SUPPLEMENT TO "ESTIMATION AND INFERENCE WITH WEAK, SEMI-STRONG, AND STRONG IDENTIFICATION"

(*Econometrica*, Vol. 80, No. 5, September 2012, 2153–2211)

BY DONALD W. K. ANDREWS AND XU CHENG

CONTENTS

 DOI: 10.3982/ECTA9456

7. OUTLINE

WE LET AC1 ABBREVIATE the main paper. This Supplemental Material includes six appendices.

Supplemental Appendix A provides (i) a verbal description of the steps in the proofs of the results in AC1, (ii) the vector β version of Assumption V1, (iii) details concerning the type 2 null-imposed (NI) robust CS, (iv) sufficient conditions for Assumptions B3, C5, C6, C1, and D1 (in that order), (v) an initial conditions adjustment to the sufficient conditions for Assumptions C1 and D1 that is useful in some time series contexts, and (vi) a brief discussion of reparametrization in the bivariate probit model with endogeneity considered in Han (2009). Sufficient conditions for other assumptions in AC1 are given in Andrews and Cheng (2011a, 2011b).

Supplemental Appendix B gives the proofs of the results in AC1, and states and proves results for the restricted estimator $\widetilde{\theta}_n$.

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

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

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

Supplemental Appendix F considers the standard linear instrumental variables regression model with one right-hand side endogenous variable. This Appendix compares the power of the robust tests introduced in AC1 with the power of the CLR test of Moreira (2003).

The notational conventions specified at the end of the Introduction to AC1 are used throughout this Supplemental Material. In addition, let $o_{p\pi}(1)$, $O_{p\pi}(1)$, and $o_{\pi}(1)$ denote terms that are $o_p(1)$, $O_p(1)$, and $o(1)$, respectively, uniformly over a parameter $\pi \in \Pi$. Thus, $X_n(\pi) = o_{p\pi}(1)$ means that $\sup_{\pi\in\Pi} \|X_n(\pi)\| = o_p(1)$, where $\|\cdot\|$ denotes the Euclidean norm. Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes indexed by $\pi \in \Pi$ for some space Π. The definition of weak convergence of R^v-valued functions on Π requires the specification of a metric d on the space $\mathcal{E}_v$ of R^v-valued functions on Π. We take d to be the uniform metric. The literature contains several definitions of weak convergence. We use any of the definitions that are compatible with the use of the uniform metric and for which the continuous mapping theorem (CMT) holds. These include the definitions employed by Pollard (1984, p. 65, 1990, p. 44) and van der Vaart and Wellner (1996, p. 17). The CMT's that correspond to these definitions are given by Pollard (1984, p. 70, 1990, p. 46) and van der Vaart and Wellner (1996, Theorem 1.3.6, p. 20). In the event of measurability issues, outer probabilities are used below implicitly in place of probabilities.

8. SUPPLEMENTAL APPENDIX A

8.1. *Description of Approach*

The criterion functions/models considered in AC1 possess the following characteristics:

(i) The criterion function does not depend on π when $\beta = 0$ (Assumption A in Section 1).

(ii) The criterion function viewed as a function of ψ with π fixed has a (stochastic) quadratic approximation w.r.t. ψ (for ψ close to the true value of ψ) for each $\pi \in \Pi$ when the true β is close to the nonidentification value 0 (Assumption C1 in Section 3.3).

(iii) The (generalized) first derivative of this quadratic expansion converges weakly as a process indexed by $\pi \in \Pi$ to a Gaussian process after suitable normalization (Assumption C3 in Section 3.3).

(iv) The (generalized) Hessian of this quadratic expansion is nonsingular asymptotically for all $\pi \in \Pi$ after suitable normalization (Assumption C4 in Section 3.3).

(v) The criterion function viewed as a function of θ has a (stochastic) quadratic approximation w.r.t. θ (for θ close to the true value) whether or not the true β is close to the nonidentification value 0 (Assumption D1 in Section 3.5).

(vi) The (generalized) first derivative of this quadratic expansion has an asymptotic normal distribution, where a matrix rescaling is employed when β is local to the nonidentification value 0 (Assumption D3 in Section 3.5).

(vii) The (generalized) Hessian of this quadratic expansion is nonsingular asymptotically, where a matrix rescaling is used when β is local to the nonidentification value 0 (Assumption D2 in Section 3.5).

Now we describe the approach used to establish the asymptotic results. The estimator $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{\zeta}_n, \widehat{\pi}_n)$ is defined to minimize a criterion function $Q_n(\theta)$ over $\theta \in \Theta$. Let $\theta_n = (\beta_n, \zeta_n, \pi_n)$ denote the true parameter.

