any sequence $\{\gamma_n\} \in \Gamma(\gamma_0)$, we have

$$
\begin{aligned}
V(\gamma_0) &= \lim_{n\to\infty} Var_{\gamma_n}\left(n^{1/2}B^{-1}(\beta_n)DQ_n(\theta_n)\right) \\
&= \lim_{n\to\infty} Var_{\gamma_n}\left(n^{-1/2}\sum_{t=1}^{n}\rho^{\dagger}_{\theta,t}(\theta_n)\right) \\
&= V^{\dagger}(\theta_0,\theta_0;\gamma_0) \\
&= Var_{\gamma_0}(\rho^{\dagger}_{\theta,t}(\theta_0)) \\
&= Diag\left\{\zeta_0^{-1}E_{\gamma_0}\left(\sum_{k=0}^{\infty}\pi_0^k Y_{t-k-1}\right)^2, (1/4)\zeta_0^{-4}E_{\gamma_0}\left(\varepsilon_t^2-\zeta_0\right)^2,\right. \\
&\qquad\qquad \left.\zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty} j\pi_0^{j-1}Y_{t-j-1}\right)^2\right\} \\
&\quad + \left(\zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_0^j Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi_0^{k-1}Y_{t-k-1}\right) \times \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},
\end{aligned} \tag{13.19}
$$

where the first equality holds because the convergence in distribution result in Assumption D3(i) is obtained by a CLT, see (13.56) below, the second equality holds by definition, and the third equality holds by strict stationarity for given γ_n, $\gamma_n \to \gamma_0$, and the continuity of $E_{\gamma_0}\rho^{\dagger}_{\theta,t}(\theta_0)\rho^{\dagger}_{\theta,t}(\theta_0)'$ in $\gamma_0 = (\theta_0, \phi_0)$, which follows straightforwardly from the form of $\rho^{\dagger}_{\theta,t}(\theta_0)$ given in (13.20) below. The last two equalities in (13.19) hold because

$$
\begin{aligned}
&\rho_{\beta,t}(\theta_0) = -\zeta_0^{-1}\varepsilon_t\sum_{j=0}^{\infty}\pi_0^j Y_{t-j-1},\ \rho_{\zeta,t}(\theta_0) = -(1/2)\zeta_0^{-2}\left(\varepsilon_t^2-\zeta_0\right), \\
&\rho^{\dagger}_{\pi,t}(\theta_0) = -\zeta_0^{-1}\varepsilon_t\sum_{j=0}^{\infty}j\pi_0^{j-1}Y_{t-j-1},\ \text{and } E_{\gamma_0}\varepsilon_t\left(\varepsilon_t^2-\zeta_0\right)Y_{t-k-1} = 0\ \forall k \geq 0,
\end{aligned} \tag{13.20}
$$

where the last equality holds because ε_t and Y_{t-j-1} are independent and $E_{\gamma_0} Y_{t-j-1} = 0$.

13.3. ARMA Example: Verification of Assumptions

Here, we verify Assumptions A, B1-B3, C1-C8, and D1-D3 for the criterion function $Q_n^\infty(\theta) = n^{-1} \sum_{t=1}^n \rho_t(\theta)$.

13.3.1. ARMA Example: Verification of Assumptions A and B1-B3

Assumption A holds immediately given the definition of $\rho_t(\theta)$ in (13.4).

Assumption B1(i) holds by the definitions of Θ and Θ^* in (4.2) and (4.3). Assumption B1(ii) holds with $\mathcal{Z}^0 = (\zeta_L^{**}, \zeta_U^{**})$, where ζ_J^{**} is between ζ_J and ζ_J^* for $J = L, U$, using the fact that $\rho_L < \pi_L$ and $\rho_U > \pi_U$ imply that, for $\theta = (\beta, \zeta, \pi) \in \Theta$, β can take values in a neighborhood of zero for any value of $\pi \in \Pi$. Assumption B1(iii) holds by the definition of Π in (4.2).

Assumption B2(i) holds by the definition of Γ in (4.4). Assumption B2(ii) holds by the definitions of Γ and Θ^* and the condition $\rho_L^* < \pi_L^* < \pi_U^* < \rho_U^*$. Assumption B2(iii) holds by the definitions of Γ and Θ^* and the conditions $\rho_L^* < \pi_L^*$ and $\pi_U^* < \rho_U^*$, which guarantee that, for $\theta = (\beta, \zeta, \pi) \in \Theta^*$, $\theta_a = (a\beta, \zeta, \pi) \in \Theta^*$ $\forall a \in [0, 1]$.

Assumption B3(i) holds with $Q(\theta; \gamma_0) = E_{\gamma_0} \rho_t(\theta)$ by the following argument. By Theorem 1 of Andrews (1992), uniform convergence in probability is implied by pointwise convergence in probability, stochastic equicontinuity, and boundedness of Θ. Pointwise convergence in probability is implied by mean square convergence. In the present case, the latter is straightforward, but tedious, to establish by writing out the square that appears in $\rho_t(\theta)$, using the expression $Y_t = \sum_{j=0}^\infty (\pi_n + \beta_n)^j (\varepsilon_{t-j-1} - \pi_n \varepsilon_{t-j-2})$ under γ_n, which is obtained by repeated substitution in (3.2), and using the moment condition $\sup_{\gamma \in \Gamma} E_\gamma |\varepsilon_t|^4 < \infty$, which appears in the definition of Γ. Because the norming is by n^{-1}, not $n^{-1/2}$, stochastic equicontinuity also is straightforward, but tedious, to establish by applying Markov's inequality and standard manipulations (along the lines of those in (13.33) below). For brevity, the details are omitted.

Assumptions B3(ii) and B3(iii) are verified using Assumption B3* and Lemma 11.1 in Appendix A. Assumption B3*(i) holds because $Q(\theta; \gamma_0)$ is a quadratic function of β and $\{\pi^j : j \geq 1\}$ and the log function is continuous on R_+. Assumption B3*(iv) holds because $\Psi(\pi) = \{\psi = (\beta, \zeta) : \beta \in [\rho_L^* - \pi, \rho_U^* - \pi]\ \&\ \zeta \in [\zeta_L^*, \zeta_U^*]\}$ is compact $\forall \pi \in \Pi$, $\Pi = [\pi_L, \pi_U]$ is compact, and Θ is compact by its definition in (4.2). Assumption B3*(v) holds because $d_H(\Psi(\pi_1), \Psi(\pi_2)) = |\pi_1 - \pi_2|$.

Assumption B3*(ii) is verified by showing that when $\beta_0 = 0$, $E_{\gamma_0}\rho_t(\psi, \pi)$ is uniquely minimized by ψ_0 $\forall \pi \in \Pi$. This holds by the following argument. When $\beta_0 = 0$, by (3.2), we have $Y_t = \pi Y_{t-1} + \varepsilon_t - \pi\varepsilon_{t-1}$ and so $Y_t = \varepsilon_t$. Thus, when $\beta_0 = 0$, we have

$$\begin{aligned}
&2E_{\gamma_0}\rho_t(\psi, \pi) - 2E_{\gamma_0}\rho_t(\psi_0, \pi) \\
&= \log\zeta + \frac{1}{\zeta}E_{\gamma_0}\left(\varepsilon_t - \beta\sum_{j=0}^{\infty}\pi^j\varepsilon_{t-j-1}\right)^2 - \log\zeta_0 - \frac{1}{\zeta_0}E_{\gamma_0}\varepsilon_t^2 \\
&= \log\zeta + \frac{\zeta_0}{\zeta} + \frac{\beta^2\zeta_0}{\zeta(1-\pi^2)} - \log\zeta_0 - 1 \\
&\geq \log(\zeta/\zeta_0) + \frac{\zeta_0}{\zeta} - 1 + \frac{\beta^2\zeta_0}{\zeta_U} \qquad (13.21)
\end{aligned}$$

using $\zeta_0 = E_{\gamma_0}\varepsilon_t^2$ $\forall t = 0, 1, ...$ The lhs is zero for $\psi = \psi_0$. The rhs is positive for $\psi = (\beta, \zeta) \neq \psi_0 = (0, \zeta_0)$ $\forall \pi \in \Pi$. This holds by writing $\zeta/\zeta_0 = 1 + x$ and noting that the function $s(x) = \log(1+x) + 1/(1+x) - 1$ is uniquely minimized over $x \in R_+$ at $x = 0$. This property of $s(x)$ holds because its derivative, $x/(1+x)^2$, is zero for $x = 0$, is strictly negative for $x < 0$, and is strictly positive for $x > 0$. Hence, Assumption B3*(ii) holds.

Next, we establish Assumption B3*(iii), i.e., $Q(\theta; \gamma_0)$ is uniquely minimized by θ_0 $\forall \gamma_0 \in \Gamma$ with $\beta_0 \neq 0$. Using (13.4), we have

$$\begin{aligned}
&2E_{\gamma_0}\rho_t(\theta) - 2E_{\gamma_0}\rho_t(\theta_0) \\
&= \log\zeta + \frac{1}{\zeta}E_{\gamma_0}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^jY_{t-j-1}\right)^2 - \log\zeta_0 - \frac{1}{\zeta_0}E_{\gamma_0}\left(Y_t - \beta_0\sum_{j=0}^{\infty}\pi_0^jY_{t-j-1}\right)^2 \\
&= \log\zeta + \frac{1}{\zeta}E_{\gamma_0}\left(\varepsilon_t - \beta\sum_{j=0}^{\infty}\pi^jY_{t-j-1} + \beta_0\sum_{j=0}^{\infty}\pi_0^jY_{t-j-1}\right)^2 - \log\zeta_0 - \frac{1}{\zeta_0}E_{\gamma_0}\varepsilon_t^2 \\
&= \left(\log(\zeta/\zeta_0) + \frac{\zeta_0}{\zeta} - 1\right) + \frac{1}{\zeta}E_{\gamma_0}\left(\beta\sum_{j=0}^{\infty}\pi^jY_{t-j-1} - \beta_0\sum_{j=0}^{\infty}\pi_0^jY_{t-j-1}\right)^2. \qquad (13.22)
\end{aligned}$$

The first term on the rhs is uniquely minimized by $\zeta = \zeta_0$ by the argument following (13.22).

We now show that the second term on the rhs of (13.22) equals zero when $(\beta, \pi) =$

(β_0, π_0) and is positive for $(\beta, \pi) \neq (\beta_0, \pi_0)$. We have

$$
\begin{aligned}
& E_{\gamma_0}\left(\sum_{j=0}^{\infty}[\beta\pi^j - \beta_0\pi_0^j]Y_{t-j-1}\right)^2 \qquad (13.23)\\
&= E_{\gamma_0}\left((\beta-\beta_0)\varepsilon_{t-1} + (\beta-\beta_0)(\rho_0 Y_{t-2} - \pi_0\varepsilon_{t-2}) + \sum_{j=1}^{\infty}[\beta\pi^j - \beta_0\pi_0^j]Y_{t-j-1}\right)^2\\
&= (\beta-\beta_0)^2\zeta_0 + E_{\gamma_0}\left((\beta-\beta_0)(\rho_0 Y_{t-2} - \pi_0\varepsilon_{t-2}) + \sum_{j=1}^{\infty}[\beta\pi^j - \beta_0\pi_0^j]Y_{t-j-1}\right)^2,
\end{aligned}
$$

where the first equality uses (3.2) and the second equality uses the independence of ε_{t-1} and $(Y_{t-2}, \varepsilon_{t-2}, ...)$ and $E\varepsilon_{t-1} = 0$. The rhs of (13.23) is zero if $\beta = \beta_0$ and is positive if $\beta \neq \beta_0$ because $\zeta_0 > 0$.

Next, we suppose $\beta = \beta_0$ $(\neq 0)$. Then, we have

$$
\begin{aligned}
& E_{\gamma_0}\left(\sum_{j=0}^{\infty}[\beta_0\pi^j - \beta_0\pi_0^j]Y_{t-j-1}\right)^2 \qquad (13.24)\\
&= \beta_0^2 E_{\gamma_0}\left((\pi-\pi_0)\varepsilon_{t-2} + (\pi-\pi_0)(\rho_0 Y_{t-3} - \pi_0\varepsilon_{t-3}) + \sum_{j=2}^{\infty}[\pi^j - \pi_0^j]Y_{t-j-1}\right)^2\\
&= (\pi-\pi_0)^2\beta_0^2\zeta_0 + \beta_0^2 E_{\gamma_0}\left((\pi-\pi_0)(\rho_0 Y_{t-3} - \pi_0\varepsilon_{t-3}) + \sum_{j=2}^{\infty}[\pi^j - \pi_0^j]Y_{t-j-1}\right)^2.
\end{aligned}
$$

The rhs of (13.24) is zero if $\pi = \pi_0$ and is positive if $\pi \neq \pi_0$ because $\zeta_0 > 0$ and $\beta_0 \neq 0$.

We conclude that when $\beta_0 \neq 0$ the second term on the rhs of (13.22) is zero iff $(\beta, \pi) = (\beta_0, \pi_0)$. Hence, Assumption B3*(iii) holds. This completes the verification of Assumption B3*.

13.3.2. ARMA Example: Verification of Assumptions C1 and D1

We verify the quadratic expansions that appear in Assumptions C1 and D1 using Lemma 11.5, which relies on Assumption Q1. Assumption Q1(i) holds with $\rho_t(\theta)$ in place of $\rho(W_t, \theta)$. (The fact that $\rho_t(\theta)$ depends on $Y_t, Y_{t-1}, ...$, rather than just W_t, does not effect the result of Lemma 11.5.) Assumption Q1(ii) holds given the form of $\rho_t(\theta)$.

Assumption Q1(iii) holds by (i) a uniform LLN for $n^{-1}\sum_{t=1}^{n}\rho_{\psi\psi,t}(\theta) - E_{\gamma_n}\rho_{\psi\psi,t}(\theta)$ over $\theta \in \Theta$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ and (ii) the convergence $\sup_{\pi\in\Pi}\sup_{\psi\in\Psi(\pi):||\psi-\psi_{0,n}||\leq\delta_n}$

$|E_{\gamma_n}\rho_{\psi\psi,t}(\psi,\pi)-E_{\gamma_n}\rho_{\psi\psi,t}(\psi_{0,n},\pi)| \to 0$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for all constants $\delta_n \to 0$. The uniform LLN holds by the same type of argument as used to verify Assumption B3(i) using the definition of $\rho_{\psi\psi,t}(\theta)$ in (13.11)-(13.13). The convergence in (ii) holds by fairly straightforward calculations. For example, for the $(1,1)$ element of $\rho_{\psi\psi,t}(\theta)$, the difference is zero for all $n \geq 1$ and hence the limit is zero. For the $(1,2)$ element of $\rho_{\psi\psi,t}(\theta)$, we have

$$\begin{aligned}
&\sup_{\pi\in\Pi} \sup_{\psi\in\Psi(\pi):||\psi-\psi_{0,n}||\leq\delta_n} |E_{\gamma_n}\rho_{\beta\zeta,t}(\psi,\pi) - E_{\gamma_n}\rho_{\beta\zeta,t}(\psi_{0,n},\pi)| \\
&= \sup_{\pi\in\Pi} \sup_{\beta:|\beta|\leq\delta_n} \left| \zeta^{-1}\beta \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \pi^j\pi^k E_{\gamma_n} Y_{t-j-1}Y_{t-k-1} \right| \\
&\leq \zeta_L^{-1}\delta_n \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \pi_+^j\pi_+^k E_{\gamma_n} Y_t^2 \to 0, \qquad (13.25)
\end{aligned}$$

where $\pi_+ = \max\{|\pi_L|, |\pi_U|\} < 1$ and $E_{\gamma_n}Y_t^2 \to E_{\gamma_0}Y_t^2 = E_{\gamma_0}\varepsilon_t^2 = \zeta_0 < \infty$.

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

$$\begin{aligned}
&B^{-1}(\beta_n)n^{-1}\sum_{t=1}^{n}\rho_{\theta\theta,t}(\theta)B^{-1}(\beta_n) \\
&= B(\beta/\beta_n)\left(n^{-1}\sum_{t=1}^{n}\left(\rho_{\theta\theta,t}^{\dagger}(\theta) + \beta^{-1}\chi_t(\theta)\right)\right)B(\beta/\beta_n) \\
&= \left(n^{-1}\sum_{t=1}^{n}\rho_{\theta\theta,t}^{\dagger}(\theta)\right)(1+o(1)) + \left(n^{-1/2}\sum_{t=1}^{n}\left(\chi_t(\theta) - E_{\gamma_n}\chi_t(\theta)\right)\right) \times \\
&\quad (n^{1/2}\beta_n)^{-1}(1+o(1)) + \left(E_{\gamma_n}\chi_t(\theta)/\beta_n\right)(1+o(1)), \qquad (13.26)
\end{aligned}$$

where $\rho_{\theta\theta,t}^{\dagger}(\theta)$ and $\chi_t(\theta)$ are defined in (13.14). In (13.26), the second equality holds because $|\beta| \leq |\beta-\beta_n| + |\beta_n| \leq (1+\delta_n)|\beta_n|$, and $\delta_n = o(1)$. By (13.26) 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 $n^{-1}\sum_{t=1}^{n}\rho_{\theta\theta,t}^{\dagger}(\theta)$ and $n^{-1/2}\sum_{t=1}^{n}(\chi_t(\theta) - E_{\gamma_n}\chi_t(\theta))$ over $\theta \in \Theta_n(\delta_n)$ and the equicontinuity of $E_{\gamma_n}\chi_t(\theta)/|\beta_n|$ over $\theta \in \Theta_n(\delta_n)$. The stochastic equicontinuity of $n^{-1}\sum_{t=1}^{n}\rho_{\theta\theta,t}^{\dagger}(\theta)$ follows by the same argument as used above to verify Assumption B3(i) with $\rho_{\theta\theta,t}^{\dagger}(\theta)$ in place of $\rho_t(\theta)$. For brevity, details are not given.

The stochastic equicontinuity of $n^{-1/2}\sum_{t=1}^{n}(\chi_t(\theta) - E_{\gamma_n}\chi_t(\theta))$ follows from the stochastic equicontinuity of terms of the form

$$v_n^*(\pi) = n^{-1/2} \sum_{t=1}^{n} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi^j k \pi^{k-1} (Y_{t-j-1} Y_{t-k-1} - E_{\gamma_n} Y_{t-j-1} Y_{t-k-1}) \qquad (13.27)$$

over $\theta \in \Theta_n(\delta_n)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, see the definition of $\chi_t(\theta)$ in (13.16). For any $\varepsilon > 0$, we have

$$\begin{aligned}
&\varepsilon^2 P_{\gamma_n} \left(\sup_{|\pi_1 - \pi_2| < \delta} |v_n^*(\pi_1) - v_n^*(\pi_2)| > \varepsilon \right) \\
&\leq E_{\gamma_n} \sup_{|\pi_1 - \pi_2| < \delta} \left(\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} k \frac{(\pi_1^{j+k-1} - \pi_2^{j+k-1})}{a_{jk}^{1/2}} \right. \\
&\quad \left. \times\, a_{jk}^{1/2} n^{-1/2} \sum_{t=1}^{n} (Y_{t-j-1} Y_{t-k-1} - E_{\gamma_n} Y_{t-j-1} Y_{t-k-1}) \right)^2 \\
&\leq \sup_{|\pi_1 - \pi_2| < \delta} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} k^2 \frac{(\pi_1^{j+k-1} - \pi_2^{j+k-1})^2}{a_{jk}} \\
&\quad \times \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{jk} E_{\gamma_n} \left(n^{-1/2} \sum_{t=1}^{n} (Y_{t-j-1} Y_{t-k-1} - E_{\gamma_n} Y_{t-j-1} Y_{t-k-1}) \right)^2 \\
&\leq \varepsilon
\end{aligned} \qquad (13.28)$$

for $\delta > 0$ sufficiently small, where $a_{jk} = \pi_{\#}^{j+k}$, $\pi_{\#}$ is some number between $\max\{|\pi_L|, |\pi_U|\}$ and 1, the first inequality holds by Markov's inequality, the second inequality holds by the Cauchy-Schwarz inequality, and the third inequality holds because (i) $\lim_{\delta \to 0} \sup_{|\pi_1 - \pi_2| < \delta} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} k^2 ((\pi_1/\pi_{\#})^{j+k-1} - (\pi_2/\pi_{\#})^{j+k-1})^2 = 0$, which can be established using the fact that $|\pi_\ell/\pi_{\#}| < 1$ for $\ell = 1, 2$ and using mean value expansions of $(\pi_1/\pi_{\#})^{j+k-1}$ around $(\pi_2/\pi_{\#})^{j+k-1}$ $\forall j, k \geq 0$, (ii) $Var_{\gamma_n}(n^{-1/2} \sum_{t=1}^{n} Y_{t-j-1} Y_{t-k-1}) \leq C$ $\forall n \geq 1$ for some $C < \infty$ by standard calculations, and (iii) $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{jk} < \infty$.

It remains to show that $\sup_{\theta_1, \theta_2 \in \Theta_n(\delta_n)} |\beta_n|^{-1} E_{\gamma_n} (\chi_t(\theta_1) - \chi_t(\theta_2)) = o(1)$. It suffices to show that $\sup_{\theta \in \Theta_n(\delta_n)} |\beta_n|^{-1} E_{\gamma_n} \chi_t(\theta) = o(1)$. For any $\theta \in \Theta_n(\delta_n)$, we have

$$\begin{aligned}
&|\beta_n|^{-1} E_{\gamma_n} \chi_t(\theta) \\
&= |\beta_n|^{-1} (E_{\gamma_n} \chi_t(\theta) - E_{\gamma_n} \chi_t(\psi_n, \pi)) + |\beta_n|^{-1} E_{\gamma_n} \chi_t(\psi_n, \pi).
\end{aligned} \qquad (13.29)$$

To show that the first term on the rhs of (13.29) is $o(1)$, we write

$$E_{\gamma_n}\chi_{\beta\pi,t}(\theta) = -\zeta^{-1}E_{\gamma_n}\left(\beta_n\sum_{j=0}^{\infty}\pi_n^j Y_{t-j-1} - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1} \text{ and}$$

$$E_{\gamma_n}\chi_{\beta\pi,t}(\psi_n,\pi) = -\zeta_n^{-1}E_{\gamma_n}\left(\beta_n\sum_{j=0}^{\infty}(\pi_n^j - \pi^j)Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1} \qquad (13.30)$$

using the definition of $\chi_{\beta\pi,t}(\theta)$ in (13.16).

For $\theta \in \Theta_n(\delta_n)$,

$$\begin{aligned}&|\zeta E_{\gamma_n}\chi_{\beta\pi,t}(\theta) - \zeta_n E_{\gamma_n}\chi_{\beta\pi,t}(\psi_n,\pi)| \\ &= \left|(\beta - \beta_n)\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^j k\pi^{k-1}E_{\gamma_n}Y_{t-j-1}Y_{t-k-1}\right| \leq \delta_n|\beta_n|C \end{aligned}\qquad (13.31)$$

for some constant $C < \infty$, where the inequality uses the definition of $\Theta_n(\delta_n)$ and $|E_{\gamma_n}Y_{t-j-1}Y_{t-k-1}| \leq E_{\gamma_n}Y_t^2 \leq C_1\ \forall n \geq 1$ for some constant $C_1 < \infty$. Combining (13.30), (13.31), and $\sup_{n\geq 1}|\zeta_n E_{\gamma_n}\chi_{\beta\pi,t}(\theta_n)| < \infty$ (which holds by standard calculations) establishes that the (3, 1) element (i.e., the $\beta\pi$ element) of the first term on the rhs of (13.29) is $o(1)$:

$$\begin{aligned}&\sup_{\theta\in\Theta_n(\delta_n)}|E_{\gamma_n}\chi_{\beta\pi,t}(\theta) - E_{\gamma_n}\chi_{\beta\pi,t}(\psi_n,\pi)| \\ &\leq \sup_{\theta\in\Theta_n(\delta_n)}\zeta^{-1}|\zeta E_{\gamma_n}\chi_{\beta\pi,t}(\theta) - \zeta_n E_{\gamma_n}\chi_{\beta\pi,t}(\psi_n,\pi)| \\ &\quad + \sup_{\theta\in\Theta_n(\delta_n)}|\zeta^{-1}(\zeta_n - \zeta)E_{\gamma_n}\chi_{\beta\pi,t}(\psi_n,\pi)| \\ &= o(|\beta_n|),\end{aligned}\qquad (13.32)$$

using $\zeta_n - \zeta = O(\delta_n|\beta_n|)$ by the definition of $\Theta_n(\delta_n)$ and $\zeta \geq \zeta_L > 0$.

The proof for the (3, 3) element (i.e., the $\pi\pi$ element) of the first term on the rhs of (13.29), which is the only other non-zero element of $\chi_t(\theta)$, is the same with $k(k-1)\pi^{k-2}$ in place of $k\pi^{k-1}$. This completes the proof that the first summand on the rhs of (13.29) is $o(1)$.

Let $c_j = |E_{\gamma_0}Y_1Y_{1+j}|$. The second summand on the rhs of (13.29) is $O(\delta_n) = o(1)$ by the following calculations: for $\theta \in \Theta_n(\delta_n)$,

$$\begin{aligned}
|\beta_n^{-1} E_{\gamma_n} \chi_{\beta\pi,t}(\psi_n,\pi)| &= \left| \beta_n^{-1}\zeta_n^{-1} E_{\gamma_n} \left(\beta_n \sum_{j=0}^{\infty} (\pi_n^j - \pi^j) Y_{t-j-1} \right) \sum_{k=1}^{\infty} k\pi^{k-1} Y_{t-k-1} \right| \\
&\leq \zeta_L^{-1} \sum_{j=1}^{\infty} |\pi^j - \pi_n^j| \sum_{k=1}^{\infty} k\pi_U^{k-1} c_{j-k} \\
&\leq C\zeta_L^{-1} \sum_{j=1}^{\infty} j\pi_U^{j-1} |\pi - \pi_n| \sum_{k=1}^{\infty} k\pi_U^{k-1} \\
&\leq \delta_n C \zeta_L^{-1} \left(\sum_{j=1}^{\infty} j\pi_U^{j-1} \right)^2 = o(1), \qquad (13.33)
\end{aligned}$$

where the equality holds by (13.30), the second inequality holds because $|\pi^j - \pi_n^j| \leq |j\pi_{n*}^{j-1}(\pi - \pi_n)| \leq j\pi_U^{j-1}|\pi - \pi_n|$ for some π_{n*} between π and π_n by a mean-value expansion and $\sup_{j\geq 1} c_j < \infty$, and the last equality holds because $\sum_{j=1}^{\infty} j\pi_U^{j-1} < \infty$ and $\delta_n = o(1)$.

For the (3, 3) element of $\chi_t(\psi_n, \pi)$, we obtain $|\beta_n^{-1} E_{\gamma_n} \chi_{\pi\pi,t}(\psi_n,\pi)| \leq |\pi - \pi_n| C^* = O(\delta_n) = o(1)$ for a constant $C^* < \infty$ by the same argument as in (13.33) with $k(k-1)\pi^{k-2}$ in place of $k\pi^{k-1}$. This concludes the proof that the second summand on the rhs of (13.29) is $o(1)$, which completes the verification of Assumption Q1(iv). In turn, this completes the verification of Assumptions C1 and D1.

13.3.3. ARMA Example: Verification of Assumptions C2-C4

Assumption C2 is verified in AC1.

The empirical process $\{G_n(\pi) : \pi \in \Pi\}$ that appears in Assumption C3 is defined in (4.12). The covariance matrix of the stochastic process $\{G(\pi;\gamma_0) : \pi \in \Pi\}$ that appears in Assumption C3 is defined in (4.14) and is derived in (13.5). The weak convergence $G_n(\cdot) \Rightarrow G(\cdot;\gamma_0)$ holds by the proof of Theorem 1(a) of Andrews and Ploberger (1996, pp. 1339-1340).

Assumption C4(i) holds by a uniform LLN for $n^{-1}\sum_{t=1}^{n}(\rho_{\psi\psi,t}(\psi_{0,n},\pi) - E_{\gamma_n}\rho_{\psi\psi,t}(\psi_{0,n},\pi))$ over $\pi \in \Pi$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ and the convergence result $\sup_{\pi\in\Pi} |E_{\gamma_n}\rho_{\psi\psi,t}(\psi_{0,n},\pi) - E_{\gamma_0}\rho_{\psi\psi,t}(\psi_0,\pi)| \to 0$. Using the definition of $\rho_{\psi\psi,t}(\psi_{0,n},\pi)$ in (4.15), the uniform LLN holds by the same sort of argument as used to prove Assumption B3(i). For brevity, the details are not given. The convergence result holds by the same calculations as in the verification of Assumption Q1(iii), see (13.25). The simplified expression for

$H(\pi;\gamma_0) = E_{\gamma_0}\rho_{\psi\psi,t}(\psi_0,\pi)$ is derived in (13.8).

Assumption C4(ii) holds because $H(\pi;\gamma_0) = Diag\{(1-\pi^2)^{-1},(2\zeta_0^2)^{-1}\}$ by (13.8), $\inf_{\pi\in\Pi}(1-\pi^2)^{-1} \geq 1$, and $\zeta^* \geq \zeta_L^* > 0$ by the definition of Θ^*.

13.3.4. ARMA Example: Verification of Assumption C5

The quantity $K_n(\theta;\gamma^*)$ that appears in Assumption C5 is

$$K_n(\theta;\gamma^*) = n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\beta^*}E_{\gamma^*}\rho_{\psi,t}(\theta) = \begin{pmatrix} \frac{\partial}{\partial\beta^*}E_{\gamma^*}\rho_{\beta,t}(\theta) \\ \frac{\partial}{\partial\beta^*}E_{\gamma^*}\rho_{\zeta,t}(\theta) \end{pmatrix}. \tag{13.34}$$

The terms on the rhs of (13.34) are calculated as follows:

$$\begin{aligned} E_{\gamma^*}\rho_{\beta,t}(\theta) &= -\zeta^{-1}E_{\gamma^*}\left(\varepsilon_t + \beta^*\sum_{j=0}^{\infty}\pi^{*j}Y_{t-j-1} - \beta\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)\sum_{k=0}^{\infty}\pi^{k}Y_{t-k-1} \\ &= -\zeta^{-1}\beta^*\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{*j}\pi^{k}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1} + \zeta^{-1}\beta\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{j}\pi^{k}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1} \end{aligned} \tag{13.35}$$

and

$$\begin{aligned} &\frac{\partial}{\partial\beta^*}E_{\gamma^*}\rho_{\beta,t}(\theta) \\ &= -\zeta^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{*j}\pi^{k}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1} - \zeta^{-1}\beta^*\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{*j}\pi^{k}\frac{\partial}{\partial\beta^*}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1} \\ &\quad +\zeta^{-1}\beta\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{j}\pi^{k}\frac{\partial}{\partial\beta^*}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1}. \end{aligned} \tag{13.36}$$

In addition, we have

$$E_{\gamma^*}\rho_{\zeta,t}(\theta) = -(1/2)\zeta^{-2}\left(E_{\gamma^*}\left(\varepsilon_t + \beta^*\sum_{j=0}^{\infty}\pi^{*j}Y_{t-j-1} - \beta\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)^2 - \zeta\right)$$

$$= -(1/2)\zeta^{-2}\left(\zeta^* - \zeta + E_{\gamma^*}\left(\beta^*\sum_{j=0}^{\infty}\pi^{*j}Y_{t-j-1} - \beta\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)^2\right) \qquad (13.37)$$

$$= -(1/2)\zeta^{-2}\left(\zeta^* - \zeta + \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(\beta^{*2}\pi^{*(j+k)} - 2\beta^*\beta\pi^{*j}\pi^k + \beta^2\pi^{j+k}\right)E_{\gamma^*}Y_{t-j-1}Y_{t-k-1}\right).$$

This gives

$$\frac{\partial}{\partial\beta^*}E_{\gamma^*}\rho_{\zeta,t}(\theta)$$

$$= -(1/2)\zeta^{-2}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(2\beta^*\pi^{*(j+k)} - 2\beta\pi^{*j}\pi^k\right)E_{\gamma^*}Y_{t-j-1}Y_{t-k-1}\right) \qquad (13.38)$$

$$-(1/2)\zeta^{-2}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(\beta^{*2}\pi^{*(j+k)} - 2\beta^*\beta\pi^{*j}\pi^k + \beta^2\pi^{j+k}\right)\frac{\partial}{\partial\beta^*}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1}.$$

>From (13.36), if $\widetilde{\gamma}_n \to \gamma_0$ with $\beta_0 = 0$ (for non-stochastic $\widetilde{\gamma}_n$) and $\psi_n \to \psi_0 = (0, \zeta_0)$, as in Assumption C5, then

$$\frac{\partial}{\partial\widetilde{\beta}_n}E_{\widetilde{\gamma}_n}\rho_{\beta,t}(\psi_0,\pi) \to -\zeta_0^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi_0^j\pi^k E_{\gamma_0}Y_{t-j-1}Y_{t-k-1}$$

$$= -\zeta_0^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi_0^j\pi^k E_{\gamma_0}\varepsilon_{t-j-1}\varepsilon_{t-k-1} = -\sum_{j=0}^{\infty}\pi_0^j\pi^j = -\frac{1}{1-\pi_0\pi}. \qquad (13.39)$$

The convergence is uniform in $\pi \in \Pi$ because (i) $|\pi| \leq \max\{|\pi_L|, |\pi_U|\} < 1\ \forall\pi \in \Pi$ and (ii) the term $(\partial/\partial\widetilde{\beta}_n)E_{\widetilde{\gamma}_n}Y_{t-j-1}Y_{t-k-1}$ is well-defined and is bounded in absolute value uniformly over $n \geq 1$. This holds because when the true parameter is $\widetilde{\gamma}_n$, we can write

$$Y_t = (\widetilde{\pi}_n + \widetilde{\beta}_n)Y_{t-1} + u_t = \sum_{j=0}^{\infty}(\widetilde{\pi}_n + \widetilde{\beta}_n)^j u_{t-j-1}, \text{ where } u_t = \varepsilon_t - \widetilde{\pi}_n\varepsilon_{t-1}, \text{ and}$$

$$\frac{\partial}{\partial\widetilde{\beta}_n}E_{\widetilde{\gamma}_n}Y_sY_t = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{\partial}{\partial\widetilde{\beta}_n}[(\widetilde{\pi}_n + \widetilde{\beta}_n)^j(\widetilde{\pi}_n + \widetilde{\beta}_n)^k]E_{\widetilde{\gamma}_n}u_{s-j-1}u_{t-k-1}. \qquad (13.40)$$

>From (13.38), if $\widetilde{\gamma}_n \to \gamma_0$ with $\beta_0 = 0$ and $\psi_n \to \psi_0 = (0, \zeta_0)$, as in Assumption C5, then

$$\frac{\partial}{\partial \widetilde{\beta}_n} E_{\widetilde{\gamma}_n} \rho_{\zeta,t}(\psi_n, \pi) \to 0 \tag{13.41}$$

due to the multiplicative terms β^*, β, β^{*2}, $\beta^* \beta$, and β^2 that appear in (13.38) and that converge to 0 when $\beta^* = \widetilde{\beta}_n \to 0$ and $\beta = \beta_n \to 0$.

Combining (13.34), (13.39), and (13.41) verifies Assumption C5(i) and C5(ii) with $K(\pi; \gamma_0) = (-(1 - \pi_0 \pi)^{-1}, 0)$. Assumption C5(iii) holds because $1 - \pi_0 \pi \neq 0 \ \forall \pi \in \Pi$.

13.3.5. ARMA Example: Verification of Assumption C6

Now, we verify Assumption C6 using Assumption C6**, which is shown in Lemma 4.1 to be sufficient for Assumption C6. Assumption C6**(i) holds because β is a scalar. Assumption C6**(ii) requires $\Omega_G(\pi_1, \pi_2; \gamma_0)$ to be positive definite $\forall \pi_1, \pi_2 \in \Pi$ with $\pi_1 \neq \pi_2$, $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$. The expression for $\Omega_G(\pi_1, \pi_2; \gamma_0)$ given in the rhs matrix in (13.10) is positive definite because the determinant of the upper left 2×2 matrix is zero iff $\pi_1 = \pi_2$ by straightforward calculations and $\zeta_0^{-4} E_{\gamma_0} (\varepsilon_t^2 - \zeta_0)^2 > 0$ by the definitions of Θ^* and Φ^* in (4.3) and (4.4). This completes the verification of Assumption C6**. Hence, Assumption C6 holds.

13.3.6. ARMA Example: Verification of Assumption C8

Here we verify Assumption C8. Suppose $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, which implies that $\beta_0 = 0$. From (13.35), we have

$$\frac{\partial}{\partial \beta} E_{\gamma^*} \rho_{\beta,t}(\theta) = \zeta^{-1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi^j \pi^k E_{\gamma^*} Y_{t-j-1} Y_{t-k-1}, \tag{13.42}$$

which leads to

$$\begin{aligned} \frac{\partial}{\partial \beta} E_{\gamma_n} \rho_{\beta,t}(\psi, \pi_n) \bigg|_{\psi = \psi_n} &= \zeta_n^{-1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_n^j \pi_n^k E_{\gamma_n} Y_{t-j-1} Y_{t-k-1} \\ \to \zeta_0^{-1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_0^j \pi_0^k E_{\gamma_0} Y_{t-j-1} Y_{t-k-1} &= \zeta_0^{-1} \sum_{j=0}^{\infty} \pi_0^{2j} E_{\gamma_0} \varepsilon_{t-j-1}^2 = \frac{1}{1 - \pi_0^2}, \end{aligned} \tag{13.43}$$

where the second to last equality uses $E_{\gamma_0} Y_{t-j-1} Y_{t-k-1} = E_{\gamma_0} \varepsilon_{t-j-1} \varepsilon_{t-k-1}$ because $\beta_0 = 0$ and $E_{\gamma_0} \varepsilon_{t-j-1} \varepsilon_{t-k-1} = 0$ for $j \neq k$ because $\{\varepsilon_t : t \leq n\}$ are mean zero and independent.

>From (13.35), we also have

$$\frac{\partial}{\partial \zeta} E_{\gamma^*} \rho_{\beta,t}(\theta) \tag{13.44}$$
$$= \zeta^{-2}\beta^* \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \pi^{*j}\pi^k E_{\gamma^*} Y_{t-j-1}Y_{t-k-1} - \zeta^{-2}\beta \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \pi^{j}\pi^k E_{\gamma^*} Y_{t-j-1}Y_{t-k-1},$$

which yields

$$\frac{\partial}{\partial \zeta} E_{\gamma_n} \rho_{\beta,t}(\psi, \pi_n)|_{\psi=\psi_n} = 0 \ \forall n \geq 1. \tag{13.45}$$

>From (13.37), we have

$$\frac{\partial}{\partial \beta} E_{\gamma^*} \rho_{\zeta,t}(\theta) = \zeta^{-2} \left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \left(\beta^* \pi^{*j}\pi^k - \beta\pi^{j+k}\right) E_{\gamma^*} Y_{t-j-1}Y_{t-k-1} \right), \tag{13.46}$$

which yields

$$\frac{\partial}{\partial \beta} E_{\gamma_n} \rho_{\zeta,t}(\psi, \pi_n)|_{\psi=\psi_n} = 0 \ \forall n \geq 1. \tag{13.47}$$

>From (13.37), we also have

$$\frac{\partial}{\partial \zeta} E_{\gamma^*} \rho_{\zeta,t}(\theta)$$
$$= \zeta^{-3} \left(\zeta^* - \zeta + \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \left(\beta^{*2}\pi^{*(j+k)} - 2\beta^*\beta\pi^{*j}\pi^k + \beta^2\pi^{j+k}\right) E_{\gamma^*} Y_{t-j-1}Y_{t-k-1} \right)$$
$$+(1/2)\zeta^{-2}, \tag{13.48}$$

which yields

$$\frac{\partial}{\partial \zeta} E_{\gamma_n} \rho_{\zeta,t}(\psi, \pi_n)|_{\psi=\psi_n} = (1/2)\zeta_n^{-2} \to (1/2)\zeta_0^{-2}. \tag{13.49}$$

Combining (13.43), (13.45), (13.47), and (13.49) gives

$$\frac{\partial}{\partial \psi'} E_{\gamma_n} D_\psi Q_n(\psi, \pi_n)|_{\psi=\psi_n} = \frac{\partial}{\partial \psi'} E_{\gamma_n} \rho_{\psi,t}(\psi, \pi_n)|_{\psi=\psi_n}$$
$$\to \begin{bmatrix} (1-\pi_0^2)^{-1} & 0 \\ 0 & (1/2)\zeta_0^{-2} \end{bmatrix} = H\left(\pi_0; \gamma_0\right), \tag{13.50}$$

where the first equality holds by (4.8). This completes the verification of Assumption C8.

13.3.7. ARMA Example: Verification of Assumption D2

Next, we verify Assumption D2. By (13.26), we have

$$
\begin{aligned}
J_n &= B^{-1}(\beta_n) n^{-1} \sum_{t=1}^{n} \rho_{\theta\theta,t}(\theta_n) B^{-1}(\beta_n) \\
&= \left(n^{-1} \sum_{t=1}^{n} \rho^{\dagger}_{\theta\theta,t}(\theta_n) \right) (1+o(1)) + \left(n^{-1/2} \sum_{t=1}^{n} (\chi_t(\theta_n) - E_{\gamma_n}\chi_t(\theta_n)) \right) \times \\
&\quad (n^{1/2}\beta_n)^{-1}(1+o(1)) + \left(E_{\gamma_n}\chi_t(\theta_n)/\beta_n \right)(1+o(1)) \\
&= \left(n^{-1} \sum_{t=1}^{n} \rho^{\dagger}_{\theta\theta,t}(\theta_n) \right) (1+o(1)) + o(1) \\
&= E_{\gamma_n}\rho^{\dagger}_{\theta\theta,t}(\theta_n) + o_p(1) \\
&\to_p E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0) = J(\gamma_0),
\end{aligned} \tag{13.51}
$$

where the third equality holds because $n^{1/2}|\beta_n| \to \infty$ for $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $E_{\gamma_n}\chi_t(\theta_n) = 0$ by the equation for $E_{\gamma_n}\chi_{\beta\pi,t}(\psi_n, \pi)$ in (13.30) evaluated at $\pi = \pi_n$ and an analogous equation for $E_{\gamma_n}\chi_{\pi\pi,t}(\psi_n, \pi)$, and $n^{-1/2}\sum_{t=1}^{n}(\chi_t(\theta_n) - E_{\gamma_n}\chi_t(\theta_n)) = O_p(1)$ because $Var_{\gamma_n}(n^{-1/2}\sum_{t=1}^{n}\chi_{\beta\pi,t}(\theta_n))^2 = O(1)$ by straightforward calculations using the fact that $\chi_{\beta\pi,t}(\theta_n) = -\zeta^{-1}\varepsilon_t \sum_{k=0}^{\infty} k\pi^{k-1}Y_{t-k-1}$ is a martingale difference sequence for $t = 1, ..., n$ and likewise for $n^{-1/2}\sum_{t=1}^{n}\chi_{\pi\pi,t}(\theta_n)$, the fourth equality holds by the mean square convergence of $n^{-1}\sum_{t=1}^{n}\rho^{\dagger}_{\theta\theta,t}(\theta_n) - E_{\gamma_n}\rho^{\dagger}_{\theta\theta,t}(\theta_n)$ to zero which holds by straightforward, but tedious, calculations that are not given here for brevity, and the convergence in the last line holds straightforwardly by the form of $\rho^{\dagger}_{\theta\theta,t}(\theta_n)$ given in (13.12)-(13.15) and $\gamma_n \to \gamma_0$.

The form of the matrix $J(\gamma_0)$ given in (4.30) is derived in (13.11)-(13.17) above.