Several steps are employed. The first three steps apply to sequences of true parameters in categories I and II of Table I.

Step 1. We consider the concentrated estimator $\widehat{\psi}_n(\pi)$ that minimizes $Q_n(\theta) = Q_n(\psi, \pi)$ over ψ for fixed $\pi \in \Pi$ and the concentrated criterion function $Q_n^c(\pi) = Q_n(\widehat{\psi}_n(\pi), \pi)$. We show that $\widehat{\psi}_n(\pi)$ is consistent for ψ_n uniformly over $\pi \in \Pi$ (Lemma 3.1). The method of proof is a variation of a standard consistency proof for extremum estimators adjusted to yield uniformity over π. The proof is analogous to that used in Andrews (1993) for estimators of structural change models in the situation where no structural change occurs.

Step 2. We employ a stochastic quadratic expansion of $Q_n(\psi, \pi)$ in ψ for given π about the nonidentification point $\psi = \psi_{0,n} = (0, \zeta_n)$, rather than the true value ψ_n, which is key. By expanding about $\psi_{0,n}$, the leading term of the expansion, $Q_n(\psi_{0,n}, \pi)$, does not depend on π because $Q_n(\beta, \zeta, \pi)$ does not depend on π when $\beta = 0$. For each $\pi \in \Pi$, we obtain a linear approximation

to $\widehat{\psi}_n(\pi)$ after centering around $\psi_{0,n}$ and rescaling (Lemma 9.2(b)). At the same time, we obtain a quadratic approximation of $Q_n^c(\pi)$ (Lemma 9.2(c)). Both results hold uniformly in π. The method employed has two steps.

The first step of the two-step method involves establishing a rate of convergence result for $\widehat{\psi}_n(\pi) - \psi_{0,n}$. The second step uses this rate of convergence result to obtain the linear approximation of $\widehat{\psi}_n(\pi) - \psi_{0,n}$ (after rescaling) and the quadratic approximation of $Q_n^c(\pi) - Q_n(\psi_{0,n}, \pi)$ (after rescaling) as a function of ψ. Because $Q_n(\psi_{0,n}, \pi)$ does not depend on π, it does not effect the behavior of $\widehat{\psi}_n(\pi)$ or $\widehat{\pi}_n$. The two-step method used here is like that used by Chernoff (1954), Pakes and Pollard (1989), and Andrews (1999), among others, except that it is carried out for a family of values π, as in Andrews (2001), rather than a single value, and the results hold uniformly over π.

Step 3. We determine the asymptotic behavior of the (generalized) first derivative of $Q_n(\psi, \pi)$ w.r.t. ψ evaluated at $\psi_{0,n}$ (Lemma 9.1). Due to the expansion about $\psi_{0,n}$, rather than about the true value ψ_n, a bias is introduced in the first derivative—its mean is not zero. The results here differ between the category I and II sequences of Table I. With category I sequences, one obtains a stochastic term (the mean zero Gaussian process $\{G(\pi) : \pi \in \Pi\}$) plus a nonstochastic term due to the bias ($K(\pi; \gamma_0)b$ in the notation of Assumption C5) and the two are of the same order of magnitude. With category II sequences, the true β_n is farther from the point of expansion 0 than with category I sequences and, in consequence, the nonstochastic bias term is of a larger order of magnitude than the stochastic term. In this case, the limit is nonstochastic.

We also determine the asymptotic behavior of the (generalized) Hessian matrix of $Q_n(\psi, \pi)$ w.r.t. ψ evaluated at $\psi_{0,n}$. It has a nonstochastic limit. There is no problem here with singularity of the Hessian because it is the Hessian for ψ only, not $\theta = (\psi, \pi)$, and ψ is identified.

For category I sequences, the results of this step combined with those of Step 2 and the condition $n^{1/2}(\psi_n - \psi_{0,n}) \to (b, 0)$ give the asymptotic distributions of (i) the concentrated estimator $\widehat{\psi}_n(\cdot)$ viewed as a stochastic process indexed by $\pi \in \Pi$, that is, $n^{1/2}(\widehat{\psi}_n(\cdot) - \psi_n) \Rightarrow \tau(\cdot)$, where $\tau(\cdot) = \tau(\cdot; \gamma_0, b)$ is a Gaussian process indexed by $\pi \in \Pi$ whose mean is nonzero unless $b = 0$, and (ii) the concentrated criterion function $Q_n^c(\cdot)$, that is, $n(Q_n^c(\cdot) - Q_n(\psi_{0,n}, \pi)) \Rightarrow \xi(\cdot)$, where $\xi(\cdot) = \xi(\cdot; \gamma_0, b)$ is a quadratic form in $\tau(\cdot)$.

For category II sequences, putting the results above together yields (i) a rate of convergence result for $\widehat{\psi}_n(\pi)$, that is, $\sup_{\pi \in \Pi} \|\widehat{\psi}_n(\pi) - \psi_{0,n}\| = O_p(\|\beta_n\|)$, that is just fast enough to obtain a rate of convergence result for $\widehat{\psi}_n - \psi_n$ in Step 6 below and (ii) the (nonstochastic) probability limit $\eta(\pi) = \eta(\pi; \gamma_0, b)$ of $Q_n^c(\pi)$ (after normalization), that is, $\|\beta_n\|^{-1}(Q_n^c(\pi) - Q_n(\psi_{0,n}, \pi)) \to_p \eta(\pi)$ uniformly over $\pi \in \Pi$.

Step 4. For category I sequences, we use $\widehat{\pi}_n = \arg\min_{\pi \in \Pi} Q_n^c(\pi)$, $n(Q_n^c(\cdot) - Q_n(\psi_{0,n}, \pi)) \Rightarrow \xi(\cdot)$ from Step 3 (where $Q_n(\psi_{0,n}, \pi)$ does not depend on π) and the continuous mapping theorem (CMT) to obtain $\widehat{\pi}_n \to_d \pi^* = \arg\min_{\pi \in \Pi} \xi(\pi)$ and $n(\inf_{\theta \in \Theta} Q_n(\theta) - Q_n(\psi_{0,n}, \pi)) = n(\inf_{\pi \in \Pi} Q_n^c(\pi) -$

$Q_n(\psi_{0,n}, \pi)) \Rightarrow \inf_{\pi \in \Pi} \xi(\pi)$. In this case, $\widehat{\pi}_n$ is not consistent. Given the asymptotic distribution of $\widehat{\pi}_n$, the result $n^{1/2}(\widehat{\psi}_n(\cdot) - \psi_n) \Rightarrow \tau(\cdot)$ from Step 3, and the CMT, we obtain the asymptotic distribution of $\widehat{\psi}_n = \widehat{\psi}_n(\widehat{\pi}_n)$, that is, $n^{1/2}(\widehat{\psi}_n - \psi_n) \to_d \tau(\pi^*)$ (Theorem 3.1). This completes the asymptotic results for $(\widehat{\psi}_n, \widehat{\pi}_n)$ for category I sequences of true parameters.

Step 5. For category II sequences, we obtain the consistency of $\widehat{\pi}_n$ by using the uniform convergence in probability of $Q_n^c(\pi)$ (after normalization) to the nonstochastic quadratic form, $\eta(\pi)$, established in Step 3, combined with the property that $\eta(\pi)$ is uniquely minimized at the limit π_0 of the true values π_n (Lemma 3.3). The vector that appears in the quadratic form $\eta(\pi)$ is the vector of biases of the (generalized) first derivative obtained in Step 3, which appears due to the expansion around $\psi_{0,n}$ rather than around ψ_n. The weight matrix of $\eta(\pi)$ is the inverse of the Hessian discussed in Step 3.

Step 6. For category II sequences, we use the rate of convergence result $\sup_{\pi \in \Pi} \|\widehat{\psi}_n(\pi) - \psi_{0,n}\| = O_p(\|\beta_n\|)$ from Step 3 and a relationship between the bias of the (generalized) first derivative and the (generalized) Hessian (w.r.t. ψ) to obtain a rate of convergence result for $\widehat{\psi}_n = \widehat{\psi}_n(\widehat{\pi}_n)$ centered at the true value ψ_n, that is, $\widehat{\psi}_n - \psi_n = o_p(\|\beta_n\|)$ (Lemmas 3.4 and 9.3).

Step 7. For category II and III sequences, we carry out stochastic quadratic expansions of $Q_n(\theta)$ about the true value θ_n. The argument proceeds as in Step 2 (but the expansion here is in θ, not in ψ with π fixed, and the expansion is about the true value). First, we obtain a rate of convergence result for $\widehat{\theta}_n - \theta_n$ and then with this rate, we obtain the asymptotic distribution of $\widehat{\theta}_n - \theta_n$ (after rescaling) using the quadratic approximation of $Q_n(\theta)$ in a particular neighborhood of θ_n. The result obtained is consistency and asymptotic normality (with mean zero) for $\widehat{\theta}_n$ with rate $n^{1/2}$ for $\widehat{\psi}_n$ for category II and III sequences, rate $n^{1/2}$ for $\widehat{\pi}_n$ for category III sequences, and rate $n^{1/2}\|\beta_n\|$ ($\ll n^{1/2}$) for $\widehat{\pi}_n$ for category II sequences (Theorem 3.2). The last rate result is due to the convergence of β_n to 0, albeit slowly. With category II sequences, $\widehat{\pi}_n$ is consistent and asymptotically normal, but has a slower rate of convergence than is standard.