Assumption D2 requires that $J(\gamma_0)$ is nonsingular. To show this, note that $J(\gamma_0) = E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0)$, as specified in (13.17), is block diagonal between its (β, π) and ζ elements. Since $(2\zeta_0^2)^{-1} > 0$ by the definition of Θ^*, it suffices to show that the 2×2 sub-matrix of $E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0)$ that corresponds to (β, π) is positive definite. The latter multiplied by ζ_0 equals

$$
E_{\gamma_0} A_t A_t', \text{ where } A_t = \begin{pmatrix} A_{1t} \\ A_{2t} \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{\infty} \pi_0^j Y_{t-j-1} \\ \sum_{j=1}^{\infty} j\pi_0^{j-1} Y_{t-j-1} \end{pmatrix}. \tag{13.52}
$$

Now, by (3.2), $Y_t = \varepsilon_t + (\pi_0 + \beta_0)Y_{t-1} - \pi_0\varepsilon_{t-1}$. Hence,

$$\begin{aligned} A_{1t} &= Y_{t-1} + \sum_{j=1}^{\infty} \pi_0^j Y_{t-j-1} = \varepsilon_{t-1} + \xi_{t-2}, \text{ where} \\ \xi_{t-2} &= (\pi_0 + \beta_0)Y_{t-2} - \pi_0\varepsilon_{t-2} + \sum_{j=1}^{\infty} \pi_0^j Y_{t-j-1} \end{aligned} \tag{13.53}$$

and ξ_{t-2} is independent of ε_{t-1}. For $\lambda = (\lambda_1, \lambda_2)' \in R^2$ with $\lambda \neq 0$, we have

$$\begin{aligned} \lambda' E_{\gamma_0} A_t A_t' \lambda &= E_{\gamma_0} \left(\lambda_1 \varepsilon_{t-1} + \lambda_1 \xi_{t-2} + \lambda_2 \sum_{j=1}^{\infty} j \pi_0^{j-1} Y_{t-j-1} \right)^2 \\ &= \lambda_1^2 E_{\gamma_0} \varepsilon_{t-1}^2 + E_{\gamma_0} \left(\lambda_1 \xi_{t-2} + \lambda_2 \sum_{j=1}^{\infty} j \pi_0^{j-1} Y_{t-j-1} \right)^2 . \end{aligned} \tag{13.54}$$

The rhs is positive if $\lambda_1 \neq 0$. Alternatively, suppose $\lambda_1 = 0$, then $\lambda_2^2 > 0$ and the rhs divided by λ_2^2 equals

$$\begin{aligned} & E_{\gamma_0} \left(\sum_{j=1}^{\infty} j \pi_0^{j-1} Y_{t-j-1} \right)^2 \\ &= E_{\gamma_0} \left(Y_{t-2} + \sum_{j=2}^{\infty} j \pi_0^{j-1} Y_{t-j-1} \right)^2 \\ &= E_{\gamma_0} \left(\varepsilon_{t-2} + (\pi_0 + \beta_0) Y_{t-3} - \pi_0 \varepsilon_{t-3} + \sum_{j=2}^{\infty} j \pi_0^{j-1} Y_{t-j-1} \right)^2 \\ &= E_{\gamma_0} \varepsilon_{t-2}^2 + E_{\gamma_0} \left((\pi_0 + \beta_0) Y_{t-3} - \pi_0 \varepsilon_{t-3} + \sum_{j=2}^{\infty} j \pi_0^{j-1} Y_{t-j-1} \right)^2 \\ &\geq \zeta_0 > 0. \end{aligned} \tag{13.55}$$

We conclude that $\lambda' E_{\gamma_0} A_t A_t' \lambda > 0 \ \forall \lambda = (\lambda_1, \lambda_2)' \in R^2$ with $\lambda \neq 0$ and, hence, $E_{\gamma_0} A_t A_t'$ is positive definite. This completes the verification that $J(\gamma_0)$ is positive definite.

13.3.8. ARMA Example: Verification of Assumption D3

Assumption D3(i) is verified as follows. By the definitions in (4.8), (4.28), and (4.29), we have

$$n^{1/2}B^{-1}(\beta_n)DQ_n(\theta_n) = n^{-1/2}\sum_{t=1}^{n} B^{-1}(\beta_n)\rho_{\theta,t}(\theta_n)$$

$$= -n^{-1/2}\sum_{t=1}^{n}\begin{pmatrix} \zeta_n^{-1}\varepsilon_t \sum_{k=0}^{\infty}\pi_n^k Y_{t-k-1} \\ (1/2)\zeta_n^{-2}\left(\varepsilon_t^2-\zeta_n\right) \\ \zeta_n^{-1}\varepsilon_t\sum_{k=0}^{\infty}k\pi_n^{k-1}Y_{t-k-1}\end{pmatrix} \rightarrow_d N(0,V(\gamma_0)), \quad (13.56)$$

where the convergence in distribution holds by a triangular array martingale difference CLT for row-wise stationary random variables, e.g., see Hall and Hyde (1980, Thm. 3.1), and $V(\gamma_0) = \lim_{n\to\infty} Var_{\gamma_n}(n^{-1/2}\sum_{t=1}^{n} B^{-1}(\beta_n)\rho_{\theta,t}(\theta_n))$. The verification of the conditions of Hall and Hyde's martingale difference CLT is essentially the same as given in the proof of Thm. 1(b) of Andrews and Ploberger (1996, p. 1339) and uses the condition $E_{\phi_n}|\zeta_n^{-1/2}\varepsilon_t|^{4+\delta} \le K < \infty$, which appears in the definition of Φ in (4.4), to verify a Lyapounov-type condition. The formula for $V(\gamma_0)$ given in (4.32) is derived in (13.18)-(13.20).

To verify Assumption D3(ii), note that the matrix $V(\gamma_0) = V^{\dagger}(\theta_0,\theta_0;\gamma_0)$ is the same as $J(\gamma_0) = E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0)$ but with $(1/4)\zeta_0^{-4}E_{\gamma_0}\left(\varepsilon_t^2-\zeta_0\right)^2$ in place of $(2\zeta_0^2)^{-1}$, see (13.17) and (13.19). Because $(1/4)\zeta_0^{-4}E_{\gamma_0}\left(\varepsilon_t^2-\zeta_0\right)^2 > 0$ by the definition of the parameter spaces Θ^* and Φ^*, the same argument as used above to show that $J(\gamma_0)$ is pd also shows that $V(\gamma_0)$ is pd. Hence, Assumption D3(ii) holds.

13.3.9. ARMA Example: Verification of Assumptions V1 and V2

Assumption V1(i) (for scalar β) holds with

$$J(\theta;\gamma_0) = Diag\left\{\zeta^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)^2,\ (2\zeta^2)^{-1},\ \zeta^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}j\pi^{j-1}Y_{t-j-1}\right)^2\right\}$$

$$+\left(\zeta^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)\sum_{k=0}^{\infty}k\pi^{k-1}Y_{t-k-1}\right)\times\begin{bmatrix}0&0&1\\0&0&0\\1&0&0\end{bmatrix} \quad (13.57)$$

by the same type of argument as used to verify Assumption B3(i). Assumption V1(i) (for scalar β) holds with $V(\theta;\gamma_0)$ defined just as $J(\theta;\gamma_0)$ is defined, but with

$$(4\zeta^2)^{-1}E_{\gamma_0}\left(\left(Y_t-\beta\sum_{j=0}^{\infty}\pi^jY_{t-j-1}\right)^2-\zeta\right)^2 \tag{13.58}$$

in place of $(2\zeta^2)^{-1}$, by the same type of argument as used to verify Assumption B3(i). This argument requires the additional condition $E_\phi|\xi_t|^{8+\delta_2}\leq K$ in the definition of Φ in (4.4).

Assumption V1(ii) holds by the functional forms of $J(\theta;\gamma_0)$ and $V(\theta;\gamma_0)$.

Next, we verify Assumption V1(iii). By definition, $\Sigma(\pi;\gamma_0)=J^{-1}(\psi_0,\pi;\gamma_0)V(\psi_0,\pi;\gamma_0)$ $J^{-1}(\psi_0,\pi;\gamma_0)$. Because the matrices $J(\theta;\gamma_0)$ and $V(\theta;\gamma_0)$ are block diagonal between the parameters (β,π) and ζ and these matrices are equal when their second rows and columns are deleted, it suffices to show that (i) Assumption V1(iii) holds for $\Sigma(\pi;\gamma_0)$ replaced by $J^{-1}(\psi_0,\pi;\gamma_0)$ with its second row and column deleted, which we call $A^{-1}(\pi)$, and (ii) the (2, 2) element of $\Sigma(\pi;\gamma_0)$, call it $\Sigma_{22}(\pi;\gamma_0)$, is in $(0,\infty)$ for all $\pi\in\Pi$. When $\beta_0=0$, we have

$$\begin{aligned}A(\pi)&=\zeta_0^{-1}E_{\gamma_0}\begin{pmatrix}\sum_{j=0}^{\infty}\pi^jY_{t-j-1}\\ \sum_{j=0}^{\infty}j\pi^{j-1}Y_{t-j-1}\end{pmatrix}\begin{pmatrix}\sum_{j=0}^{\infty}\pi^jY_{t-j-1}\\ \sum_{j=0}^{\infty}j\pi^{j-1}Y_{t-j-1}\end{pmatrix}'\\&=\begin{pmatrix}\sum_{j=0}^{\infty}\pi^{2j} & \sum_{j=0}^{\infty}j\pi^{2j-1}\\ \sum_{j=0}^{\infty}j\pi^{2j-1} & \sum_{j=0}^{\infty}j^2\pi^{2(j-1)}\end{pmatrix}\end{aligned} \tag{13.59}$$

where the first equality holds by (13.57) and the second equality holds because $Y_t=\varepsilon_t$ under γ_0 when $\beta_0=0$, which is the case in Assumption V1(iii). We have: $||A(\pi)||<\infty$ because $|\pi|<1$ $\forall\pi\in\Pi$. In addition, $\det(A(\pi))>0$ because

$$\left(\sum_{j=0}^{\infty}j\pi^{2j-1}\right)^2<\left(\sum_{j=0}^{\infty}\pi^{2j}\right)\left(\sum_{j=0}^{\infty}j^2\pi^{2(j-1)}\right)\ \forall\pi\in\Pi \tag{13.60}$$

by the Cauchy-Schwarz inequality. This implies $\lambda_{\min}(A^{-1}(\pi))>0$ and $\lambda_{\max}(A^{-1}(\pi))<\infty$ $\forall\pi\in\Pi$. Next, when $\beta_0=0$, using (13.57) and (13.58), we have $\Sigma_{22}(\pi;\gamma_0)=(2\zeta_0^2)(4\zeta_0^2)^{-1}E_{\gamma_0}(Y_t^2-\zeta_0)^2(2\zeta_0^2)=\zeta_0^2E_{\gamma_0}(\varepsilon_t^2-\zeta_0)^2$, which lies in $(0,\infty)$ because $\zeta_0=Var(\varepsilon_t)>0$ and $E_{\gamma_0}\varepsilon_t^4<\infty$. 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

$\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$. 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, imply that Assumption V2 holds.

13.4. Proof of the ARMA Initial Conditions Lemma

Proof of Lemma 13.1. To prove part (a), we write

$$\begin{aligned} 2\zeta_L Q_n^{IC}(\theta) &= 2\zeta_L |Q_n^\infty(\theta) - Q_n(\theta)| \\ &\leq \left| n^{-1} \sum_{t=1}^n [(A_t - B_t)^2 - A_t^2] \right| = \left| n^{-1} \sum_{t=1}^n [-2A_t B_t + B_t^2] \right| \\ &\leq 2 \left(n^{-1} \sum_{t=1}^n A_t^2 \right)^{1/2} \left(n^{-1} \sum_{t=1}^n B_t^2 \right)^{1/2} + n^{-1} \sum_{t=1}^n B_t^2, \end{aligned} \tag{13.61}$$

where

$$A_t = A_t(\theta) = Y_t - \beta \sum_{j=0}^{t-1} \pi^j Y_{t-j-1} \text{ and } B_t = B_t(\theta) = \beta \sum_{j=t}^{\infty} \pi^j Y_{t-j-1}. \tag{13.62}$$

Hence, to show part (a), it suffices to show that under $\{\gamma_n\} \in \Gamma(\gamma_0)$ $\forall \gamma_0 \in \Gamma$,

$$\sup_{\theta \in \Theta} n^{-1} \sum_{t=1}^n A_t^2(\theta) = O_p(1) \text{ and } \sup_{\theta \in \Theta} n^{-1} \sum_{t=1}^n B_t^2(\theta) = o_p(1). \tag{13.63}$$

To show (13.63), we have

$$\begin{aligned} n^{-1} \sum_{t=1}^n B_t^2(\theta) &= \beta^2 n^{-1} \sum_{t=1}^n \left(\sum_{j=t}^{\infty} \pi^j Y_{t-j-1} \right)^2 = \beta^2 n^{-1} \sum_{t=1}^n \left(\sum_{k=0}^{\infty} \pi^{t+k} Y_{-k-1} \right)^2 \\ &\leq n^{-1} \beta_U^2 \sum_{t=1}^{\infty} \pi_+^{2t} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_+^{j+k} |Y_{-j-1} Y_{-k-1}|, \end{aligned} \tag{13.64}$$

where the second equality holds by change of variables with $k = j - t$, $\beta_U = \max\{\rho_U - \pi_L, \pi_U - \rho_L\}$, and $\pi_+ = \max\{|\pi_L|, |\pi_U|\}$. Using (13.64), we obtain

$$E_{\gamma_n} \sup_{\theta \in \Theta} n^{-1} \sum_{t=1}^n B_t^2(\theta) \leq n^{-1} \beta_U^2 \sum_{t=1}^{\infty} \pi_+^{2t} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_+^{j+k} E_{\gamma_n} Y_1^2 \to 0, \tag{13.65}$$

where the inequality uses $E_{\gamma_n}|Y_{-j-1}Y_{-k-1}| \leq \sup_{n\geq 1} E_{\gamma_n} Y_1^2 \leq C < \infty$ by the Cauchy-Schwarz inequality and stationarity.

Next, we have

$$
\begin{aligned}
& E_{\gamma_n} \sup_{\theta \in \Theta} n^{-1} \sum_{t=1}^{n} A_t^2(\theta) \leq \sup_{t \geq 1} E_{\gamma_n} \sup_{\theta \in \Theta} A_t^2(\theta) \\
\leq & \; 2 \sup_{t \geq 1} E_{\gamma_n} Y_t^2 + 2 \sup_{t \geq 1} E_{\gamma_n} \sup_{\theta \in \Theta} \left(\beta \sum_{j=0}^{t-1} \pi^j Y_{t-j-1} \right)^2 \\
\leq & \; 2 \sup_{n,t \geq 1} E_{\gamma_n} Y_t^2 + 2\beta_U^2 \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_+^{j+k} \sup_{n,t \geq 1, j,k \geq 0} E_{\gamma_n} |Y_{t-j-1} Y_{t-k-1}| < \infty. \qquad (13.66)
\end{aligned}
$$

This completes the proof of part (a).

Next, we establish part (b). By (13.61) and (13.62),

$$A_t(\psi_{0,n}, \pi) = Y_t, \;\; B_t(\psi_{0,n}, \pi) = 0, \;\text{ and }\; Q_n^{IC}(\psi_{0,n}, \pi) = 0. \qquad (13.67)$$

Hence, for part (b), it suffices to show that

$$\sup_{\pi \in \Pi} \sup_{\psi \in \Psi(\pi): ||\psi - \psi_{0,n}|| \leq \delta_n} \frac{|a_n^2(\gamma_n) Q_n^{IC}(\psi, \pi)|}{(1 + ||a_n(\gamma_n)(\psi - \psi_{0,n})||)^2} = o_p(1) \qquad (13.68)$$

for all constants $\delta_n \to 0$. The lhs of (13.68) is less than or equal to

$$\sup_{\theta \in \Theta: |\beta| \leq \delta_n} |n Q_n^{IC}(\theta)| = o_p(1), \qquad (13.69)$$

where the equality holds by (13.61) and (13.64)-(13.66) because (13.64) and (13.65) hold with β_U replaced by δ_n and $\delta_n \to 0$.

Lastly, we establish part (c). It suffices to show that

$$\sup_{\theta \in \Theta_n(\delta_n)} |Q_n^{IC}(\theta) - Q_n^{IC}(\theta_n)| = o_p(n^{-1}) \qquad (13.70)$$

for all $\delta_n \to 0$, where $\Theta_n(\delta_n) = \{\theta \in \Theta : \|\psi - \psi_n\| \leq \delta_n |\beta_n| \text{ and } |\pi - \pi_n| \leq \delta_n\}$.

Let $A_{t,n} = A_t(\theta_n)$ and $B_{t,n} = B_t(\theta_n)$.

First, suppose $\zeta = \zeta_n$. Then, using (13.61), we have

$$\begin{aligned}
& 2\zeta_L |Q_n^{IC}(\theta) - Q_n^{IC}(\theta_n)| \\
\leq\ & 2\zeta_L |Q_n^{\infty}(\theta) - Q_n(\theta) - Q_n^{\infty}(\theta_n) + Q_n(\theta_n)| \\
\leq\ & \left| n^{-1} \sum_{t=1}^{n} [-2A_t B_t + 2A_{t,n} B_{t,n} + B_t^2 - B_{t,n}^2] \right| \\
\leq\ & \left| n^{-1} \sum_{t=1}^{n} [-2A_t (B_t - B_{t,n}) - 2(A_t - A_{t,n}) B_{t,n} + B_t^2 - B_{t,n}^2] \right| \qquad (13.71) \\
\leq\ & 2n^{-1} \sum_{t=1}^{n} |A_t| \cdot |B_t - B_{t,n}| + 2n^{-1} \sum_{t=1}^{n} |A_t - A_{t,n}| \cdot |B_{t,n}| + \left| n^{-1} \sum_{t=1}^{n} (B_t^2 - B_{t,n}^2) \right|,
\end{aligned}$$

where the first inequality uses $\zeta = \zeta_n$.

To bound the first two terms on the rhs of (13.71), we have

$$\begin{aligned}
\sup_{\theta \in \Theta_n(\delta_n)} |A_t(\theta)| &\leq |Y_t| + \beta_U \sum_{j=0}^{\infty} \pi_+^{j-1} |Y_{t-j-1}|, \\
A_t(\theta) - A_t(\theta_n) &= -(\beta - \beta_n) \sum_{j=0}^{t-1} \pi^j Y_{t-j-1} - \beta_n \sum_{j=0}^{t-1} (\pi^j - \pi_n^j) Y_{t-j-1}, \\
\sup_{\theta \in \Theta_n(\delta_n)} |A_t(\theta) - A_t(\theta_n)| &\leq |\beta - \beta_n| \sum_{j=0}^{\infty} \pi_+^j |Y_{t-j-1}| + \beta_U \sum_{j=0}^{\infty} |\pi^j - \pi_n^j| \cdot |Y_{t-j-1}| \\
&\leq \delta_n \beta_U \sum_{j=1}^{\infty} [\pi_+^j + j \pi_+^{j-1}] |Y_{t-j-1}|, \qquad (13.72)
\end{aligned}$$

where the last inequality holds by mean-value expansions of π^j around π_n^j for $j \geq 1$ and

$\pi_+ = \max\{|\pi_L|, |\pi_U|\}$, and

$$B_t(\theta) = \beta \sum_{j=t}^{\infty} \pi^j Y_{t-j-1} = \beta \sum_{k=0}^{\infty} \pi^{t+k} Y_{-k-1},$$
$$|B_t(\theta) - B_t(\theta_n)| \leq \left| (\beta - \beta_n) \sum_{k=0}^{\infty} \pi^{t+k} Y_{-k-1} + \beta_n \sum_{k=0}^{\infty} (\pi^{t+k} - \pi_n^{t+k}) Y_{-k-1} \right|$$
$$\leq \delta_n \beta_U \sum_{k=0}^{\infty} \pi_+^{t+k} |Y_{-k-1}| + |\pi - \pi_n| \beta_U \sum_{k=0}^{\infty} (t+k) \pi_+^{t+k-1} |Y_{-k-1}|, \text{ and}$$
$$\sup_{\theta \in \Theta_n(\delta_n)} |B_t(\theta) - B_t(\theta_n)| \leq \delta_n \beta_U \pi_+^t \sum_{k=0}^{\infty} [\pi_+^k + (t+k) \pi_+^{k-1}] |Y_{-k-1}|, \tag{13.73}$$

where the second equality holds by change of variables and the second inequality holds by mean-value expansions of π^{t+k} around π_n^{t+k} for $k \geq 0$.

Using (13.72) and (13.73), we have the following bound on the expectation of the supremum over $\theta \in \Theta_n(\delta_n)$ of the first term on the rhs of (13.71):

$$2E_{\gamma_n} \sup_{\theta \in \Theta_n(\delta_n)} n^{-1} \sum_{t=1}^{n} |A_t(\theta)| \cdot |B_t(\theta) - B_t(\theta_n)|$$
$$\leq 2n^{-1} \delta_n \sum_{t=1}^{\infty} \pi_+^t \sum_{k=0}^{\infty} [\pi_+^k + \beta_U (t+k) \pi_+^{k-1}] E_{\gamma_n} |Y_t Y_{-k-1}| \tag{13.74}$$
$$+2n^{-1} \delta_n \beta_U \sum_{t=1}^{\infty} \pi_+^t \sum_{j=0}^{\infty} \pi_+^{j-1} \sum_{k=0}^{\infty} [\pi_+^k + \beta_U (t+k) \pi_+^{k-1}] E_{\gamma_n} |Y_{t-j-1} Y_{-k-1}| = o(n^{-1})$$

using $E_{\gamma_n} |Y_{t-j-1} Y_{-k-1}| \leq \sup_{n \geq 1} E_{\gamma_n} Y_1^2 \leq C < \infty$ and $\pi_+ \in (0,1)$. By Markov's inequality, (13.74) implies that the lhs quantity with E_{γ_n} deleted is $o_p(n^{-1})$, as desired.

Similarly, using (13.72) and (13.73), we have the following bound on the expectation of the supremum over $\theta \in \Theta_n(\delta_n)$ of the second term on the rhs of (13.71):

$$E_{\gamma_n} \sup_{\theta \in \Theta_n(\delta_n)} n^{-1} \sum_{t=1}^{n} |A_t(\theta) - A_t(\theta_n)| \cdot |B_t(\theta_n)| \tag{13.75}$$
$$\leq n^{-1} \delta_n \beta_U^2 \sum_{t=1}^{\infty} \pi_+^t \sum_{j=1}^{\infty} [\pi_+^j + j \pi_+^{j-1}] \sum_{k=0}^{\infty} \pi_+^k \sup_{n, t \geq 1, j, k \geq 0} E_{\gamma_n} |Y_{t-j-1} Y_{-k-1}| = o(n^{-1}).$$

Hence, the lhs of (13.75) with E_{γ_n} deleted is $o_p(n^{-1})$.

Next, we consider the third term on the rhs of (13.71):

$$\begin{aligned}
&n^{-1}\sum_{t=1}^{n}(B_t^2(\theta)-B_t^2(\theta_n)) \\
&=\beta^2 n^{-1}\sum_{t=1}^{n}\left(\sum_{k=0}^{\infty}\pi^{t+k}Y_{-k-1}\right)^2-\beta_n^2 n^{-1}\sum_{t=1}^{n}\left(\sum_{k=0}^{\infty}\pi_n^{t+k}Y_{-k-1}\right)^2 \qquad (13.76)\\
&=(\beta^2-\beta_n^2)n^{-1}\sum_{t=1}^{n}\left(\sum_{j=0}^{\infty}\pi^{t+j}Y_{-j-1}\right)^2+\beta_n^2 n^{-1}\sum_{t=1}^{n}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(\pi^{t+j+k}-\pi_n^{t+j+k})Y_{-j-1}Y_{-k-1}.
\end{aligned}$$

The supremum over $\theta\in\Theta_n(\delta_n)$ of the absolute value of the first term on the rhs of (13.76) is $O_p(\sup_{\theta\in\Theta_n(\delta_n)}|\beta^2-\beta_n^2|n^{-1})=o_p(n^{-1})$ by calculations analogous to those in (13.64) and (13.65). The expectation of the supremum over $\theta\in\Theta_n(\delta_n)$ of the absolute value of the second term on the rhs of (13.76) is bounded by

$$\beta_U^2 n^{-1}\sum_{t=1}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sup_{|\pi-\pi_n|\le\delta_n}|\pi^{t+j+k}-\pi_n^{t+j+k}|\cdot\sup_{n\ge1}E_{\gamma_n}Y_1^2=o(n^{-1}). \qquad (13.77)$$

The equality in (13.77) holds because

$$\begin{aligned}
&\sum_{t=1}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sup_{|\pi-\pi_n|\le\delta_n}|\pi^{t+j+k}-\pi_n^{t+j+k}| \\
&\le\sup_{|\pi-\pi_n|\le\delta_n}|\pi-\pi_n|\sum_{t=1}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(t+j+k)\pi_+^{t+j+k-1}=o(1), \qquad (13.78)
\end{aligned}$$

where the inequality holds by mean-value expansions of π^{t+j+k} around π_n^{t+j+k} for $t\ge1$, $j,k\ge0$ and the equality holds because $\pi_+\in(0,1)$. Equation (13.77) implies that the supremum over $\theta\in\Theta_n(\delta_n)$ of the absolute value of the second term on the rhs of (13.76) is $o_p(n^{-1})$. Hence, we conclude that the supremum over $\theta\in\Theta_n(\delta_n)$ of the absolute value of the lhs of (13.76), which is the third summand in (13.71), is $o_p(n^{-1})$.