For category II sequences, the results in this step are complicated by two issues. First, the (generalized) Hessian matrix for θ with the standard normalization is singular asymptotically because $\beta_n \to 0$ and the random criterion function $Q_n(\theta)$ becomes more flat w.r.t. π for β in a neighborhood of β_n the closer is β_n to 0. This requires a matrix rescaling of the Hessian based on the magnitude of $\|\beta_n\|$. Second, the quadratic approximation of the criterion function w.r.t. θ around the true value θ_n only holds for θ close enough to θ_n; specifically, only for $\theta \in \Theta_n(\delta_n) = \{\theta \in \Theta : \|\psi - \psi_n\| \leq \delta_n \|\beta_n\| \ \& \ \|\pi - \pi_n\| \leq \delta_n\}$ for constants $\delta_n \to 0$. Thus, ψ needs to be very close to the true value ψ_n for the quadratic approximation to hold. It is for this reason that the rate of convergence result $\widehat{\psi}_n - \psi_n = o_p(\|\beta_n\|)$ in Step 6 is a key result. The quadratic approximation requires $\theta \in \Theta_n(\delta_n)$ because for such $\theta = (\beta, \zeta, \pi)$, we have $\|\beta\|/\|\beta_n\| = 1 + o(1)$ and, hence, the rescaling that enters the Hessian is

asymptotically equivalent whether it is based on β or the true value β_n. (For example, see the verification of Assumption Q1(iv) for the LS example in (12.17) to see that the restriction $\theta \in \Theta_n(\delta_n)$ is required for the quadratic approximation to hold in this example.)

Step 8. We obtain the asymptotic null distributions of t test statistics for linear and nonlinear restrictions using the asymptotic distributions of the estimators described in Steps 1–7 plus asymptotic results for the variance matrix and standard error estimators upon which the test statistics depend (Theorem 4.1). The latter exhibit nonstandard behavior for category I sequences because $\widehat{\pi}_n$ is random even in the limit. These results yield the asymptotic null rejection probabilities and coverage probabilities of the standard t test for category I–III sequences.

For category I sequences, the asymptotic distribution of the t statistic for a linear or nonlinear restriction that involves both π and ψ is found to depend only on the randomness in $\widehat{\pi}_n$ and not on the randomness in $\widehat{\psi}_n$. This occurs because the former is of a larger order of magnitude than the latter. When a restriction does not involve π, then the asymptotic null distribution of the t statistic for category I sequences usually still depends on the (asymptotically nonstandard) randomness of $\widehat{\pi}_n$ through the standard deviation estimator and implicitly through the effect of the randomness of $\widehat{\pi}_n$ on the asymptotic distribution of $\widehat{\psi}_n = \widehat{\psi}_n(\widehat{\pi}_n)$.

Step 9. Next we consider the QLR test for restrictions of the form $r(\theta) = (r_1(\psi), r_2(\pi))$. The results of Step 4 give half of the asymptotic distribution of the QLR statistic for category I sequences, namely, $n(\inf_{\theta\in\Theta} Q_n(\theta) - Q_n(\psi_{0,n}, \pi)) \Rightarrow \inf_{\pi\in\Pi} \xi(\pi)$; the results of Step 7 provide half for category II and III sequences. The requisite other halves of the asymptotic null distributions of the QLR statistic are similar, but minimization is subject to the restrictions $r(\theta) = v_n$, where $v_n = r(\theta_n)$ is the true value of the restrictions. That is, one needs to establish the asymptotic distributions of $n(\inf_{\theta\in\Theta_r(v_n)} Q_n(\theta) - Q_n(\psi_{0,n}, \pi))$, where $\Theta_r(v_n) = (\theta \in \Theta : r(\theta) = v_n\}$ (Theorems 4.2 and 4.3). Determining these asymptotic distributions is noticeably more complicated than in the unrestricted case and requires innovations to the arguments given in Steps 1–7.

First, for category I sequences, the restrictions can affect the values that π can take on. In consequence, the effective parameter space for π becomes a set of the form $\Pi_r(v_{n,1})$, where $v_{n,1} = r_1(\psi_n)$, which is sample-size dependent, rather than Π. This requires a new version of the standard arg max/min theorem (see van der Vaart and Wellner (1996, Lemma 3.2.1)). The new version is given in Lemma 9.10 below. To apply this lemma, we need to define and analyze a concentrated restricted estimator $\widetilde{\psi}_n(\pi, v_{1,n})$ that is defined for all $\pi \in \Pi$ so as to determine its asymptotic behavior on $\Pi_r(v_{n,1}) \subset \Pi$.

Second, because the criterion function $Q_n(\theta)$ is not necessarily smooth (to allow for quantile estimators, etc.), one cannot use standard methods based on pointwise Taylor expansions to determine the asymptotic behavior

of $\widetilde{\psi}_n(\pi, v_{1,n})$. Instead, one has to approximate the sample-size-dependent restricted parameter space for ψ given π, denoted $\widetilde{\Psi}_n(\pi, v_{1,n})$, by a linear subspace defined by the derivatives of the restrictions. This uses the Chernoff (1954) set approximation idea, modified by Andrews (1999) to allow for data-dependent sequences of sets, and modified further by Andrews (2001) to allow for dependence on a parameter π.

Third, the quadratic expansion about $\psi_{0,n}$, rather than the true value ψ_n, in the restricted analogue of Step 2 causes new complications. With the unrestricted concentrated estimator $\widehat{\psi}_n(\pi)$, a key inequality, $a_n^2(\gamma_n)(Q_n(\widehat{\psi}_n(\pi), \pi) - Q_n(\psi_{0,n}, \pi)) \leq o_{p\pi}(1)$ (see (9.11) below), is obtained from the definition of $\widehat{\psi}_n(\pi)$, that is, $Q_n(\widehat{\psi}_n(\pi), \pi) \leq \inf_{\psi \in \Psi(\pi)} Q_n(\psi, \pi) + o_{p\pi}(n^{-1})$ in (3.2), combined with $\psi_{0,n} \in \Psi(\pi)$. However, it is not necessarily the case that $\psi_{0,n}$ lies in the restricted parameter space $\widetilde{\Psi}_n(\pi, v_{1,n})$. Hence, the previous argument fails. Instead, using a new argument, we establish a slightly weaker inequality, $a_n^2(\gamma_n)(Q_n(\widetilde{\psi}_n(\pi), \pi) - Q_n(\psi_{0,n}, \pi))) \leq O_{p\pi}(1)$ (see (9.81) below), which turns out to be sufficient.

The complications that arise in the proofs for the restricted concentrated estimator $\widetilde{\psi}_n(\pi, v_{1,n})$ are responsible for our treatment of restrictions of the form $r(\theta) = (r_1(\psi), r_2(\pi))$, rather than more general functions of θ.

Step 10. Using the asymptotic results from Steps 8 and 9 for category I–III sequences of true parameters, combined with an argument that such sequences determine the asymptotic size of tests and CS's (viz., Lemma 2.1 in Section 2), we obtain a formula for the asymptotic size of standard t and QLR tests and CS's (Theorem 4.4). Their behavior under category I sequences determines whether a test overrejects asymptotically and whether a CS undercovers asymptotically. Under category II and III sequences, they perform asymptotically as desired.

Step 11. We introduce LF and data-dependent robust critical values that yield tests and CI's that have correct asymptotic size, even in the presence of identification failure and weak identification in part of the parameter space (Theorem 5.1). The adjusted critical values employ the asymptotic formulae derived in Steps 8–10.

8.2. *Assumption V1 for Vector* β

The asymptotic behavior of the t statistic relies on Assumption V1, which concerns the variance matrix estimator. This assumption differs, depending on whether β is a scalar or a vector. The scalar version in stated in AC1. Here we state the vector version. When β is a vector (i.e., $d_\beta > 1$), we reparametrize β as $(\|\beta\|, \omega)$, where $\omega = \beta/\|\beta\|$ if $\beta \neq 0$ and, by definition, $\omega = 1_{d_\beta}/\|1_{d_\beta}\|$ with $1_{d_\beta} = (1, \ldots, 1) \in R^{d_\beta}$ if $\beta = 0$. Correspondingly, θ is reparametrized as $\theta^+ = (\|\beta\|, \omega, \zeta, \pi)$. Let $\Theta^+ = \{\theta^+ : \theta^+ = (\|\beta\|, \beta/\|\beta\|, \zeta, \pi), \theta \in \Theta\}$. Let $\widehat{\theta}_n^+$ and θ_0^+ be the counterparts of $\widehat{\theta}_n$ and θ_0 after reparametrization.