This completes the verification of (13.70) for the case where $\zeta=\zeta_n$.

Lastly, we consider the case where $\zeta\neq\zeta_n$. We have

$$|Q_n^{IC}(\theta)-Q_n^{IC}(\theta_n)|=|Q_n^{IC}(\theta)-Q_n^{IC}(\beta_n,\zeta,\pi_n)|+|Q_n^{IC}(\beta_n,\zeta,\pi_n)-Q_n^{IC}(\beta_n,\zeta_n,\pi_n)|. \qquad (13.79)$$

The proof of part (c) for the case where $\zeta=\zeta_n$ gives $\sup_{\theta\in\Theta_n(\delta_n)}|Q_n^{IC}(\theta)-Q_n^{IC}(\beta_n,\zeta,\pi_n)|$

$= o_p(n^{-1})$. It remains to show

$$\sup_{\theta\in\Theta_n(\delta_n)} |Q_n^{IC}(\beta_n,\zeta,\pi_n) - Q_n^{IC}(\beta_n,\zeta_n,\pi_n)| = o_p(n^{-1}). \tag{13.80}$$

We have

$$\begin{aligned} Q_n^{IC}(\beta_n,\zeta,\pi_n) &= Q_n(\beta_n,\zeta,\pi_n) - Q_n^{\infty}(\beta_n,\zeta,\pi_n) \\ &= \frac{1}{2\zeta} n^{-1}\sum_{t=1}^{n}\left(Y_t - \beta_n\sum_{j=0}^{t-1}\pi_n^j Y_{t-j-1}\right)^2 - \frac{1}{2\zeta}n^{-1}\sum_{t=1}^{n}\varepsilon_t^2 \qquad (13.81)\\ &= \frac{1}{2\zeta} n^{-1}\sum_{t=1}^{n}\left(\varepsilon_t + \beta_n\sum_{j=t}^{\infty}\pi_n^j Y_{t-j-1}\right)^2 - \frac{1}{2\zeta}n^{-1}\sum_{t=1}^{n}\varepsilon_t^2 \\ &= \frac{1}{\zeta}n^{-1}\sum_{t=1}^{n}\varepsilon_t\beta_n\sum_{k=0}^{\infty}\pi_n^{t+k}Y_{-k-1} + \frac{1}{2\zeta}n^{-1}\sum_{t=1}^{n}\left(\beta_n\sum_{k=0}^{\infty}\pi_n^{t+k}Y_{-k-1}\right)^2. \end{aligned}$$

The quantity $Q_n^{IC}(\beta_n,\zeta_n,\pi_n)$ is the same, but with ζ_n in place of ζ. Hence,

$$\begin{aligned} &|Q_n^{IC}(\beta_n,\zeta,\pi_n) - Q_n^{IC}(\beta_n,\zeta_n,\pi_n)| \qquad (13.82)\\ \le\ & \frac{|\zeta-\zeta_n|}{\zeta\zeta_n}\left|n^{-1}\sum_{t=1}^{n}\varepsilon_t\beta_n\sum_{k=0}^{\infty}\pi_n^{t+k}Y_{-k-1}\right| + \frac{|\zeta-\zeta_n|}{2\zeta\zeta_n}n^{-1}\sum_{t=1}^{n}\left(\beta_n\sum_{k=0}^{\infty}\pi_n^{t+k}Y_{-k-1}\right)^2. \end{aligned}$$

We have

$$\begin{aligned} &E_{\gamma_n}\sup_{\theta\in\Theta_n(\delta_n)}\left|n^{-1}\sum_{t=1}^{n}\varepsilon_t\beta_n\sum_{k=0}^{\infty}\pi_n^{t+k}Y_{-k-1}\right| \\ \le\ & n^{-1}\beta_U\sum_{t=1}^{\infty}\sum_{k=0}^{\infty}\pi_+^{t+k}\sup_{n\ge1,k\ge0}E_{\gamma_n}|\varepsilon_tY_{-k-1}| = O(n^{-1}), \qquad (13.83) \end{aligned}$$

where $\pi_+ = \max\{|\pi_L|,|\pi_U|\}$, and

$$\begin{aligned} &E_{\gamma_n}n^{-1}\sum_{t=1}^{n}\left(\beta_n\sum_{k=0}^{\infty}\pi_n^{t+k}Y_{-k-1}\right)^2 \\ \le\ & n^{-1}\beta_U^2\sum_{t=1}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi_+^{t+j}\pi_+^{t+k}\sup_{n\ge1,j,k\ge0}E_{\gamma_n}|Y_{-j-1}Y_{-k-1}| = O(n^{-1}). \qquad (13.84) \end{aligned}$$

Equations (13.83) and (13.84) and Markov's inequality, coupled with (13.82) and $\sup_{\theta \in \Theta_n(\delta_n)} |\zeta - \zeta_n| \leq \delta_n = o(1)$, establish (13.80), which completes the proof of part (c). □

14. Appendix D: Additional ARMA(1, 1) Monte Carlo Simulations

This Appendix provides details concerning the ARMA(1, 1) simulations computations. It also provides additional simulation results.

14.1. Simulation Details

To achieve an approximately stationary start-up, the first innovation is set equal to 0 and the first 200 realizations of the process are discarded. For purposes of speed, matrix/vector calculations are employed to compute the time series Y_t and the log likelihood. In these calculations, lags are truncated at 100.

The matlab function *fmincon* is used in all cases where optimization is required. When the optimization is in more than one dimension, such as with the finite-sample unconstrained optimization, six independent random starting values are used. The random starting values are uniformly distributed in the parameter space of the parameters. When the optimization is one dimensional, such as with the asymptotic results and with the finite-sample constrained optimization, the starting value for the *fmincon* function is obtained by a grid search. In all cases, the grids divide the optimization parameter space into 50 intervals of equal length.

For the finite-sample and asymptotic results for both the MA and AR parameters, the constrained and unconstrained criterion functions often are found to have multiple local minimum for small values of $|b|$. Hence, the grid search and multiple starting values are useful.

In all figures concerning the MA parameter π for which the x axis is b or $|b|$, such as Figures 6-10 of AC1, the discrete values of b for which computations are made run from 0 to -20 (although only values from 0 to -15 are reported), with a grid of 0.1 for b between 0 and -5, a grid of 0.2 for b between -5 and -10, and a grid of 1 for b between -10 and -20. For the analogous figures concerning the AR parameter ρ, the same grids are used but the b values are non-negative.

For the finite-sample simulations concerning the MA parameter, for each b, the true value of β is $\beta_n = -b/\sqrt{n}$ and the AR parameter is $\rho_n = \pi_0 + \beta_n = \pi_0 - b/\sqrt{n}$. The value of b is restricted such that ρ_n belongs to its true parameter space, i.e., $\rho_n \in [-0.85, 0.85]$. Note that the b values are negative. Positive values of b also could be considered, but if π_0 is positive, then the range of positive b values is more restricted (by the requirement that $\rho_n \in [-0.85, 0.85]$) than the range of negative b values.

For the finite-sample simulations concerning the AR parameter, for each b, the true value of β is $\beta_n = b/\sqrt{n}$ and the MA parameter is $\pi_n = \rho_0 - \beta_n = \pi_0 - b/\sqrt{n}$. The value of b is restricted such that π_n belongs to its true parameter space, i.e., $\pi_n \in [-0.8, 0.8]$.

In Figures 1 and 2 of AC1 and Figure S-1 below, the asymptotic density of the ML estimator of the MA parameter π is given by $\pi^*(\gamma_0, b)$ $(= \arg\min_{\pi \in \Pi} \xi(\pi; \gamma_0, b))$ for $b = 0$, -2, -4, and -12. Similarly, in Figures S-9 to S-11 below, the asymptotic density of the ML estimator of the AR parameter $\rho = \pi + \beta$ is given by $\pi^*(\gamma_0, b)$ for $b = 0$, 2, 4, and 12 (because its asymptotic distribution is the same as that of the MA parameter when $|b| < \infty$).

In Figure 3 of AC1, the asymptotic density of the ML estimator of β centered at the true value is equal to the first element of $\tau(\pi^*(\gamma_0, b); \gamma_0, b)$ divided by $n^{1/2}$ with $n = 250$, so that it has the same scale as the finite-sample ($n = 250$) estimator. In this ARMA example, the first element of $\tau(\pi^*(\gamma_0, b); \gamma_0, b)$ equals

$$-(1 - \pi^2)\left(\sum_{j=0}^{\infty} \pi^j Z_j - (1 - \pi_0 \pi)^{-1} b\right) + b. \tag{14.1}$$

Figures that give densities for the estimators of π and ρ are constructed using histograms with 40 bins. Figures that give densities for the estimator of β and for the test statistics use 100 bins. The areas under the histograms equal one.

14.2. Additional Simulation Results

Figures S-1 to S-9 provide additional results concerning the MA parameter π. Figures S-10 to S-23 provide results concerning the AR parameter ρ. Tables S-I to S-VI provide results for CI's concerning π and CI's concerning ρ. See AC1 for some discussion of the results in these Figures and Tables.

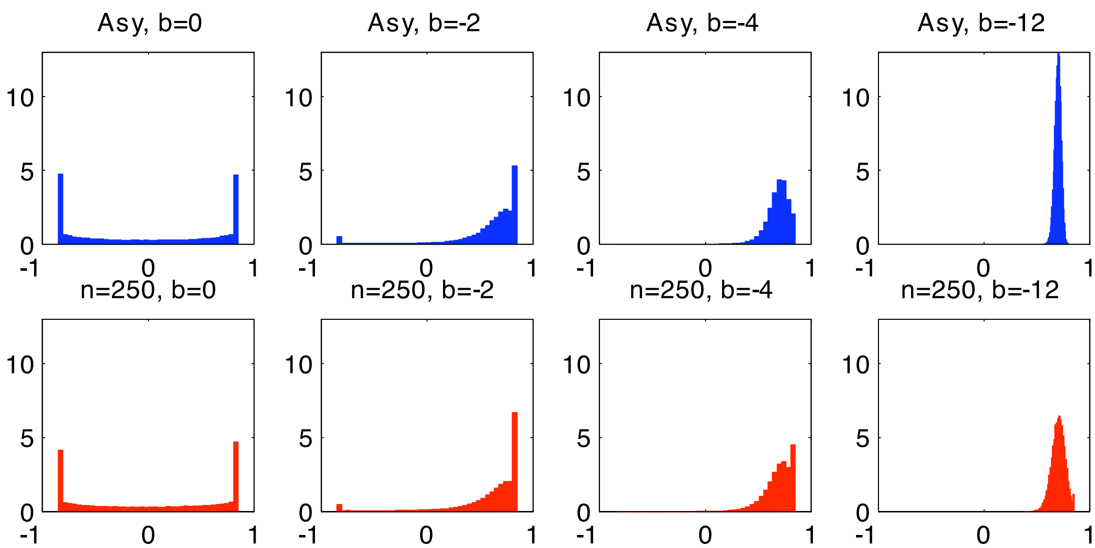

Figure S-1. Asymptotic and Finite-Sample (n=250) Densities of the Estimator of the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.7$.

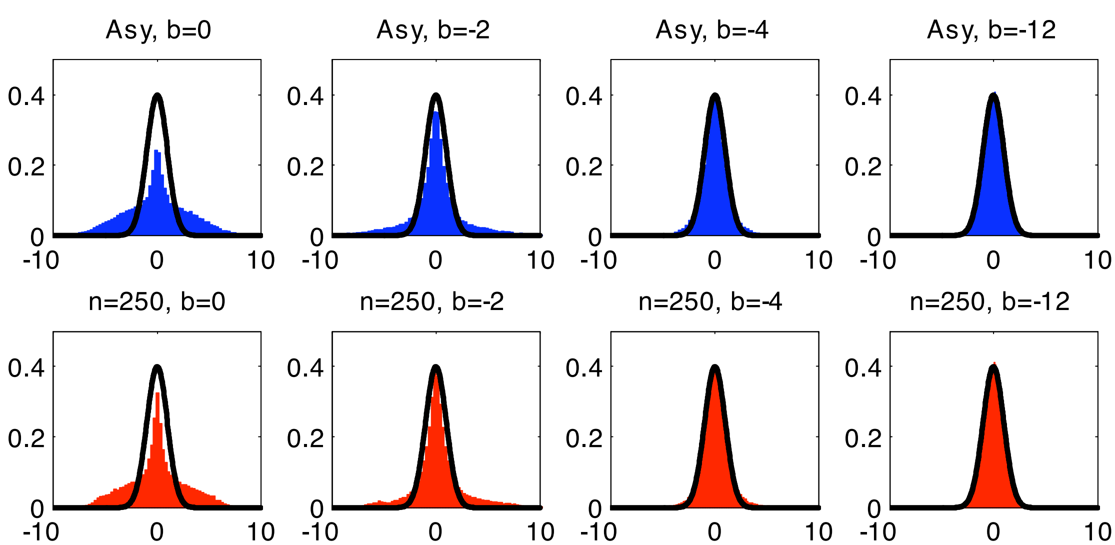

Figure S-2. Asymptotic and Finite-Sample (n=250) Densities of the t Statistic for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0$ and the Standard Normal Density (Black Line).

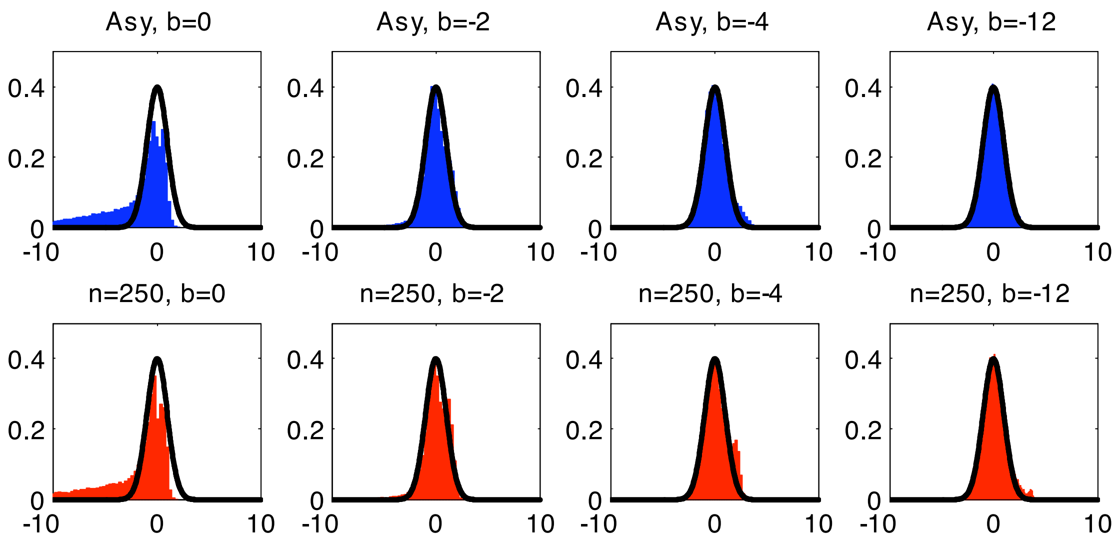

Figure S-3. Asymptotic and Finite-Sample (n=250) Densities of the t Statistic for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.7$ and the Standard Normal Density (Black Line).

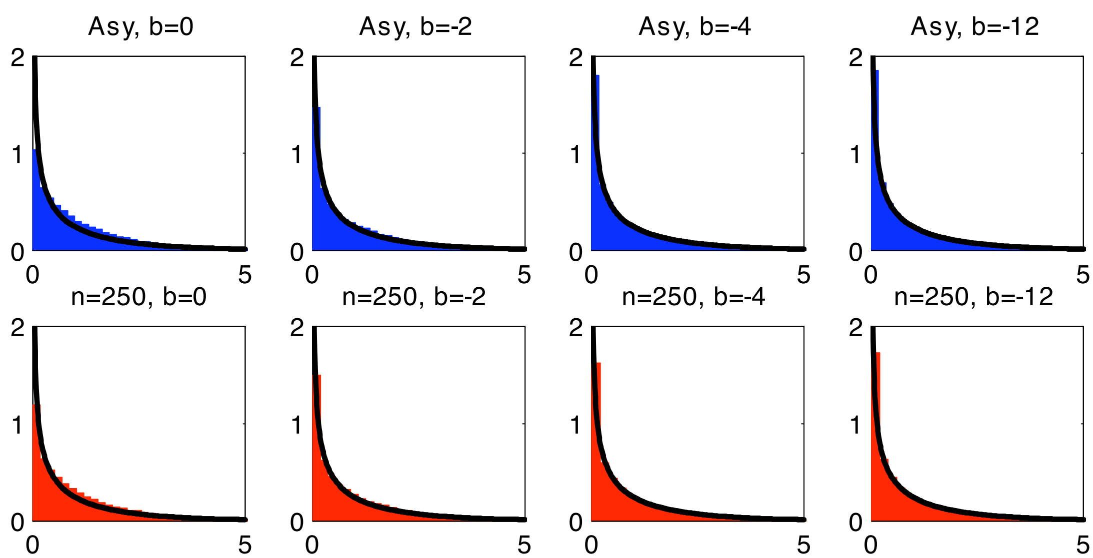

Figure S-4. Asymptotic and Finite-Sample (n=250) Densities of the QLR Statistic for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0$ and the χ^2_1 Density (Black Line).

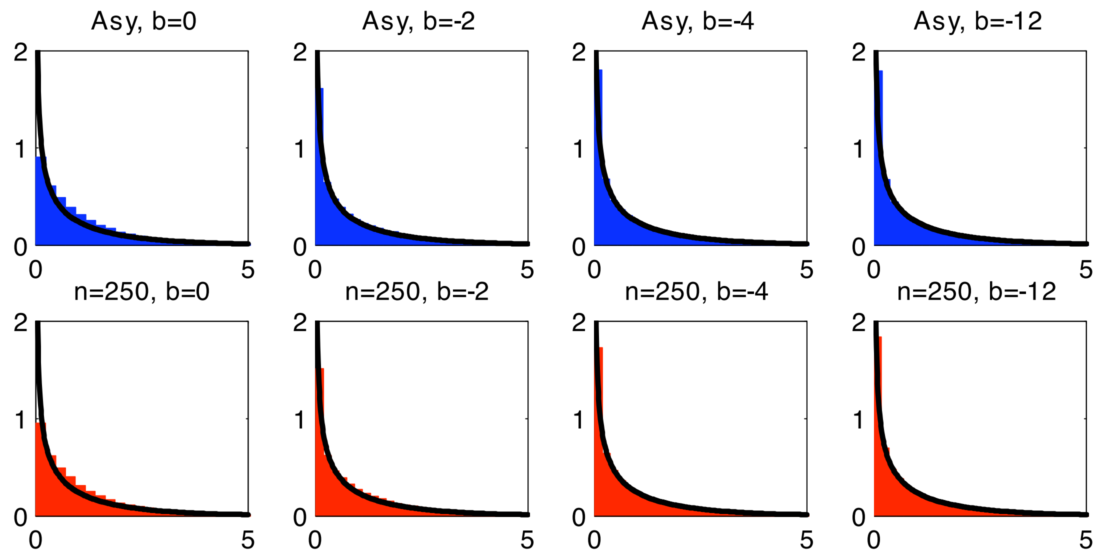

Figure S-5. Asymptotic and Finite-Sample (n=250) Densities of the QLR Statistic for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.7$ and the χ_1^2 Density (Black Line).

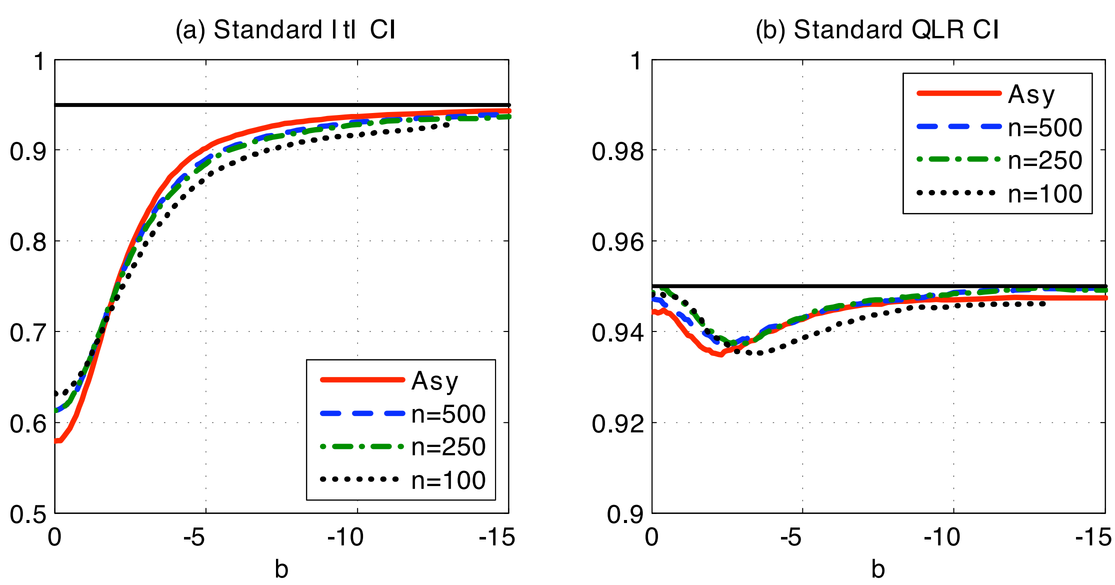

Figure S-6. Coverage Probabilities of Standard $|t|$ and QLR CI's for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.4$.