When β is a vector, let $J(\theta^+;\gamma_0)$ and $V(\theta^+;\gamma_0)$ denote some nonstochastic $d_\theta \times d_\theta$ matrix-valued functions such that $J(\theta_0^+;\gamma_0) = J(\gamma_0)$ and $V(\theta_0^+;\gamma_0) = V(\gamma_0)$. Let

$$\Sigma(\theta^+;\gamma_0) = J^{-1}(\theta^+;\gamma_0)V(\theta^+;\gamma_0)J^{-1}(\theta^+;\gamma_0), \tag{8.1}$$
$$\Sigma(\pi,\omega;\gamma_0) = \Sigma(\|\beta_0\|,\omega,\zeta_0,\pi;\gamma_0).$$

Let $\Sigma_{\beta\beta}(\pi,\omega;\gamma_0)$ denote the upper left $d_\beta \times d_\beta$ submatrix of $\Sigma(\pi,\omega;\gamma_0)$.

Assumption V1 below applies when β is a vector.

ASSUMPTION V1—Vector β: (i) $\widehat{J}_n = \widehat{J}_n(\widehat{\theta}_n^+)$ *and* $\widehat{V}_n = \widehat{V}_n(\widehat{\theta}_n^+)$ *for some (stochastic)* $d_\theta \times d_\theta$ *matrix-valued functions* $\widehat{J}_n(\theta^+)$ *and* $\widehat{V}_n(\theta^+)$ *on* Θ^+ *that satisfy* $\sup_{\theta^+\in\Theta^+}\|\widehat{J}_n(\theta^+) - J(\theta^+;\gamma_0)\| \to_p 0$ *and* $\sup_{\theta^+\in\Theta^+}\|\widehat{V}_n(\theta^+) - V(\theta^+;\gamma_0)\| \to_p 0$ *under* $\{\gamma_n\} \in \Gamma(\gamma_0,0,b)$ *with* $\|b\| < \infty$.[70]

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

(iii) $\lambda_{\min}(\Sigma(\pi,\omega;\gamma_0)) > 0$ *and* $\lambda_{\max}(\Sigma(\pi,\omega;\gamma_0)) < \infty$ $\forall\pi \in \Pi$, $\forall\omega \in R^{d_\beta}$ *with* $\|\omega\| = 1$, $\forall\gamma_0 \in \Gamma$ *with* $\beta_0 = 0$.

(iv) $P(\tau_\beta(\pi^*(\gamma_0,b);\gamma_0,b) = 0) = 0$ $\forall\gamma_0 \in \Gamma$ *with* $\beta_0 = 0$ *and* $\forall b$ *with* $\|b\| < \infty$.[71]

When β is a vector, the matrix $\overline{\Sigma}(\pi;\gamma_0,b)$ is defined differently from the scalar β case. It is defined as

$$\overline{\Sigma}(\pi;\gamma_0,b) = \Sigma(\pi,\omega^*(\pi;\gamma_0,b);\gamma_0), \quad \text{where} \tag{8.2}$$
$$\omega^*(\pi;\gamma_0,b) = \tau_\beta(\pi;\gamma_0,b)/\|\tau_\beta(\pi;\gamma_0,b)\|.$$

The upper left $d_\psi \times d_\psi$ block of $\overline{\Sigma}(\pi;\gamma_0,b)$, denoted $\overline{\Sigma}_{\psi\psi}(\pi;\gamma_0,b)$, appears in the denominator of the asymptotic t statistic in (4.5). The lower right $d_\pi \times d_\pi$ block of $\overline{\Sigma}(\pi;\gamma_0,b)$, denoted $\overline{\Sigma}_{\pi\pi}(\pi;\gamma_0,b)$, appears in the denominator of the asymptotic t statistic in (4.6).

With the changes above, Theorems 4.1, 4.4(a), and 5.1(a) hold for the t statistic and t statistic-based CI in the vector β case.

8.3. *Details for the Type 2 Robust CS With NI Critical Values*

The type 2 NI robust critical value is defined by replacing H with $H(v)$ (defined in (5.2)) in (5.8) and in the definitions of $h_{\max}$ and $b_{\max}$, which are then denoted $b_{\max}(v)$ and $h_{\max}(v)$. The set H_1 is replaced with $H_1(v) =$

[70]The functions $J(\theta^+;\gamma_0)$ and $V(\theta^+;\gamma_0)$ do not depend on ω_0, only γ_0.

[71]Assumption V1 (vector β) differs from Assumption V1 (scalar β) because in the vector β case Assumption V1(ii) (scalar β) (i.e., continuity in θ) often fails, but Assumption V1(ii) (vector β) (i.e., continuity in θ^+) holds.