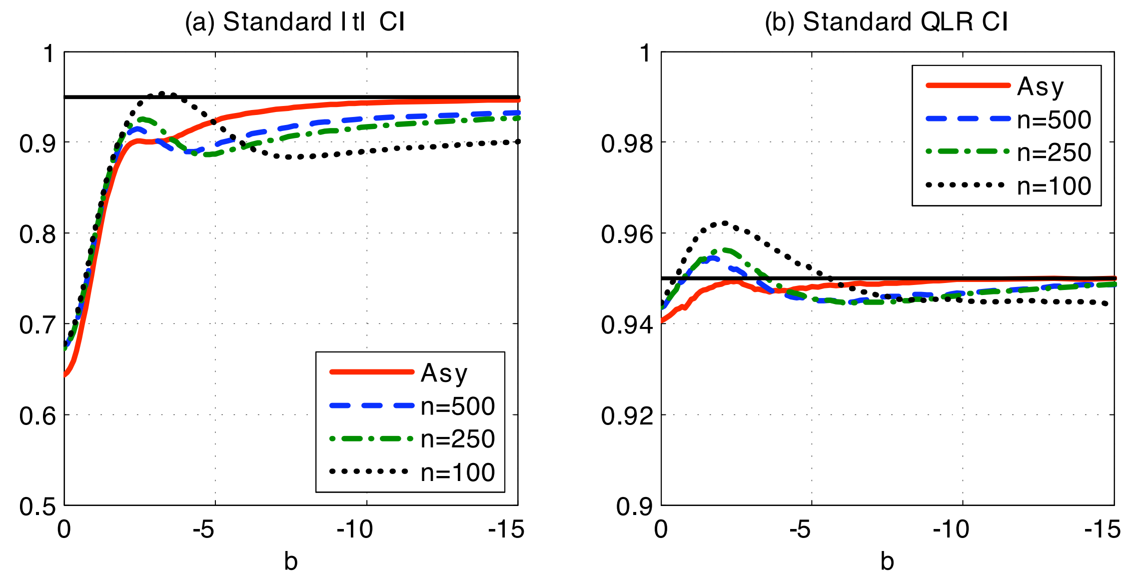

Figure S-7. Coverage Probabilities of Standard $|t|$ and QLR CI's for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.7$.

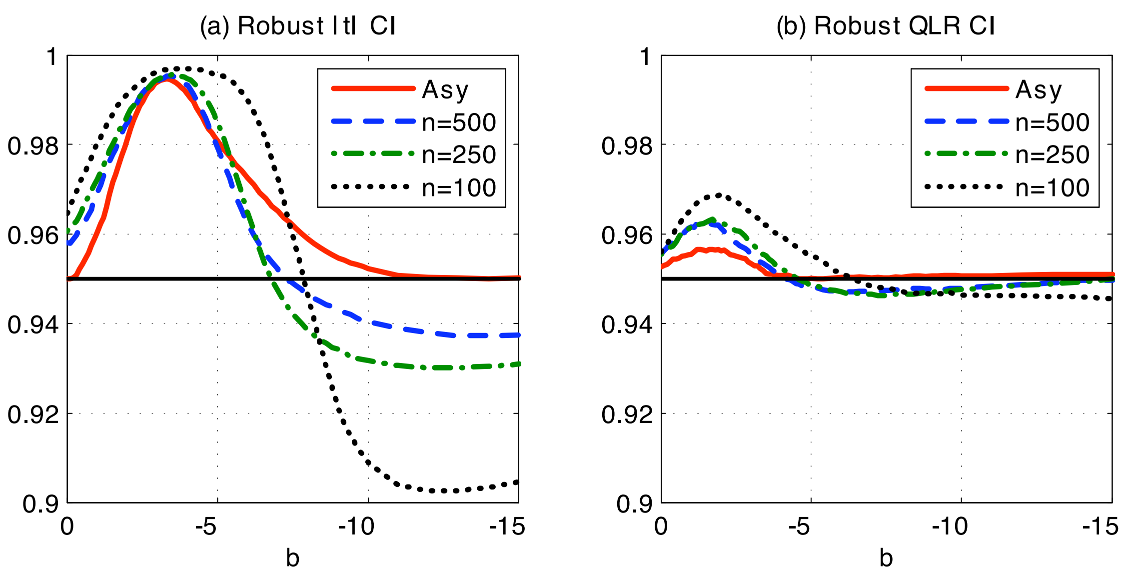

Figure S-8. Coverage Probabilities of Robust $|t|$ and QLR CI's for the MA Parameter π in the ARMA(1, 1) Model when $\pi_0 = 0.7$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

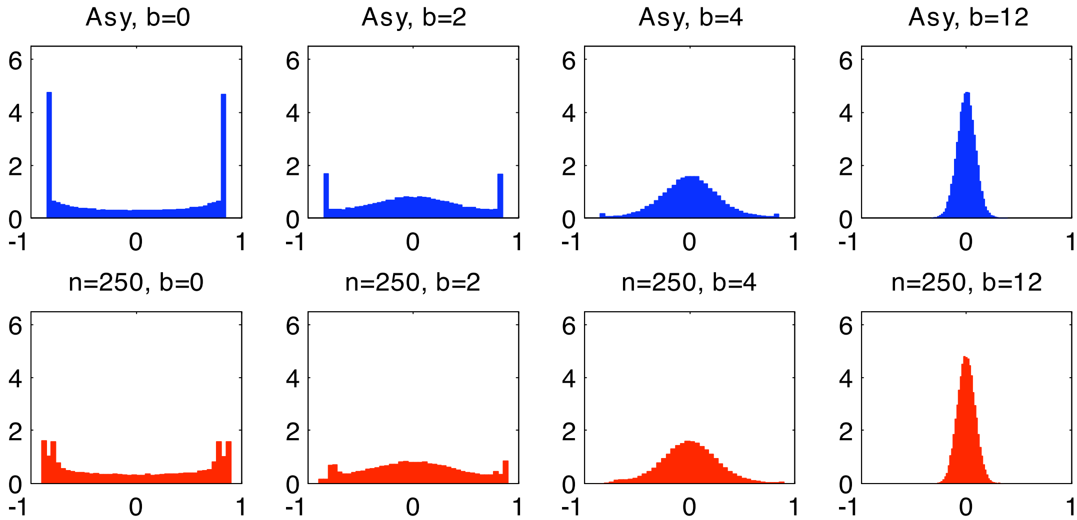

Figure S-9. Asymptotic and Finite-Sample (n=250) Densities of the Estimator of the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0$.

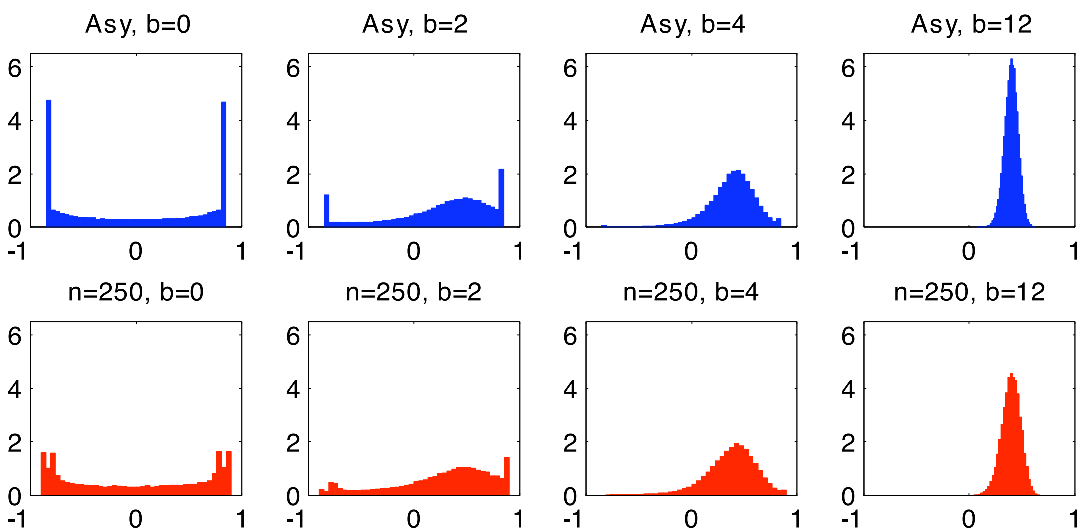

Figure S-10. Asymptotic and Finite-Sample (n=250) Densities of the Estimator of the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.4$.

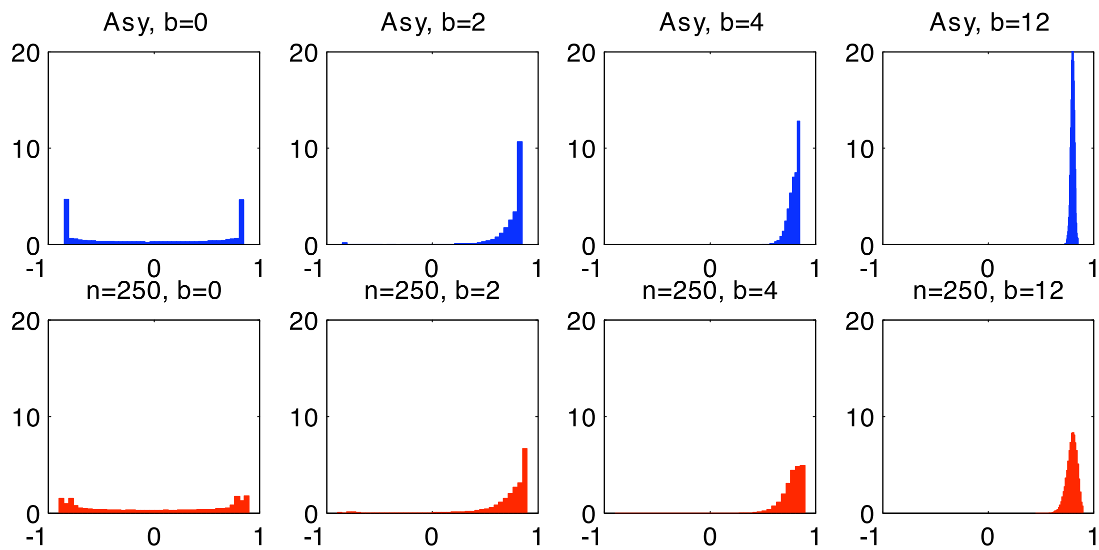

Figure S-11. Asymptotic and Finite-Sample (n=250) Densities of the Estimator of the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.8$.

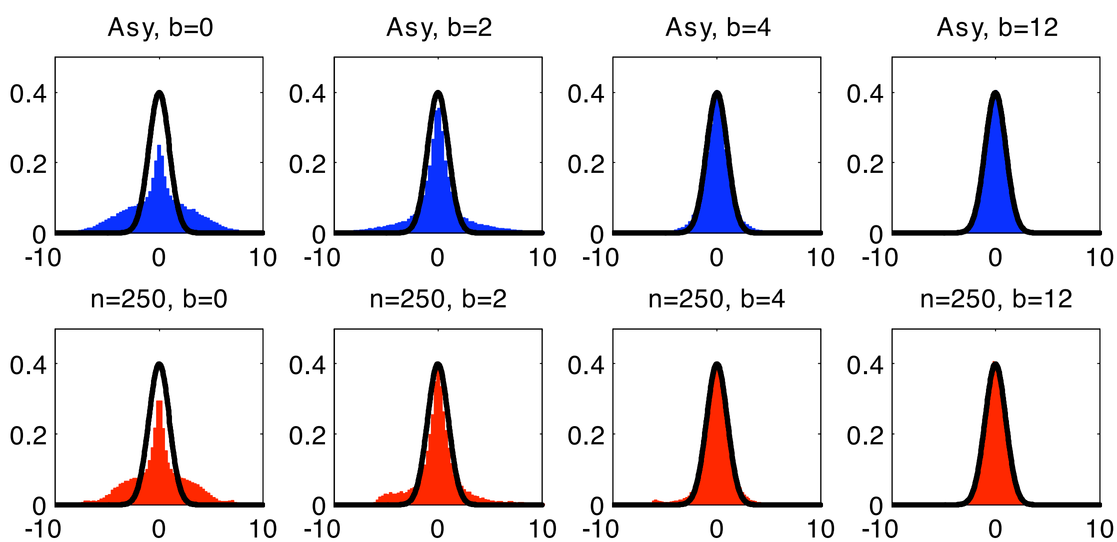

Figure S-12. Asymptotic and Finite-Sample (n=250) Densities of the t Statistic for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0$ and the Standard Normal Density (Black Line).

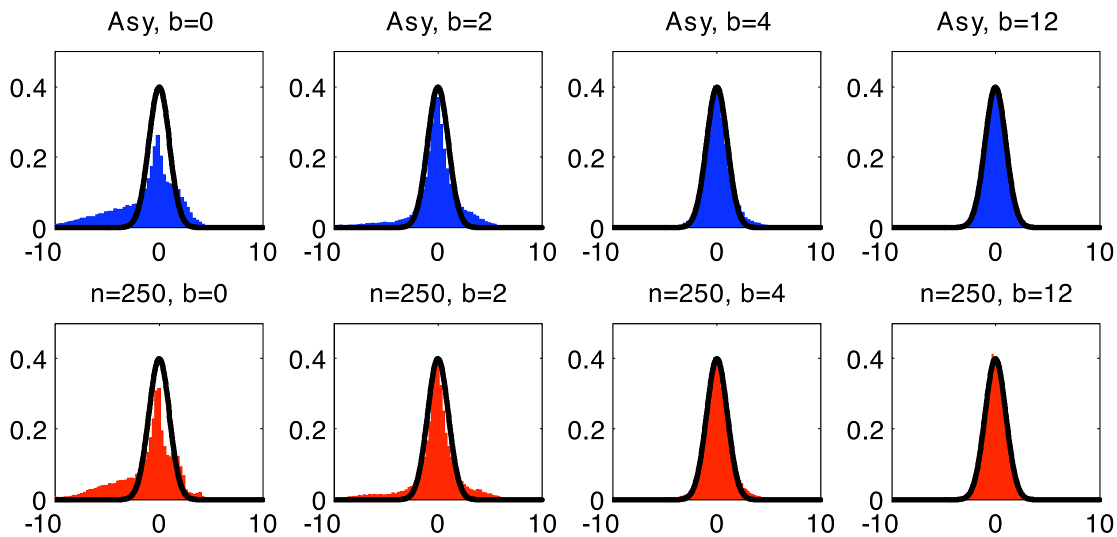

Figure S-13. Asymptotic and Finite-Sample (n=250) Densities of the t Statistic for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.4$ and the Standard Normal Density (Black Line).

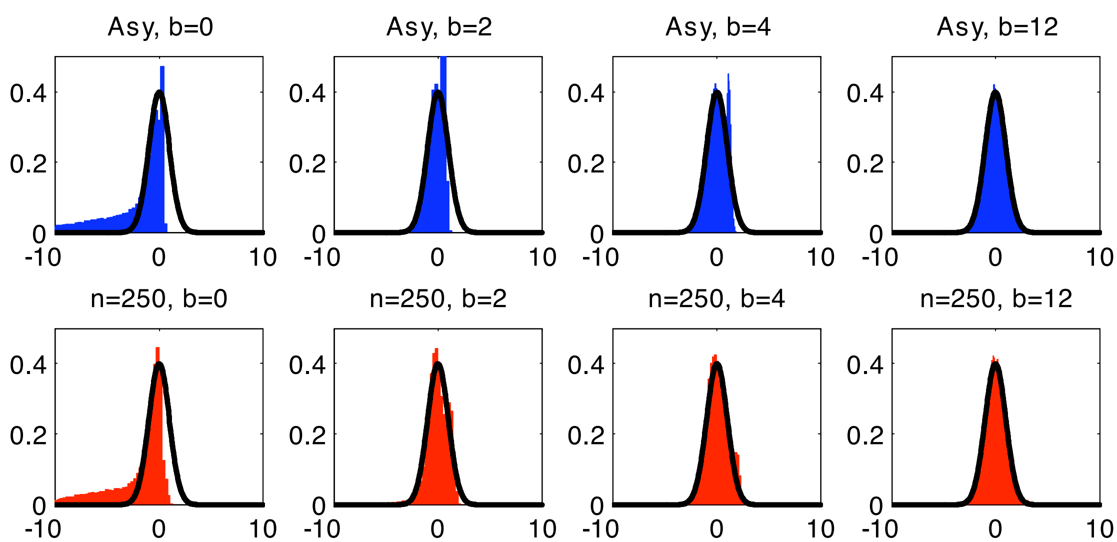

Figure S-14. Asymptotic and Finite-Sample (n=250) Densities of the t Statistic for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.8$ and the Standard Normal Density (Black Line).

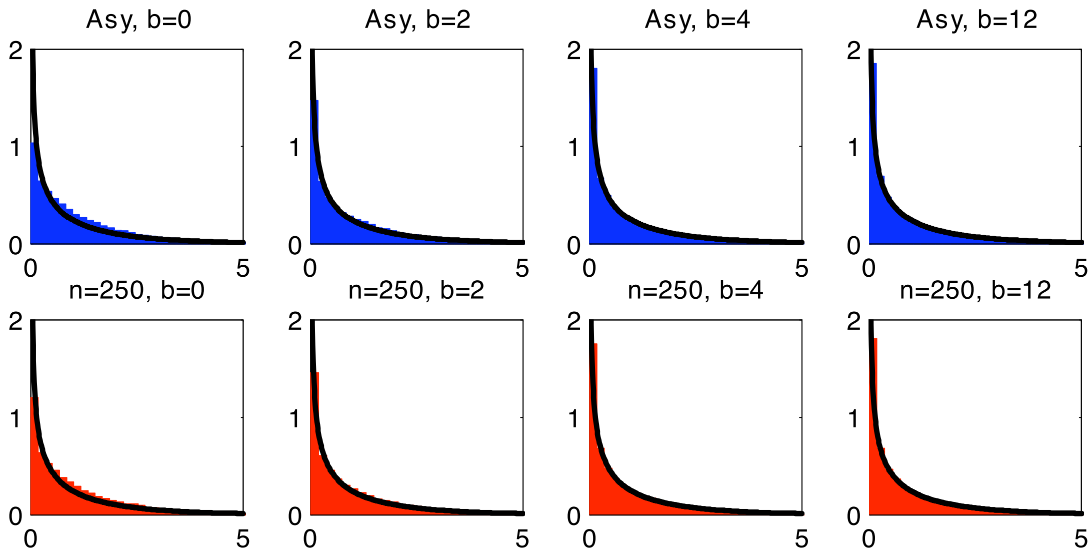

Figure S-15. Asymptotic and Finite-Sample (n=250) Densities of the QLR Statistic for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0$ and the χ_1^2 Density (Black Line).

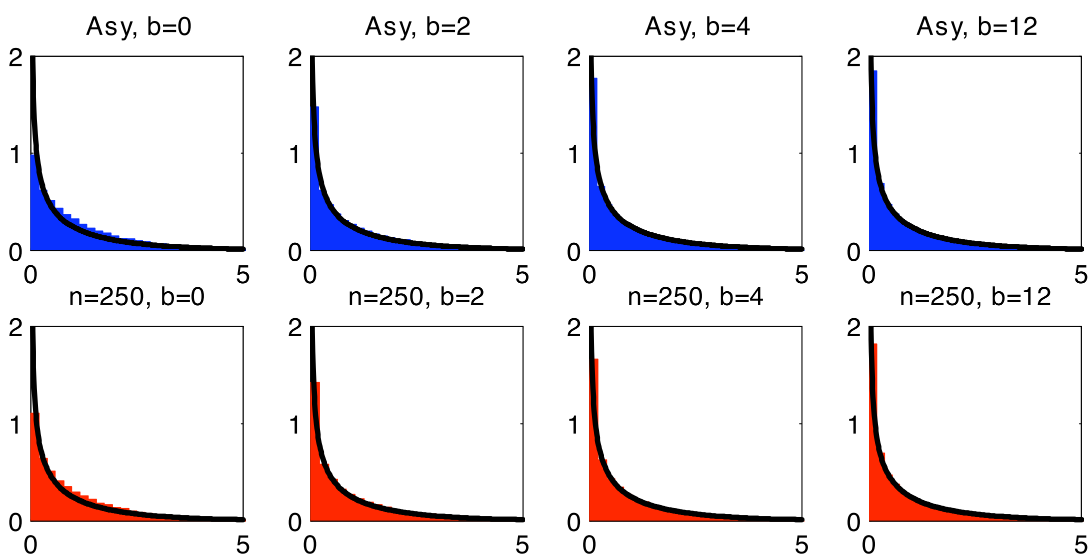

Figure S-16. Asymptotic and Finite-Sample (n=250) Densities of the QLR Statistic for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.4$ and the χ_1^2 Density (Black Line).

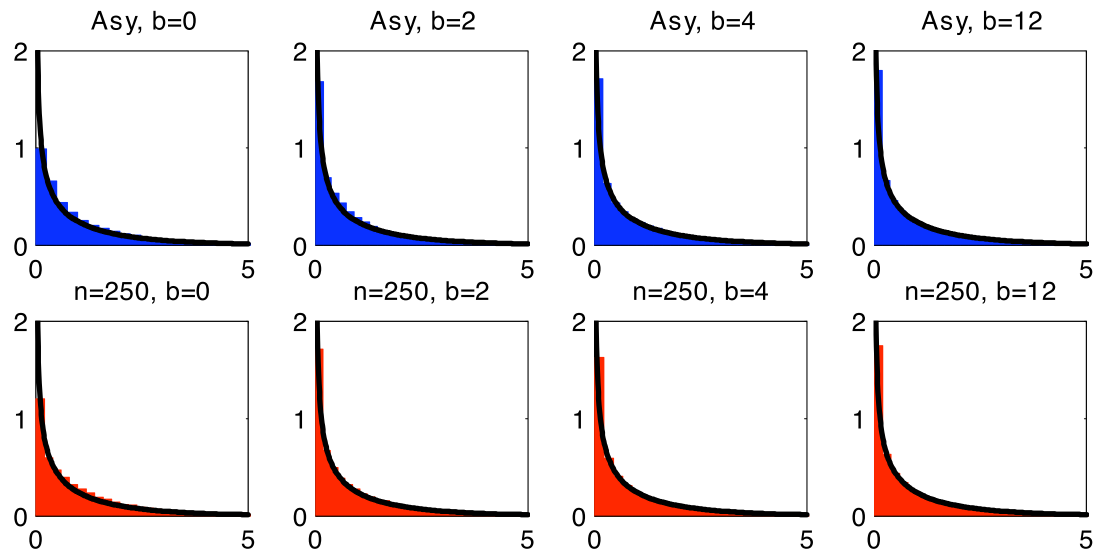

Figure S-17. Asymptotic and Finite-Sample (n=250) Densities of the QLR Statistic for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.8$ and the χ^2_1 Density (Black Line).

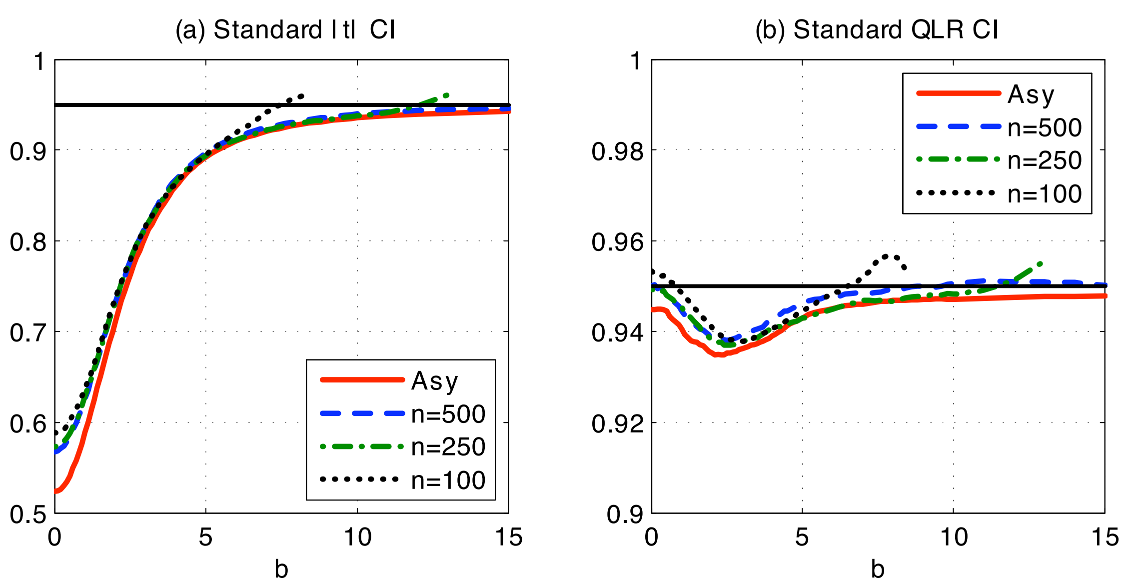

Figure S-18. Coverage Probabilities of Standard $|t|$ and QLR CI's for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0$.

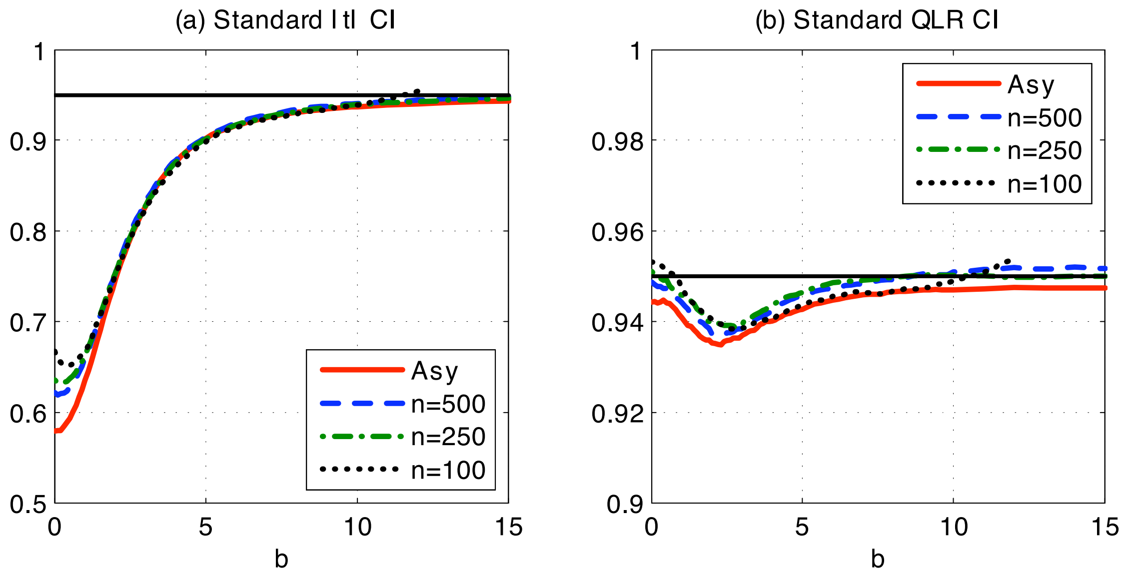

Figure S-19. Coverage Probabilities of Standard $|t|$ and QLR CI's for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.4$.

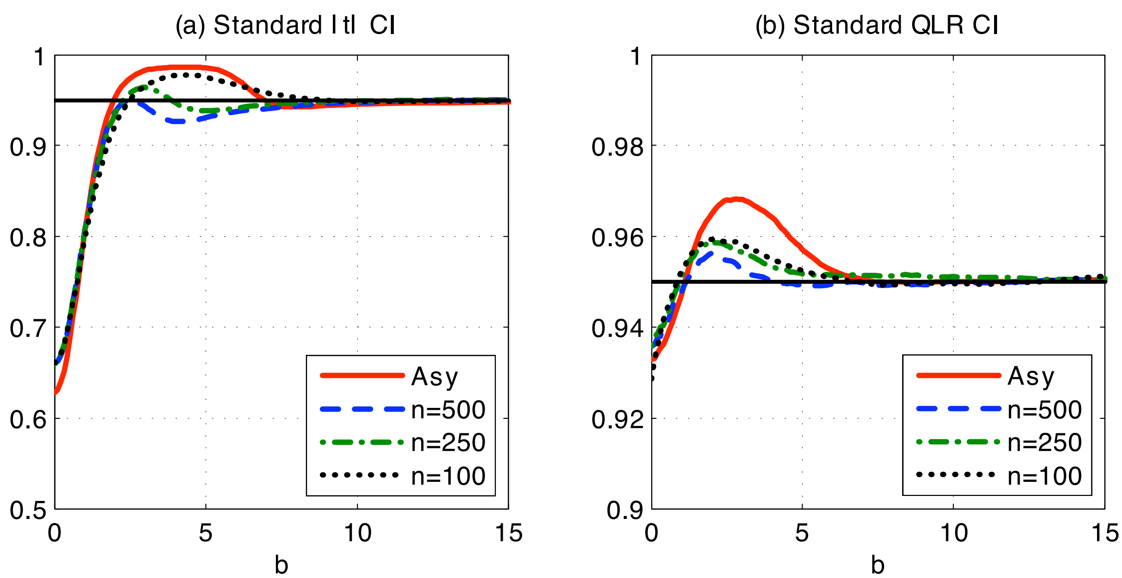

Figure S-20. Coverage Probabilities of Standard $|t|$ and QLR CI's for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.8$.

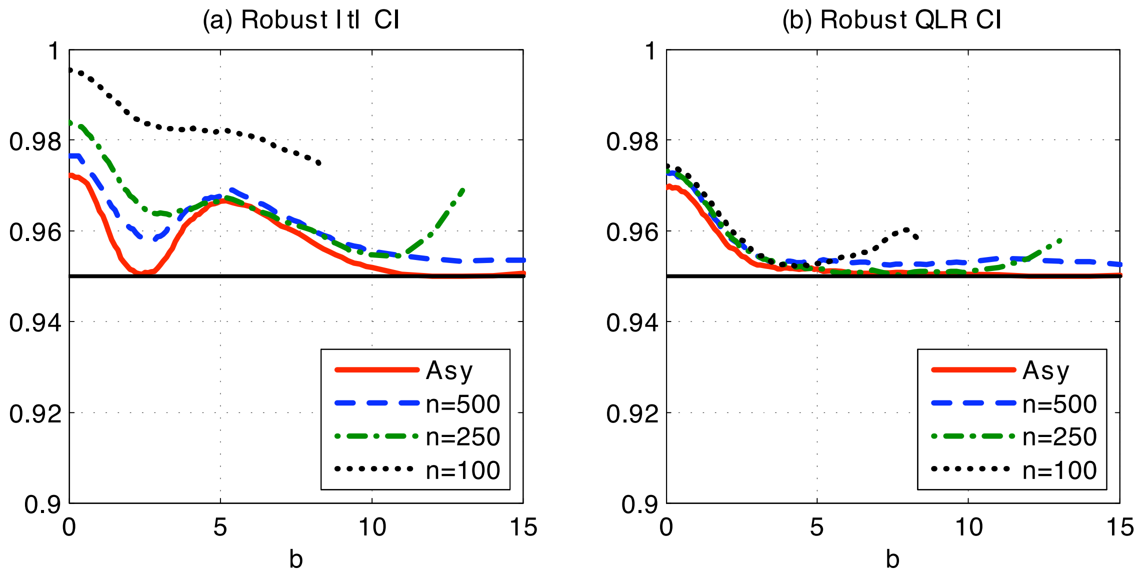

Figure S-21. Coverage Probabilities of Robust $|t|$ and QLR CI's for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

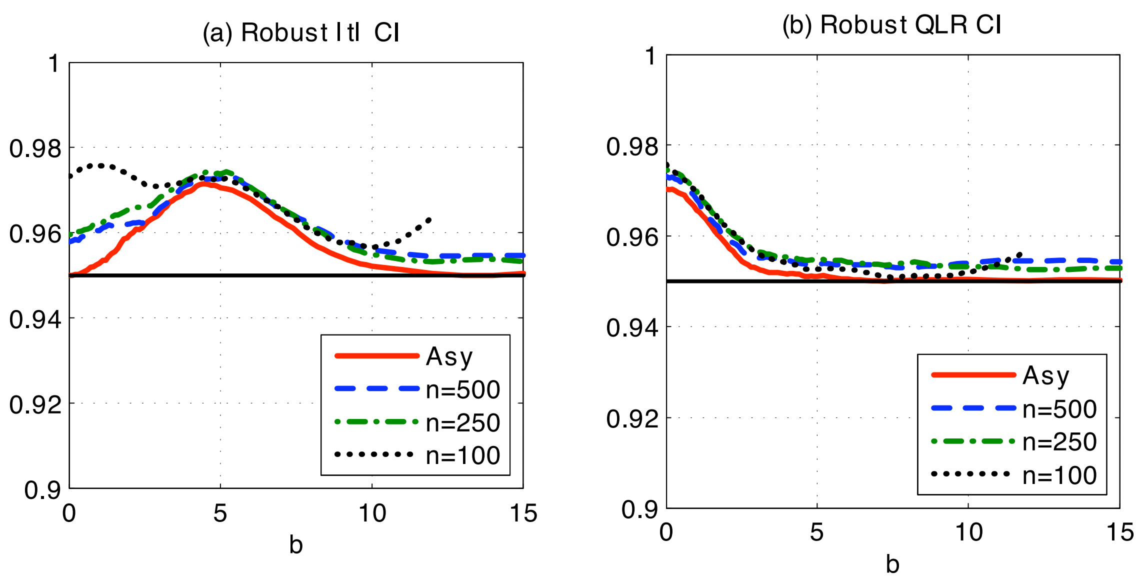

Figure S-22. Coverage Probabilities of Robust $|t|$ and QLR CI's for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.4$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

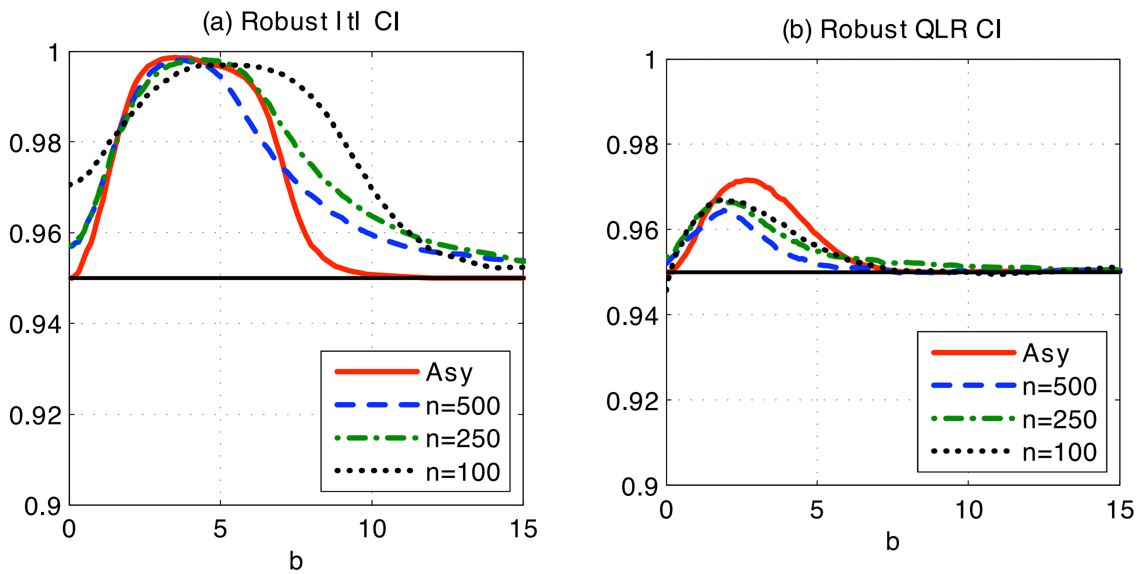

Figure S-23. Coverage Probabilities of Robust $|t|$ and QLR CI's for the AR Parameter ρ in the ARMA(1, 1) Model when $\rho_0 = 0.8$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

Table S-I. Finite-Sample Coverage Probabilities (Minimum over b) of Nominal 95% CI's for π and ρ in the ARMA(1, 1) Model, $n = 100$, 500

		$\|t\|$			QLR		
		Std	LF	Rob	Std	LF	Rob
$n = 100$							
MA	$\pi_0 = 0.0$	0.572	0.970	0.956	0.936	0.950	0.950
	$\pi_0 = 0.4$	0.630	0.971	0.933	0.935	0.951	0.948
	$\pi_0 = 0.7$	0.678	0.972	0.903	0.944	0.953	0.946
AR	$\rho_0 = 0.0$	0.589	0.982	0.974	0.938	0.954	0.953
	$\rho_0 = 0.4$	0.651	0.982	0.957	0.938	0.953	0.952
	$\rho_0 = 0.8$	0.661	0.982	0.952	0.929	0.947	0.946
$n = 500$							
MA	$\pi_0 = 0.0$	0.565	0.956	0.951	0.935	0.951	0.951
	$\pi_0 = 0.4$	0.613	0.958	0.946	0.937	0.952	0.951
	$\pi_0 = 0.7$	0.676	0.959	0.937	0.944	0.953	0.947
AR	$\rho_0 = 0.0$	0.567	0.965	0.953	0.938	0.952	0.953
	$\rho_0 = 0.4$	0.619	0.962	0.955	0.937	0.952	0.953
	$\rho_0 = 0.8$	0.662	0.961	0.953	0.936	0.952	0.950

Table S-II. Finite-Sample False Coverage Probabilities of Robust $|t|$ CI's for the MA Parameter π for Different Values of κ in the ARMA(1, 1) Model, $n = 500$

	$\pi_0 = 0.0$				$\pi_0 = 0.4$				$\pi_0 = 0.7$				
b	-2	-5	-10	$-\infty$	-2	-5	-10	$-\infty$	-2	-5	-10	$-\infty$	Avg
π_{H_0}	0.800	0.740	0.220	0.110	0.000	0.000	0.210	0.293	0.000	0.410	0.580	0.623	
LF	0.968	0.994	1.000	1.000	0.928	0.957	0.997	1.000	0.760	0.958	1.000	1.000	0.964
κ													
0.00	0.944	0.395	0.483	0.490	0.912	0.628	0.506	0.512	0.682	0.433	0.491	0.504	0.582
0.50	0.944	0.395	0.483	0.490	0.912	0.628	0.506	0.512	0.682	0.433	0.491	0.504	0.582
1.00	0.944	0.395	0.483	0.490	0.911	0.627	0.506	0.512	0.681	0.433	0.491	0.504	0.581
1.50	0.947	0.415	0.483	0.490	0.911	0.627	0.506	0.512	0.681	0.444	0.493	0.503	0.584
1.75	0.954	0.455	0.484	0.490	0.911	0.627	0.507	0.511	0.680	0.465	0.496	0.503	0.590
2.00	0.958	0.498	0.486	0.489	0.916	0.641	0.508	0.509	0.697	0.490	0.500	0.503	0.600
2.25	0.962	0.544	0.490	0.488	0.917	0.659	0.511	0.508	0.706	0.516	0.504	0.503	0.609
2.50	0.964	0.594	0.495	0.487	0.919	0.680	0.515	0.508	0.718	0.545	0.510	0.503	0.620
2.75	0.966	0.643	0.501	0.486	0.921	0.706	0.520	0.507	0.731	0.576	0.517	0.503	0.631
3.00	0.967	0.694	0.508	0.485	0.924	0.731	0.525	0.506	0.739	0.609	0.524	0.502	0.643
4.00	0.968	0.870	0.547	0.482	0.928	0.831	0.555	0.504	0.758	0.751	0.560	0.503	0.688
5.00	0.968	0.963	0.610	0.480	0.928	0.909	0.603	0.502	0.760	0.878	0.619	0.503	0.727
6.00	0.968	0.990	0.707	0.480	0.928	0.946	0.671	0.501	0.760	0.940	0.697	0.503	0.758
8.00	0.968	0.994	0.936	0.479	0.928	0.957	0.851	0.501	0.760	0.958	0.889	0.506	0.811
10.00	0.968	0.994	0.999	0.477	0.928	0.957	0.974	0.499	0.760	0.958	0.988	0.514	0.835

Table S-III. Finite-Sample False Coverage Probabilities of Robust QLR CI's for the MA Parameter π for Different Values of κ in the ARMA(1, 1) Model, $n = 500$

	$\pi_0 = 0.0$				$\pi_0 = 0.4$				$\pi_0 = 0.7$				
b	-2	-5	-10	$-\infty$	-2	-5	-10	$-\infty$	-2	-5	-10	$-\infty$	Avg
π_{H_0}	0.800	0.410	0.200	0.048	0.000	0.010	0.205	0.290	0.000	0.460	0.570	0.615	
LF	0.678	0.510	0.546	0.524	0.876	0.524	0.546	0.552	0.594	0.531	0.539	0.533	0.579
κ													
0.00	0.669	0.497	0.509	0.485	0.887	0.505	0.508	0.510	0.620	0.513	0.511	0.508	0.560
0.50	0.669	0.496	0.509	0.485	0.887	0.505	0.508	0.510	0.619	0.513	0.511	0.508	0.560
1.00	0.669	0.496	0.509	0.485	0.886	0.505	0.508	0.510	0.618	0.513	0.511	0.508	0.560
1.50	0.669	0.496	0.509	0.485	0.886	0.504	0.508	0.510	0.617	0.512	0.511	0.508	0.560
1.75	0.669	0.496	0.509	0.485	0.886	0.504	0.508	0.510	0.616	0.512	0.511	0.508	0.560
2.00	0.671	0.496	0.509	0.485	0.885	0.504	0.508	0.510	0.615	0.512	0.511	0.508	0.560
2.25	0.673	0.495	0.509	0.485	0.884	0.504	0.508	0.510	0.612	0.512	0.511	0.508	0.559
2.50	0.675	0.495	0.509	0.485	0.882	0.504	0.508	0.510	0.609	0.512	0.511	0.508	0.559
2.75	0.676	0.495	0.509	0.485	0.880	0.504	0.508	0.510	0.605	0.511	0.511	0.508	0.559
3.00	0.677	0.494	0.509	0.485	0.878	0.504	0.508	0.510	0.601	0.511	0.511	0.508	0.558
4.00	0.678	0.499	0.509	0.485	0.876	0.510	0.508	0.509	0.595	0.516	0.511	0.508	0.559
5.00	0.678	0.505	0.510	0.485	0.876	0.519	0.509	0.508	0.594	0.524	0.512	0.507	0.561
6.00	0.678	0.509	0.513	0.485	0.876	0.523	0.511	0.507	0.594	0.530	0.513	0.506	0.562
8.00	0.678	0.510	0.523	0.485	0.876	0.524	0.522	0.507	0.594	0.531	0.520	0.506	0.565
10.00	0.678	0.510	0.541	0.485	0.876	0.524	0.540	0.507	0.594	0.531	0.534	0.506	0.569

Table S-IV. Finite-Sample False Coverage Probabilities of 95% Least Favorable and Robust (with $\kappa = 1.5$) $|t|$ and QLR CI's for the AR parameter ρ in the ARMA(1, 1) Model, $n = 500$

	$\rho_0 = 0.0$				$\rho_0 = 0.4$				$\rho_0 = 0.8$				
b	2	5	10	∞	2	5	10	∞	2	5	10	∞	Avg
ρ_{H_0}	0.800	0.400	0.200	0.110	0.000	0.000	0.200	0.287	0.200	0.625	0.700	0.730	
$\|t\|$													
LF	0.97	0.99	1.00	1.00	0.94	0.97	1.00	1.00	0.69	1.00	1.00	1.00	0.96
Rob	0.93	0.77	0.54	0.56	0.93	0.65	0.49	0.50	0.58	0.57	0.45	0.47	0.62
QLR													
LF	0.66	0.52	0.53	0.53	0.88	0.52	0.54	0.54	0.48	0.49	0.51	0.52	0.56
Rob	0.65	0.50	0.50	0.49	0.89	0.50	0.50	0.50	0.51	0.48	0.49	0.49	0.54

Table S-V. Finite-Sample False Coverage Probabilities of Robust $|t|$ CI's for the AR Parameter ρ for Different Values of κ in the ARMA(1, 1) Model, $n = 500$

	$\rho_0 = 0.0$				$\rho_0 = 0.4$				$\rho_0 = 0.8$				
b	2	5	10	∞	2	5	10	∞	2	5	10	∞	Avg
ρ_{H_0}	0.800	0.725	0.212	0.117	0.000	0.000	0.200	0.287	0.075	0.595	0.705	0.735	
LF	0.967	0.990	1.000	1.000	0.942	0.973	0.999	1.000	0.588	0.995	1.000	1.000	0.955
κ													
0.00	0.925	0.400	0.495	0.504	0.932	0.656	0.492	0.497	0.501	0.445	0.482	0.517	0.573
0.50	0.925	0.399	0.495	0.504	0.932	0.656	0.492	0.497	0.501	0.445	0.482	0.517	0.572
1.00	0.925	0.399	0.495	0.504	0.932	0.655	0.492	0.497	0.501	0.445	0.482	0.517	0.572
1.50	0.930	0.416	0.495	0.504	0.930	0.655	0.492	0.497	0.500	0.457	0.484	0.517	0.575
1.75	0.941	0.454	0.496	0.504	0.926	0.655	0.493	0.496	0.498	0.476	0.487	0.517	0.581
2.00	0.948	0.496	0.497	0.503	0.929	0.670	0.494	0.495	0.506	0.503	0.491	0.516	0.590
2.25	0.953	0.543	0.500	0.502	0.932	0.688	0.497	0.494	0.520	0.536	0.495	0.516	0.600
2.50	0.958	0.591	0.504	0.502	0.936	0.708	0.502	0.493	0.537	0.566	0.501	0.515	0.612
2.75	0.961	0.635	0.510	0.501	0.938	0.731	0.506	0.492	0.552	0.600	0.507	0.515	0.623
3.00	0.963	0.688	0.517	0.500	0.940	0.756	0.511	0.491	0.564	0.635	0.513	0.515	0.635
4.00	0.967	0.851	0.556	0.498	0.941	0.859	0.542	0.490	0.585	0.794	0.551	0.515	0.681
5.00	0.967	0.951	0.615	0.497	0.942	0.935	0.590	0.487	0.588	0.922	0.612	0.515	0.720
6.00	0.967	0.982	0.709	0.496	0.942	0.965	0.664	0.486	0.588	0.986	0.696	0.516	0.750
8.00	0.967	0.990	0.923	0.497	0.942	0.973	0.851	0.485	0.588	0.995	0.908	0.519	0.803
10.00	0.967	0.990	0.997	0.501	0.942	0.973	0.978	0.484	0.588	0.995	0.997	0.529	0.829

15. Appendix E: Nonlinear Regression Example

In this section, we illustrate the verification of the assumptions in AC1 in a second example, a cross-section nonlinear regression model. We also show that the framework of Stock and Wright (2000) does not apply to this example.