$\{(b, \gamma_0) : (b, \gamma_0) \in H(v)\ \&\ \|b\| \leq \sup_{v_0 \in V_r} \|b_{\max}(v_0)\| + D\}$.[72] The constants Δ_1, Δ_2, $\Delta_1(h)$, and $\Delta_2(h)$ in (5.8) are then denoted $\Delta_1(v)$, $\Delta_2(v)$, $\Delta_1(h, v)$, and $\Delta_2(h, v)$. By definition, for any $v \in V_r$, $\mathrm{NRP}(\Delta_1(v), \Delta_2(v); h) \leq \alpha$ for all $h \in H(v)$. The NI robust critical value is denoted $\widehat{c}_{\mathcal{T},1-\alpha,n}(v)$.

For example, consider the construction of a type 2 robust CS with NI critical values for the parameter π. For each value of $v \in \Pi$, one first obtains the LF critical value $c^{\mathrm{LF}}_{\mathcal{T},1-\alpha}(v)$, and then one calculates $\Delta_1(v)$ and $\Delta_2(v)$ based on $c^{\mathrm{LF}}_{\mathcal{T},1-\alpha}(v)$ and the asymptotic distribution of $\mathcal{T}_n$ and A_n under the null $H_0 : \pi_0 = v$.

A plug-in version of the type 2 robust critical value requires the replacement of H with $\widehat{H}_n$ throughout (5.8), where $\widehat{H}_n$ is defined as in Section 5.1. Similarly, a plug-in version of the type 2 NI robust critical value is defined like the type 2 NI robust critical value, but with H replaced with $H(v) \cap \widehat{H}_n$ throughout.

Note that for a type 2 robust CS with NI critical values for β, under semi-strong or strong identification, $\Delta_1(v) \to 0$ and $\Delta_2(v) \to 0$ as $\|b\| \to \infty$, and the NI robust critical value converges to the standard critical value.

For $h \in H$ and $v \in V_r$, define

$$\widehat{c}_{\mathcal{T},1-\alpha}(h, v) = \begin{cases} c^{\mathrm{LF}}_{\mathcal{T},1-\alpha}(v) + \Delta_1(v), & \text{if } A(h) \leq \kappa, \\ c_{\mathcal{T},1-\alpha}(\infty) + \Delta_2(v) \\ \quad + [c^{\mathrm{LF}}_{\mathcal{T},1-\alpha}(v) + \Delta_1(v) - c_{\mathcal{T},1-\alpha}(\infty) - \Delta_2(v)] \\ \quad \times s(A(h) - \kappa), & \text{if } A(h) > \kappa, \end{cases} \tag{8.3}$$

where the random variable $A(h)$ is defined in (5.6). It is shown in the proof of Theorem 5.1 that the asymptotic distribution of $\widehat{c}_{\mathcal{T},1-\alpha,n}(v)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $\|b\| < \infty$ is the distribution of $\widehat{c}_{\mathcal{T},1-\alpha}(h, v)$.

Theorem 5.1 uses the following d.f. continuity condition.

ASSUMPTION NI-ROB2: (i) $P(\mathcal{T}(h) = \widehat{c}_{\mathcal{T},1-\alpha}(h, v)) = 0\ \forall h \in H(v), \forall v \in V_r$.

(ii) *For some* $v \in V_r$, $\Delta_2(v) = 0$ *or* $\mathrm{NRP}(\Delta_1(v), \Delta_2(v); h^*) = \alpha$ *for some point* $h^* \in H(v)$, *where* $\Delta_1(v)$ *and* $\Delta_2(v)$ *are defined after* (5.8).

8.4. *Assumption B3*

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

ASSUMPTION B3*: (i) $Q(\theta; \gamma_0)$ *is continuous on* $\Theta\ \forall \gamma_0 \in \Gamma$.

[72] In the definition of $H_1(v)$, the upper bound on $\|b\|$ does not vary with v, which improves the smoothness of $\Delta_1(v)$ as a function of v.

(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 8.1: *Assumption* B3* *implies Assumption* B3(ii) *and* (iii).

8.5. *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* $f_{W_i}(w; \gamma^*)$ *w.r.t. 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)

$$\int_{\mathcal{W}} \sup_{\theta \in \Theta} \|m(w, \theta)\| \cdot \max_{i \leq 1} \Big\{ \sup_{\gamma \in N(\gamma^*, \delta)} \|f_{\beta, W_i}(w; \gamma) / f_{W_i}(w; \gamma)\| \cdot \sup_{\gamma \in N(\gamma^*, \delta)} |f_{W_i}(w; \gamma)| \Big\} d\mu(w) < \infty,$$

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 8.2: *Assumption* C5* *implies that Assumption* C5 *holds with*

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

$$K(\theta;\gamma^*)=\int_{\mathcal{W}} m(w,\theta)f_{\beta,W}(w;\gamma^*)'\,d\mu(w).$$