15.1. Nonlinear Regression Model

This example is a cross-section nonlinear regression model estimated by LS. The model is

$$Y_i = \beta \cdot h\left(X_i, \pi\right) + Z_i'\zeta + U_i \text{ for } i = 1, ..., n, \quad (15.1)$$

where $h(X_i, \pi) \in R$ is known up to the finite-dimensional parameter $\pi \in R^{d_\pi}$. When the true value of β is 0, (15.1) becomes a linear model and π is not identified.

Table S-VI. Finite-Sample False Coverage Probabilities of Robust QLR CI's for the AR Parameter ρ for Different Values of κ in the ARMA(1, 1) Model, $n = 500$

	$\rho_0 = 0.0$				$\rho_0 = 0.4$				$\rho_0 = 0.8$				
b	2	5	10	∞	2	5	10	∞	2	5	10	∞	Avg
ρ_{H_0}	0.800	0.400	0.200	0.110	0.000	0.000	0.200	0.287	0.200	0.625	0.700	0.730	
LF	0.662	0.517	0.533	0.535	0.883	0.520	0.538	0.537	0.477	0.489	0.511	0.518	0.560
κ													
0.00	0.654	0.504	0.497	0.494	0.896	0.504	0.501	0.501	0.513	0.480	0.487	0.489	0.543
0.50	0.654	0.504	0.497	0.494	0.896	0.503	0.501	0.501	0.512	0.480	0.487	0.489	0.543
1.00	0.654	0.504	0.497	0.494	0.895	0.503	0.501	0.501	0.511	0.480	0.487	0.489	0.543
1.50	0.654	0.503	0.497	0.494	0.894	0.502	0.501	0.501	0.510	0.480	0.487	0.489	0.543
1.75	0.655	0.503	0.497	0.494	0.894	0.502	0.501	0.502	0.509	0.480	0.487	0.489	0.543
2.00	0.656	0.503	0.497	0.494	0.893	0.502	0.501	0.502	0.506	0.480	0.487	0.489	0.542
2.25	0.658	0.503	0.497	0.494	0.891	0.502	0.501	0.502	0.502	0.480	0.487	0.489	0.542
2.50	0.659	0.502	0.497	0.494	0.889	0.502	0.501	0.502	0.498	0.480	0.487	0.489	0.542
2.75	0.660	0.502	0.497	0.494	0.888	0.502	0.501	0.502	0.494	0.480	0.486	0.489	0.541
3.00	0.661	0.502	0.497	0.494	0.886	0.502	0.501	0.502	0.489	0.480	0.485	0.489	0.540
4.00	0.662	0.506	0.497	0.493	0.883	0.508	0.502	0.501	0.479	0.480	0.485	0.488	0.540
5.00	0.662	0.512	0.498	0.493	0.883	0.515	0.502	0.499	0.477	0.484	0.485	0.488	0.541
6.00	0.662	0.516	0.500	0.493	0.883	0.519	0.504	0.499	0.477	0.488	0.486	0.488	0.543
8.00	0.662	0.517	0.510	0.492	0.883	0.520	0.513	0.499	0.477	0.489	0.493	0.488	0.545
10.00	0.662	0.517	0.528	0.492	0.883	0.520	0.531	0.498	0.477	0.489	0.505	0.488	0.549

Suppose the support of X_i for all $\gamma \in \Gamma$ is contained in a set $\mathcal{X}$. We assume here that $h(x, \pi)$ is twice continuously differentiable wrt π, $\forall \pi \in \Pi$, $\forall x \in \mathcal{X}$, although the general theory of AC1 allows for non-smooth functions. Let $h_\pi(x, \pi) \in R^{d_\pi}$ and $h_{\pi\pi}(x, \pi) \in R^{d_\pi \times d_\pi}$ denote the first-order and second-order partial derivatives of $h(x, \pi)$ wrt π.

The LS sample criterion function is

$$Q_n(\theta) = n^{-1} \sum_{i=1}^{n} U_i^2(\theta)/2, \text{ where } U_i(\theta) = Y_i - \beta h(X_i, \pi) - Z_i' \zeta. \tag{15.2}$$

When $\beta = 0$, the residual $U_i(\theta)$ and the criterion function $Q_n(\theta)$ do not depend on π. Hence, Assumption A holds for this example.

15.2. Parameter Space

In this example, the random variables $\{(X_i, Z_i, U_i) : i = 1, ..., n\}$ are i.i.d. with distribution $\phi \in \Phi^*$, where Φ^* is a compact metric space with some metric that induces weak convergence. The parameter of interest is $\theta = (\beta, \zeta, \pi)$ and the nuisance parameter is ϕ, which is infinite dimensional. The true parameter space for θ is

$$\Theta^* = \mathcal{B}^* \times \mathcal{Z}^* \times \Pi^*, \text{ where } \mathcal{B}^* = [-b_1^*, b_2^*] \subset R \tag{15.3}$$

with $b_1^* \geq 0$, $b_2^* \geq 0$, b_1^* and b_2^* are not both equal to 0, $\mathcal{Z}^*$ $(\subset R^{d_\zeta})$ is compact, and Π^* $(\subset R^{d_\pi})$ is compact. For any $\theta^* \in \Theta^*$, the true parameter space for ϕ is

$$\begin{aligned}
&\Phi^*(\theta^*) = \{\phi \in \Phi^* : E_\phi(U_i|X_i, Z_i) = 0 \text{ a.s.}, \; E_\phi(U_i^2|X_i, Z_i) = \sigma^2(X_i, Z_i) > 0 \text{ a.s.}, \\
&E_\phi\left(\sup_{\pi\in\Pi} ||h\left(X_i, \pi\right)||^{4+\varepsilon} + \sup_{\pi\in\Pi} ||h_\pi\left(X_i, \pi\right)||^{4+\varepsilon} + \sup_{\pi\in\Pi} ||h_{\pi\pi}\left(X_i, \pi\right)||^{2+\varepsilon}\right) \leq C, \\
&||h_{\pi\pi}(X_i, \pi_1) - h_{\pi\pi}(X_i, \pi_2)|| \leq M(X_i)||\pi_1 - \pi_2|| \; \forall \pi_{1,}\pi_2 \in \Pi \text{ for some function} \\
&M(X_i), \; E_\phi M(X_i)^{2+\varepsilon} \leq C, \; E_\phi|U_i|^{4+\varepsilon} \leq C, \; E_\phi \left\|Z_i\right\|^{4+\varepsilon} \leq C, \\
&P_\phi(a'(h(X_i, \pi_1), h\left(X_i, \pi_2\right), Z_i) = 0) < 1, \; \forall \pi_1, \pi_2 \in \Pi \text{ with } \pi_1 \neq \pi_2, \; \forall a \in R^{d_\zeta+2} \\
&\text{with } a \neq 0, \; \lambda_{\min}(E_\phi(h(X_i, \pi), Z_i')'(h(X_i, \pi), Z_i')) \geq \varepsilon \; \forall \pi \in \Pi, \text{ and} \\
&\lambda_{\min}(E_\phi d_i\left(\pi\right) d_i(\pi)') \geq \varepsilon \; \forall \pi \in \Pi\}
\end{aligned} \tag{15.4}$$

for some constants $C < \infty$ and $\varepsilon > 0$, and by definition $d_i(\pi) = (h\left(X_i, \pi\right), Z_i, h_\pi\left(X_i, \pi\right))'$. The moment conditions are needed to ensure the uniform convergence of various sample averages. The other conditions are for the identification of β and ζ and the identification of π when $\beta \neq 0$.

Given the definitions above, the true parameter space Γ is of the form in (3.4). Thus, Assumption B2(i) holds immediately. Assumption B2(ii) follows from the form of $\mathcal{B}^*$ given in (15.3). Assumption B2(iii) follows from the form of $\mathcal{B}^*$ and the fact that Θ^* is a product space and $\Phi^*(\theta^*)$ does not depend on β^*. Hence, the true parameter space Γ satisfies Assumption B2.

The LS estimator of θ minimizes $Q_n(\theta)$ over $\theta \in \Theta$. The optimization parameter space Θ takes the form

$$\Theta = \mathcal{B} \times \mathcal{Z} \times \Pi, \text{ where } \mathcal{B} = [-b_1, b_2] \subset R \tag{15.5}$$

with $b_1 > b_1^*$, $b_2 > b_2^*$, $\mathcal{Z}$ $(\subset R^{d_\zeta})$ is compact, Π $(\subset R^{d_\pi})$ is compact, $\mathcal{Z}^* \in int(\mathcal{Z})$, and $\mathcal{B}^* \in int(\mathcal{B})$. Given these conditions, Assumptions B1(i) and B1(iii) follow immediately. Assumption B1(ii) holds by taking $\delta < \min\{b_1^*, b_2^*\}$ and $\mathcal{Z}^0 = int(\mathcal{Z})$.

15.3. Criterion Function Limit Assumption

In this example, the function $Q(\theta;\gamma_0)$ in Assumption B3(i) is

$$Q(\theta;\gamma_0) = E_{\phi_0} U_i^2/2 + E_{\phi_0}(\beta_0 h(X_i,\pi_0) + Z_i'\zeta_0 - \beta h(X_i,\pi) - Z_i'\zeta)^2/2, \tag{15.6}$$

where $\gamma_0 = (\beta_0, \zeta_0, \pi_0, \phi_0)$ and E_{ϕ_0} denotes expectation when the distribution of (X_i, Z_i, U_i) is ϕ_0. The uniform convergence in Assumption B3(i) holds by the following uniform WLLN given the moment and smoothness conditions in $\Phi^*(\theta^*)$ in (15.3).

Lemma 15.1. *Suppose* (i) $\{W_i : i \geq 1\}$ *is an i.i.d. sequence under* F_{γ^*} *for all* $\gamma^* \in \Gamma$, (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) Θ *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* Θ.

Comments. 1. The centering term in Lemma 15.1 is $E_{\gamma_0} s(W_i,\theta)$, rather than $E_{\gamma_n} s(W_i, \theta)$.

2. The proof of Lemma 15.1 is given in AC2.

Next, we verify Assumption B3* given in Appendix A, which is a set of sufficient conditions for Assumptions B3(ii) and B3(iii). Assumption B3*(i) holds with $Q(\theta;\gamma_0)$ defined in (15.6) by the continuity of $h(x,\pi)$ in π, the moment conditions in (15.4), and the DCT. Assumptions B3*(iv) and B3*(v) hold because $\Psi(\pi) = \mathcal{B} \times \mathcal{Z}$ is compact and does not depend on π. To verify Assumption B3*(ii), we need that when $\beta_0 = 0$,

$$Q(\psi,\pi;\gamma_0) - Q(\psi_0,\pi;\gamma_0) = E_{\phi_0}(\beta h(X_i,\pi) + Z_i'(\zeta_0 - \zeta))^2/2 > 0 \tag{15.7}$$

$\forall \psi \neq \psi_0$, $\forall \pi \in \Pi$. The inequality in (15.7) holds unless

$$P_{\phi_0}(\beta h(X_i,\pi) + Z_i'(\zeta_0 - \zeta) = 0) = 1 \tag{15.8}$$

for some $\psi \neq \psi_0$ and $\pi \in \Pi$. But $P_{\phi_0}(a'(h(X_i,\pi),Z_i) = 0) < 1$ for all $a \in R^{d_\zeta+1}$ and $a \neq 0$ by (15.4). Hence, (15.8) cannot hold for any $(\beta,\zeta) \neq (0,\zeta_0)$. This completes the verification of Assumption B3*(ii).

To verify Assumption B3*(iii), we need that when $\beta_0 \neq 0$,

$$Q(\theta;\gamma_0) - Q(\theta_0;\gamma_0) = E_{\phi_0}(\beta h(X_i,\pi) - \beta_0 h(X_i,\pi_0) + Z_i'(\zeta_0 - \zeta))^2/2 > 0 \qquad (15.9)$$

$\forall \theta \neq \theta_0$. The inequality in (15.9) holds unless

$$P_{\phi_0}(\beta_0 h(X_i,\pi_0) - \beta h(X_i,\pi) + Z_i'(\zeta_0 - \zeta) = 0) = 1 \qquad (15.10)$$

for some $\theta \neq \theta_0$. Because $P_{\phi_0}(a'(h(X_i,\pi), h(X_i,\pi_0), Z_i) = 0) < 1$ for all $\pi \neq \pi_0$ and $a \neq 0$ by (15.4), the condition $\beta_0 \neq 0$ implies that (15.10) cannot hold for any θ such that $\pi \neq \pi_0$. When $\pi = \pi_0$, (15.10) becomes

$$P_{\phi_0}((\beta_0 - \beta)h(X_i,\pi_0) + Z_i'(\zeta_0 - \zeta) = 0) = 1. \qquad (15.11)$$

Because $P_{\phi_0}(a'(h(X_i,\pi),Z_i) = 0) < 1$ for all $a \in R^{d_\zeta+1}$ and $a \neq 0$ by (15.4), equation (15.11) cannot hold for $(\beta,\zeta) \neq (\beta_0,\zeta_0)$. This completes the verification of Assumption B3*.

15.4. Close to $\boldsymbol{\beta} = \mathbf{0}$ Assumptions

15.4.1. Assumptions C1 and D1

The sample criterion function $Q_n(\theta)$ is a smooth sample average:

$$Q_n(\theta) = n^{-1}\sum_{i=1}^{n} \rho(W_i,\theta), \text{ where } \rho(W_i,\theta) = U_i^2(\theta)/2 \text{ and } W_i = (Y_i, X_i, Z_i')'. \qquad (15.12)$$

In consequence, we verify Assumptions C1 and D1 by verifying Assumption Q1 of Appendix A. The latter is sufficient for the Assumptions C1 and D1 by Lemma 11.5 of Appendix A (given Assumptions B1 and B2).

The first- and second-order partial derivatives of $\rho(W_i,\theta)$ wrt to ψ are

$$\rho_\psi(W_i,\theta) = -U_i(\theta)d_{\psi,i}(\pi) \text{ and } \rho_{\psi\psi}(W_i,\theta) = d_{\psi,i}(\pi)d_{\psi,i}(\pi)', \text{ where}$$
$$d_{\psi,i}(\pi) = (h(X_i,\pi), Z_i')'. \qquad (15.13)$$

Thus, by Lemma 11.5, we verify that Assumption C1 holds with

$$D_\psi Q_n(\theta) = -n^{-1}\sum_{i=1}^{n} U_i(\theta)d_{\psi,i}(\pi) \text{ and } D_{\psi\psi}Q_n(\theta) = n^{-1}\sum_{i=1}^{n} d_{\psi,i}(\pi)d_{\psi,i}(\pi)'. \quad (15.14)$$

The first- and second-order partial derivatives of $\rho(W_i, \theta)$ wrt to θ are

$$\begin{aligned}
\rho_\theta(W_i,\theta) &= -U_i(\theta)B(\beta)d_i(\pi) \text{ and} \\
\rho_{\theta\theta}(W_i,\theta) &= -U_i(\theta)D_i(\theta) + B(\beta)d_i(\pi)d_i(\pi)'B(\beta), \text{ where} \\
d_i(\pi) &= (h(X_i,\pi), Z_i', h_\pi(X_i,\pi)')', \\
D_i(\theta) &= \begin{bmatrix} 0 & 0_{1\times d_\zeta} & h_\pi(X_i,\pi)' \\ 0_{d_\zeta\times 1} & 0_{d_\zeta\times d_\zeta} & 0_{d_\zeta\times d_\pi} \\ h_\pi(X_i,\pi) & 0_{d_\pi\times d_\zeta} & h_{\pi\pi}(X_i,\pi)\beta \end{bmatrix},
\end{aligned} \quad (15.15)$$

and $B(\beta)$ depends on β, not $||\beta||$, because β is a scalar. Hence, by Lemma 11.5, we verify that Assumption D1 holds with

$$\begin{aligned}
DQ_n(\theta) &= -n^{-1}\sum_{i=1}^{n} U_i(\theta)B(\beta)d_i(\pi) \text{ and} \\
D^2Q_n(\theta) &= n^{-1}\sum_{i=1}^{n}\left(B(\beta)d_i(\pi)d_i(\pi)'B(\beta) - U_i(\theta)D_i(\theta)\right)
\end{aligned} \quad (15.16)$$

by Lemma 11.5 in AC1-SM.[47]

Now, verify Assumption Q1. Assumptions Q1(i) and Q1(ii) hold immediately. Assumption Q1(iii) holds because $\rho_{\psi\psi}(W_i,\theta)$ does not depend on ψ. Now we verify Assumption Q1(iv). By (15.13), verification of Assumption Q1(iv) is equivalent to showing the stochastic equicontinuity (SE) of $n^{-1}\sum_{i=1}^n U_i(\theta)h_\pi(X_i,\pi)/\beta_n$, $n^{-1}\sum_{i=1}^n U_i(\theta)h_{\pi\pi}(X_i,\pi) \times \beta/\beta_n^2$, and $n^{-1}\sum_{i=1}^n B(\beta/\beta_n)d_i(\pi)d_i(\pi)'B(\beta/\beta_n)$ over $\theta \in \Theta_n(\delta_n)$. We now show the SE of these three terms under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$.

[47] This example illustrates why defining $B(\beta)$ using β, not $||\beta||$, is preferred in the scalar β case. If $B(\beta)$ is defined with $||\beta||$ in place of β, then $d_i(\pi)$ needs to be replaced by $d_i(\beta,\pi) = (h(X_i,\pi), Z_i', sgn(\beta)h_\pi(X_i,\pi)')'$. The appearance of $sgn(\beta)$ complicates matters because it introduces a dependence of $d_i(\beta,\pi)$ on β, which otherwise does not appear, and it is a discontinuous function of β.

The first term is

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n} U_i(\theta)h_\pi(X_i,\pi)/\beta_n \qquad (15.17)\\
&= \left(n^{-1/2}\sum_{i=1}^{n} U_i h_\pi(X_i,\pi)\right)/\left(n^{1/2}\beta_n\right) + \left(n^{-1}\sum_{i=1}^{n} h(X_i,\pi_n)\, h_\pi(X_i,\pi)\right) - \\
&\left(n^{-1}\sum_{i=1}^{n} h(X_i,\pi)\, h_\pi(X_i,\pi)\right)\beta/\beta_n + n^{-1}\sum_{i=1}^{n} Z_i'(\zeta_n-\zeta)\, h_\pi(X_i,\pi)/\beta_n.
\end{aligned}
$$

Note that for $\theta \in \Theta_n(\delta_n)$, we have $|\beta/\beta_n| = 1 + o(1)$ and $(\zeta - \zeta_n)/\beta_n = o(1)$ because $||\psi - \psi_n|| \leq \delta_n|\beta_n|$ and $\delta_n \to 0$. Hence, under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, the SE of $n^{-1}\sum_{i=1}^n U_i(\theta)h_\pi(X_i,\pi)/\beta_n$ is implied by the SE of (i) $n^{-1/2}\sum_{i=1}^n U_i h_\pi(X_i,\pi)$ on $\pi \in \Pi$, (ii) $n^{-1}\sum_{i=1}^n h(X_i,\pi)h_\pi(X_i,\overline{\pi})$ on $(\pi,\overline{\pi}) \in \Pi\times\Pi$, and (iii) $n^{-1}\sum_{i=1}^n Z_i h_\pi(X_i,\pi)'$ on $\pi \in \Pi$. The SE of (i) holds by Theorems 1 and 2 of Andrews (1994) using the type II class with envelope function $B(W_i) = U_i \sup_{\pi\in\Pi}||h_{\pi\pi}(X_i,\pi)||$, the moment conditions in (15.4), and the compactness of Π. The SE of (ii) and (iii) follows from Lemma 15.1.

Similarly, we can show the SE of $n^{-1}\sum_{i=1}^n U_i(\theta)h_{\pi\pi}(X_i,\pi)\beta/\beta_n^2$ by replacing $h_\pi(X_i,\pi)$ with $h_{\pi\pi}(X_i,\pi)$ in the foregoing argument and using $|\beta/\beta_n| = 1 + o(1)$. To verify the SE of $n^{-1/2}\sum_{i=1}^n U_i h_{\pi\pi}(X_i,\pi)$ on $\pi \in \Pi$ (element by element), we use the type II class in Andrews (1994) with envelope function $B(W_i) = U_i M(X_i)$ and the Lipschitz condition in (15.4). The SE of $n^{-1}\sum_{i=1}^n h(X_i,\pi)h_{\pi\pi}(X_i,\overline{\pi})$ and $n^{-1}\sum_{i=1}^n Z_i h_{\pi\pi}(X_i,\pi)'$ follows from Lemma 15.1.

Finally, the SE of $n^{-1}\sum_{i=1}^n B(\beta/\beta_n)d_i(\pi)d_i(\pi)'B(\beta/\beta_n)$ follows from Lemma 15.1 using $|\beta/\beta_n| = 1 + o(1)$. This completes the verification of Assumption Q1.

15.4.2. Assumption C2

Assumption C2(i) holds in this example with

$$m(W_i,\theta) = -U_i(\theta)d_{\psi,i}(\pi). \qquad (15.18)$$

Assumption C2(ii) holds because $E_{\gamma^*}m(W_i,\theta^*) = -E_{\gamma^*}U_i(h(X_i,\pi^*),Z_i')' = 0\ \forall\gamma^* \in \Gamma$. Assumption C2(iii) holds because $E_{\gamma^*}m(W_i,\psi^*,\pi) = -E_{\gamma^*}(U_i + \beta^* h(X_i,\pi^*) - \beta^* h(X_i,\pi)) \times (h(X_i,\pi),Z_i')' = 0\ \forall\pi \in \Pi$ when $\beta^* = 0$.

15.4.3. Assumption C3

To verify Assumption C3, we have

$$\begin{aligned} U_i(\psi_{0,n}, \pi) &= Y_i - Z_i'\zeta_n = U_i + \beta_n h(X_i, \pi_n) \text{ and} \qquad (15.19) \\ G_n(\pi) &= -n^{-1/2} \sum_{i=1}^{n} (U_i d_{\psi,i}(\pi) + \beta_n [h(X_i, \pi_n) d_{\psi,i}(\pi) - E_{\phi_n} h(X_i, \pi_n) d_{\psi,i}(\pi)]). \end{aligned}$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $G_n(\pi) \Rightarrow G(\pi; \gamma_0)$, where $G(\pi; \gamma_0)$ is a Gaussian process with bounded continuous sample paths and covariance kernel $\Omega(\pi_1, \pi_2; \gamma_0) = E_{\phi_0} U_i^2 d_{\psi,i}(\pi_1) d_{\psi,i}(\pi_2)'$. This weak convergence follows from Andrews (1994, p. 2251) because (i) Π is compact, (ii) the finite-dimensional convergence holds by the CLT for a triangular array of row-wise i.i.d. random variables, where the Lindeberg condition holds by the $L^{2+\delta}$-boundedness of its summands, and $\beta_n \to 0$, and (iii) the stochastic equicontinuity (SE) holds by applying the type II class (Lipschitz functions) using the differentiability of $h(x, \pi)$ in π.

15.4.4. Assumption C4

Assumption C4(i) holds in this example with

$$H(\pi; \gamma_0) = E_{\phi_0} d_{\psi,i}(\pi) d_{\psi,i}(\pi)' \qquad (15.20)$$

by applying a uniform LLN for drifting true distributions, specifically, Lemma 15.1, to $n^{-1} \sum_{i=1}^{n} d_{\psi,i}(\pi) d_{\psi,i}(\pi)$. The continuity of $H(\pi; \gamma_0)$ is implied by the continuity of $h(X_i, \pi)$ in π, $E_{\phi_0} \sup_{\pi \in \Pi} ||d_{\psi,i}(\pi) d_{\psi,i}(\pi)'|| < \infty$, and the DCT. Assumption C4(ii) follows immediately from the conditions in (15.4).

15.4.5. Assumption C5

To verify Assumption C5(i), we have

$$\begin{aligned} K_n(\theta; \gamma^*) &= \frac{\partial}{\partial \beta^*} E_{\phi^*} m(W_i, \theta) = -\frac{\partial}{\partial \beta^*} E_{\phi^*} (Y_i - \beta h(X_i, \pi) - Z_i'\zeta) d_{\psi,i}(\pi) \\ &= -\frac{\partial}{\partial \beta^*} E_{\phi^*} (U_i + \beta^* h(X_i, \pi^*) - \beta h(X_i, \pi) - Z_i'(\zeta - \zeta^*)) d_{\psi,i}(\pi) \\ &= -E_{\phi^*} h(X_i, \pi^*) d_{\psi,i}(\pi). \qquad (15.21) \end{aligned}$$

Next, we verify that Assumptions C5(ii) and C5(iii) hold with

$$K(\pi;\gamma_0) = K(\psi_0,\pi;\gamma_0) = -E_{\phi_0} h(X_i,\pi_0) d_{\psi,i}(\pi). \tag{15.22}$$

They hold provided $E_{\phi_n} h(X_i,\pi_1)d_{\psi,i}(\pi_2) \to E_{\phi_0} h(X_i,\pi_1)d_{\psi,i}(\pi_2)$ uniformly over $(\pi_1,\pi_2) \in \Pi\times\Pi$ as $\phi_n \to \phi_0$ and $E_{\phi_0} h(X_i,\pi_1)d_{\psi,i}(\pi_2)$ is continuous in (π_1,π_2). The continuity holds by the continuity of $h(X_i,\pi_1)d_{\psi,i}(\pi_2)$ in (π_1,π_2), $E_{\phi_0}\sup_{(\pi_1,\pi_2)\in\Pi\times\Pi}||h(X_i,\pi_1) d_{\psi,i}(\pi_2)|| < \infty$, and the DCT. By Lemma 8.2 in AC2, the uniform convergence follows from the pointwise convergence and the equicontinuity of $E_{\phi^*} h(X_i,\pi_1)d_{\psi,i}(\pi_2)$ in (π_1,π_2) over $\phi^* \in \Phi^*(\theta^*)$. The pointwise convergence $E_{\phi_n} h(X_i,\pi_1)d_{\psi,i}(\pi_2) \to E_{\phi_0} h(X_i,\pi_1)d_{\psi,i}(\pi_2)$ holds by the convergence in distribution of ϕ_n to ϕ_0 (since $\phi_n \to \phi_0$ and the metric on Φ^* induces weak convergence) and the $L^{1+\delta}$ boundedness of $h(X_i,\pi_1)d_{\psi,i}(\pi_2)$ under $\phi \in \Phi^*$, i.e., $\sup_{\phi\in\Phi^*} E_\phi||h(X_i,\pi_1)d_{\psi,i}(\pi_2)||^{1+\delta} \leq C < \infty$ (e.g., see Theorem 2.20 and Example 2.21 of van der Vaart (1998)). Equicontinuity holds because $h(X_i,\pi_1)d_{\psi,i}(\pi_2)$ is partially differentiable in (π_1,π_2) and the partial derivatives are uniformly bounded, i.e., $E_{\phi^*}\sup_{(\pi_1,\pi_2)\in\Pi\times\Pi}(||h_\pi(X_i,\pi_1)'d_{\psi,i}(\pi_2)||+||h(X_i,\pi_1)(\partial d_{\psi,i}(\pi_2)/\partial\pi')||) \leq C$ for some $C<\infty$ for all $\phi^* \in \Phi^*(\theta^*)$.

15.4.6. Assumption C6

Next, we verify Assumption C6**. Assumption C6**(i) holds because β is a scalar. By the discussion following (15.19), $a'(G_1(\pi_1), G_1(\pi_2), G_2)$ has variance $E_{\phi_0} U_i^2 d_a^2(\pi_1,\pi_2)$, where $d_a(\pi_1,\pi_2) = a'(h(X_i,\pi_1), h(X_i,\pi_2), Z_i)$. By the conditions in (15.4), $P_{\phi_0}(d_a(\pi_1,\pi_2) = 0) < 1$ $\forall a \in R^{d_\varsigma+2}$ with $a \neq 0$, $\forall \pi_1 \neq \pi_2$, $\forall \phi_0 \in \Phi^*(\theta_0)$, and $E_{\phi_0}(U_i^2|X_i,Z_i) > 0$ a.s. Hence, $E_{\phi_0} U_i^2 d_a^2(\pi_1,\pi_2) > 0$ $\forall a \neq 0$ and Assumption C6**(ii) holds.

15.4.7. Assumption C7

We verify Assumption C7 as follows. Given the form of $H(\pi;\gamma_0)$ and $K(\pi;\gamma_0)$ in (15.20) and (15.22), respectively, we have

$$\begin{aligned} & K(\pi;\gamma_0)'H^{-1}(\pi;\gamma_0)K(\pi;\gamma_0) \\ = {} & [E_{\phi_0} h(X_i,\pi_0)d_{\psi,i}(\pi)]'[E_{\phi_0} d_{\psi,i}(\pi)d_{\psi,i}(\pi)']^{-1}[E_{\phi_0} d_{\psi,i}(\pi)h(X_i,\pi_0)] \leq E_{\phi_0} h^2(X_i,\pi_0), \end{aligned} \tag{15.23}$$

where the inequality holds by the matrix Cauchy-Schwarz inequality in Tripathi (1999). The "$\leq$" holds as an equality if and only if $h(X_i,\pi_0)a_1 + d_{\psi,i}(\pi)'a_2 = 0$ with probability

1 for some $a_1 \in R$, $a_2 \in R^{d_\zeta+1}$, and $(a_1, a_2') \neq 0$. The "$\leq$" holds as an equality uniquely at $\pi = \pi_0$ because for any $\pi \neq \pi_0$, $P_{\phi_0}(c'(h(X_i, \pi_0), h(X_i, \pi), Z_i) = 0) < 1$ for any $c \neq 0$ by (15.4). This completes the verification of Assumption C7.

15.4.8. Assumption C8

Lastly, we verify Assumption C8. To verify Assumption C8, we have

$$(\partial/\partial\psi')E_{\gamma_n} D_\psi Q_n(\psi, \pi_n)|_{\psi=\psi_n} = E_{\phi_n} d_{\psi,i}(\pi_n) d_{\psi,i}(\pi_n)' \tag{15.24}$$

by the form of $D_\psi Q_n(\theta_n)$ given in (15.14) of AC1. Assumption C8 holds provided $E_{\phi_n} d_{\psi,i}(\pi) d_{\psi,i}(\pi)'$ converges to $E_{\phi_0} d_{\psi,i}(\pi) d_{\psi,i}(\pi)'$ uniformly over $\pi \in \Pi$ and $E_{\phi_0} d_{\psi,i}(\pi)$ $d_{\psi,i}(\pi)'$ is continuous in π. This holds by the same argument as in the verification of Assumption C5 above by replacing $h(X_i, \pi_1) d_{\psi,i}(\pi_2)$ with $d_{\psi,i}(\pi) d_{\psi,i}(\pi)'$. The smoothness and moment conditions are satisfied by the conditions in (15.4) of AC1.

15.5. Distant from $\beta = 0$ Assumptions

15.5.1. Assumption D2

To verify Assumption D2 with $D^2 Q_n(\theta)$ given in (15.16), we have

$$J_n = n^{-1} \sum_{i=1}^{n} d_i(\pi_n) d_i(\pi_n)' - \tag{15.25}$$
$$(n^{1/2}\beta_n)^{-1} \begin{bmatrix} 0 & 0_{1\times d_\zeta} & n^{-1/2}\sum_{i=1}^n U_i h_\pi(X_i, \pi_n)' \\ 0_{d_\zeta \times 1} & 0_{d_\zeta \times d_\zeta} & 0_{d_\zeta \times d_\pi} \\ n^{-1/2}\sum_{i=1}^n U_i h_\pi(X_i, \pi_n) & 0_{d_\pi \times d_\zeta} & n^{-1/2}\sum_{i=1}^n U_i h_{\pi\pi}(X_i, \pi) \end{bmatrix}.$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $n^{-1}\sum_{i=1}^n d_i(\pi_n) d_i(\pi_n)' \rightarrow_p E_{\phi_0} d_i(\pi_0) d_i(\pi_0)'$ because $n^{-1}\sum_{i=1}^n d_i(\pi) d_i(\pi)' \rightarrow_p E_{\phi_0} d_i(\pi) d_i(\pi)'$ uniformly over $\pi \in \Pi$ by Lemma 15.1 in AC1-SM and the continuity of $E_{\phi_0} d_i(\pi) d_i(\pi)'$ in π. The second line of (15.25) is $o_p(1)$ because $n^{1/2}|\beta_n| \rightarrow \infty$, $n^{-1/2}\sum_{i=1}^n U_i h_\pi(X_i, \pi_n)' = O_p(1)$, and $n^{-1/2}\sum_{i=1}^n U_i h_{\pi\pi}(X_i, \pi_n) = O_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$. The latter two terms are $O_p(1)$ by the CLT for a triangular array of row-wise i.i.d. random variables under the moment conditions in (15.4). Hence, Assumption D2 holds with the matrix

$$J(\gamma_0) = E_{\phi_0} d_i(\pi_0) d_i(\pi_0)', \tag{15.26}$$

which is nonsingular by the conditions in (15.4).

15.5.2. Assumption D3

To verify Assumption D3 in this example, we have

$$n^{1/2}B^{-1}(\beta_n)DQ_n(\theta_n) = -n^{-1/2}\sum_{i=1}^{n} U_i d_i(\pi_n) \rightarrow_d N(0_{d_\theta}, V(\gamma_0)), \text{ where}$$
$$V(\gamma_0) = E_{\phi_0}U_i^2 d_i(\pi_0)d_i(\pi_0)'. \tag{15.27}$$

The convergence in distribution holds by the CLT for a triangular array of row-wise i.i.d. random variables. Assumption D3(ii) holds because $E_{\phi_0}d_i(\pi_0)d_i(\pi_0)'$ is non-singular and $E_{\phi_0}(U_i^2|X_i, Z_i) > 0$ a.s. by (15.4).

15.6. Key Quantities

In this example, the components of the stochastic processes $\xi(\pi;\gamma_0, b)$ and $\tau(\pi;\gamma_0, b)$, the function $\eta(\pi;\gamma_0,\omega_0)$, and the matrices $J(\gamma_0)$ and $V(\gamma_0)$ that appear in the asymptotic results in Section 5 of AC1 are

$$\begin{aligned}
H(\pi;\gamma_0) &= E_{\phi_0}d_{\psi,i}(\pi)d_{\psi,i}(\pi)', \\
K(\pi;\gamma_0) &= -E_{\phi_0}h(X_i,\pi_0)d_{\psi,i}(\pi), \\
\Omega(\pi_1,\pi_2;\gamma_0) &= E_{\phi_0}U_i^2 d_{\psi,i}(\pi_1)d_{\psi,i}(\pi_2)', \\
J(\gamma_0) &= E_{\phi_0}d_i(\pi_0)d_i(\pi_0)', \\
V(\gamma_0) &= E_{\phi_0}U_i^2 d_i(\pi_0)d_i(\pi_0)', \text{ where} \\
d_{\psi,i}(\pi) &= (h(X_i,\pi), Z_i')', \;\; d_i(\pi) = \left(h\left(X_i,\pi\right), Z_i, h_\pi\left(X_i,\pi\right)\right)',
\end{aligned} \tag{15.28}$$

and $G(\pi;\gamma_0)$ is a mean zero Gaussian process with covariance kernel $\Omega(\pi_1,\pi_2;\gamma_0)$.

15.7. Variance Matrix Estimators

In this example, we estimate $J(\gamma_0)$ and $V(\gamma_0)$ by $\widehat{J}_n = \widehat{J}_n(\widehat{\theta}_n)$ and $\widehat{V}_n = \widehat{V}_n(\widehat{\theta}_n)$, respectively, where

$$\begin{aligned}
\widehat{J}_n(\theta) &= n^{-1}\sum_{i=1}^n d_i(\pi)\, d_i(\pi)' \text{ and} \\
\widehat{V}_n(\theta) &= n^{-1}\sum_{i=1}^n U_i^2(\theta) d_i(\pi)\, d_i(\pi)' = n^{-1}\sum_{i=1}^n U_i^2 d_i(\pi)\, d_i(\pi)' \\
&\quad +2n^{-1}\sum_{i=1}^n U_i\left(\beta_n h(X_i,\pi_n) - \beta h(X_i,\pi) + (\zeta_n-\zeta)'Z_i\right) d_i(\pi)\, d_i(\pi)' \\
&\quad +n^{-1}\sum_{i=1}^n \left(\beta_n h(X_i,\pi_n) - \beta h(X_i,\pi) + (\zeta_n-\zeta)'Z_i\right)^2 d_i(\pi)\, d_i(\pi)'. \qquad (15.29)
\end{aligned}$$

These variance matrix estimators are used to construct t and Wald statistics and also to construct the identification-category-selection statistic A_n in (7.4) of AC1.

Assumption V1(i) (scalar β) holds with

$$\begin{aligned}
&J(\theta;\gamma_0) = E_{\phi_0} d_i(\pi)\, d_i(\pi)' \text{ and } V(\theta;\gamma_0) = E_{\phi_0} U_i^2 d_i(\pi)\, d_i(\pi)' \\
&+E_{\phi_0}\left(\beta_0 h(X_i,\pi_0) - \beta h(X_i,\pi) + (\zeta_0-\zeta)'Z_i\right)^2 d_i(\pi) d_i(\pi)', \qquad (15.30)
\end{aligned}$$

by Lemma 15.1 using the conditions in (15.4). Assumption V1(ii) holds by the continuity of $h(x,\pi)$ and $h_\pi(x,\pi)$ in π and the moment conditions in (15.4).

The quantity $\Sigma(\pi;\gamma_0)$ in (6.4) takes the form

$$\Sigma(\pi;\gamma_0) = \left(E_{\phi_0} d_i(\pi)\, d_i(\pi)'\right)^{-1} E_{\phi_0} U_i^2 d_i(\pi)\, d_i(\pi)' \left(E_{\phi_0} d_i(\pi)\, d_i(\pi)'\right)^{-1}. \qquad (15.31)$$

Given this, Assumption V1(iii) holds by the nonsingularity conditions in (15.4).

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.

15.8. Failure of Assumption C of Stock and Wright (2000)

In this section, we show that the main assumption of Stock and Wright (2000) (SW), Assumption C, fails for the GMM estimator based on the nonlinear LS first-order conditions in the nonlinear regression model of (15.1). The implication is that the range of applicability of this paper and that of SW are different, as discussed in the Introduction of AC1. In particular, in SW the estimator criterion function cannot be indexed by parameters that determine the strength of identification, whereas in this paper it does.

Consider the model in (15.1) and, for simplicity, suppose no $Z_i'\zeta$ summand appears:

$$Y_i = \beta \cdot h\left(X_i, \pi\right) + U_i. \tag{15.32}$$

The parameters (β, π) in our notation correspond to (β, α) in SW. That is, β is strongly identified and π $(= \alpha)$ is potentially weakly identified. We switch notation from π to α and back whenever it is convenient. To generate weak identification of π in (15.32), suppose the true parameters are $\gamma_n = (\beta_n, \pi_0, \phi_0)$, where $\beta_n = Cn^{-1/2}$ for $n \geq 1$ for some $0 < C < \infty$. The nonlinear LS first-order conditions yield the following moment conditions: When $(\beta, \pi) = (\beta_n, \pi_0)$,

$$E_{\gamma_n}(Y_i - \beta h\left(X_i, \pi\right)) \begin{pmatrix} h\left(X_i, \pi\right) \\ h_\pi\left(X_i, \pi\right) \end{pmatrix} = 0_2. \tag{15.33}$$

To apply SW's results, one takes their $Z_t = 1\ \forall t$ and their moment function $\phi_t(\theta)$ to equal the function in (15.33), where their t, T, θ correspond to our $i, n, (\beta, \pi)$, respectively.

SW's population moments $\widetilde{m}_T(\alpha, \beta)$ equal the following:

$$\begin{aligned} \widetilde{m}_T(\alpha, \beta) &= E_{\gamma_n}(Y_i - \beta h\left(X_i, \pi\right)) \begin{pmatrix} h\left(X_i, \pi\right) \\ h_\pi\left(X_i, \pi\right) \end{pmatrix} \\ &= E_{\phi_0}(\beta_n h\left(X_i, \pi_0\right) - \beta h\left(X_i, \pi\right)) \begin{pmatrix} h\left(X_i, \pi\right) \\ h_\pi\left(X_i, \pi\right) \end{pmatrix}. \end{aligned} \tag{15.34}$$

Next, SW use an identity $\widetilde{m}_T(\alpha,\beta) = \widetilde{m}_T(\alpha_0,\beta_n) + \widetilde{m}_{1T}(\alpha,\beta) + \widetilde{m}_2(\beta)$, where

$$\begin{aligned}
\widetilde{m}_{1T}(\alpha,\beta) &= \widetilde{m}_T(\alpha,\beta) - \widetilde{m}_T(\alpha_0,\beta) \\
&= E_{\phi_0}(\beta_n h(X_i,\pi_0) - \beta h(X_i,\pi)) \begin{pmatrix} h(X_i,\pi) \\ h_\pi(X_i,\pi) \end{pmatrix} \\
&\quad - E_{\phi_0}(\beta_n h(X_i,\pi_0) - \beta h(X_i,\pi_0)) \begin{pmatrix} h(X_i,\pi_0) \\ h_\pi(X_i,\pi_0) \end{pmatrix} \\
&= A_{1n}(\pi) + A_2(\pi,\beta), \qquad (15.35)
\end{aligned}$$

where

$$\begin{aligned}
A_{1n}(\pi) &= n^{-1/2} C \cdot E_{\phi_0} h(X_i,\pi_0) \begin{pmatrix} h(X_i,\pi) - h(X_i,\pi_0) \\ h_\pi(X_i,\pi) - h_\pi(X_i,\pi_0) \end{pmatrix} \text{ and} \qquad (15.36) \\
A_2(\pi,\beta) &= \beta E_{\phi_0}\left[h(X_i,\pi_0) \begin{pmatrix} h(X_i,\pi_0) \\ h_\pi(X_i,\pi_0) \end{pmatrix} - h(X_i,\pi) \begin{pmatrix} h(X_i,\pi) \\ h_\pi(X_i,\pi) \end{pmatrix} \right].
\end{aligned}$$

The first component, $A_{1n}(\pi)$, of $\widetilde{m}_{1T}(\alpha,\beta)$ has the form required by Assumption C(i) of SW. It is $n^{-1/2}$ times a function, call it $s_n(\pi)$, that has a limit as $n \to \infty$ uniformly over π that is continuous and bounded and equals 0 when $\pi = \pi_0$. (In fact, in the present case, $s_n(\pi)$ does not depend on n so the limit holds trivially.)

However, the second component, $A_2(\pi,\beta)$, does not have the form specified in Assumption C(i). It does not depend on n and is not identically zero. In consequence, Assumption C(i) of SW fails in this example.

In words, SW state "The key idea in this paper, made precise in Assumption C below, is to treat $\widetilde{m}_2(\beta)$ as large for β outside β_0, but $\widetilde{m}_{1T}(\alpha,\beta)$ as small for all α and β," see p. 1060 of SW. As shown in (15.35)-(15.36), in this example, $\widetilde{m}_{1T}(\alpha,\beta)$ is not small for all α and β. The same feature arises in other examples in which a parameter that determines the strength of identification appears in the estimator criterion function.

References

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241-257.

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821-856.

——— (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. McFadden. Amsterdam: North-Holland.

——— (1999): "Estimation When a Parameter Is on a Boundary," *Econometrica*, 67, 1341-1383.

——— (2001): "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683-734.

Andrews, D. W. K. and X. Cheng (2008a): "Maximum Likelihood Estimation and Inference with Weak, Semi-strong, and Strong Identification," manuscript in preparation, Cowles Foundation, Yale University.

——— (2008b): "GMM Estimation and Inference with Weak, Semi-strong, and Strong Identification," manuscript in preparation, Cowles Foundation, Yale University.

Andrews, D. W. K. and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica,* 62, 1383–1414.

——— (1996): "Testing for Serial Correlation Against an ARMA(1, 1) Process," *Journal of the American Statistical Association,* 91, 1331-1342.

Arcones, M. A. and B. Yu (1994): "Central Limit Theorems for Empirical and U-processes of Stationary Mixing Sequences," *Journal of Theoretical Probability*, 7, 47–71.

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 54, 573-578.

Doukhan, P., P. Massart, and E. Rio (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annals of the Institute of Henri Poincaré*, 31, 393–427.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Applications.* New York: Academic Press.

Hansen, B. E. (1996): "Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays," *Econometric Theory*, 12, 347-359.

Kim, J. and D. Pollard (1990): "Cube Root Asymptotics" *Annals of Statistics,* 17, 191-219.

Pollard, D. (1985): "New Ways to Prove Central Limit Theorems," *Econometric Theory*, 1, 295-313.

Stock, J. H. and J. H. Wright (2000): "GMM with Weak Instruments," *Econometrica,* 68, 1055-1096.

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: Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes.* New York: Springer.