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

8.6. *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 $m_2(W_i,\theta)\in R^{d_\xi}$ does not depend on π when $\beta=0$. In consequence, we can partition the following quantities and obtain certain subquantities that do not depend on π:

$$H(\pi;\gamma_0)=\begin{bmatrix}H_{11}(\pi) & H_{12}(\pi)\\ H_{21}(\pi) & H_{22}\end{bmatrix},\quad G(\pi;\gamma_0)=\begin{pmatrix}G_1(\pi)\\ G_2\end{pmatrix},\tag{8.4}$$
$$K(\pi;\gamma_0)=\begin{pmatrix}K_1(\pi)\\ K_2\end{pmatrix},$$

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

$$\begin{aligned}
G_1^*(\pi;\gamma_0)&=G_1(\pi)-H_{12}(\pi)H_{22}^{-1}G_2,\\
K_1^*(\pi;\gamma_0)&=K_1(\pi)-H_{12}(\pi)H_{22}^{-1}K_2,\\
H_{11}^*(\pi;\gamma_0)&=H_{11}(\pi)-H_{12}(\pi)H_{22}^{-1}H_{12}(\pi)',\\
\xi_1(\pi;\gamma_0,b)&=-\frac{1}{2}(G_1^*(\pi;\gamma_0)+K_1^*(\pi;\gamma_0)b)'H_{11}^*(\pi;\gamma_0)^{-1}\\
&\qquad\times(G_1^*(\pi;\gamma_0)+K_1^*(\pi;\gamma_0)b),\\
\xi_2(\gamma_0,b)&=-\frac{1}{2}(G_2+K_2b)'H_{22}^{-1}(G_2+K_2b),\\
\eta_1(\pi;\gamma_0,\omega_0)&=-\frac{1}{2}\omega_0'K_1^*(\pi;\gamma_0)'H_{11}^*(\pi;\gamma_0)^{-1}K_1^*(\pi;\gamma_0)\omega_0,\\
\eta_2(\gamma_0,\omega_0)&=-\frac{1}{2}\omega_0'K_2'H_{22}^{-1}K_2\omega_0.
\end{aligned}\tag{8.5}$$

LEMMA 8.3: *Suppose Assumptions* C1(iii) *and* C2–C5 *hold. Then the following equalities hold*:

(a) $\xi(\pi;\gamma_0,b)=\xi_1(\pi;\gamma_0,b)+\xi_2(\gamma_0,b)$.

(b) $\eta(\pi;\gamma_0,\omega_0)=\eta_1(\pi;\gamma_0,\omega_0)+\eta_2(\gamma_0,\omega_0)$.

COMMENT: By Lemma 8.3, 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, that is, 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 8.3 and an argument similar to that used to prove Lemma 2.6 of Kim and Pollard (1990; KP) (see Lemma 9.13 below), we obtain the following sufficient condition for Assumption C6 when β is a scalar.[73]

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

LEMMA 8.4: *Assumption* C6* *implies Assumption* C6.

Next, we provide a primitive sufficient condition for Assumption C6*. We partition the covariance kernel $\Omega(\pi_1, \pi_2; \gamma_0)$ in Assumption C3 analogously to $H(\pi; \gamma_0)$ and obtain

$$\Omega(\pi_1, \pi_2; \gamma_0) = \begin{bmatrix} \Omega_{11}(\pi_1, \pi_2; \gamma_0) & \Omega_{12}(\pi_1; \gamma_0) \\ \Omega_{12}(\pi_2; \gamma_0)' & \Omega_{22}(\gamma_0) \end{bmatrix}, \tag{8.6}$$

where $\Omega_{22}(\gamma_0) \in R^{d_\zeta \times d_\zeta}$ does not depend on π. For any $\pi_1, \pi_2 \in \Pi$ and $\pi_1 \neq \pi_2$, $(G_1(\pi_1), G_1(\pi_2), G_2)'$ is normally distributed with mean zero and covariance matrix

$$\Omega_G(\pi_1, \pi_2; \gamma_0) = \begin{bmatrix} \Omega_{11}(\pi_1, \pi_1; \gamma_0) & \Omega_{11}(\pi_1, \pi_2; \gamma_0) & \Omega_{12}(\pi_1; \gamma_0) \\ \Omega_{11}(\pi_2, \pi_1; \gamma_0) & \Omega_{11}(\pi_2, \pi_2; \gamma_0) & \Omega_{12}(\pi_2; \gamma_0) \\ \Omega_{12}(\pi_1; \gamma_0)' & \Omega_{12}(\pi_2; \gamma_0)' & \Omega_{22}(\gamma_0) \end{bmatrix}. \tag{8.7}$$

Typically, the covariance matrix $\Omega_G(\pi_1, \pi_2; \gamma_0)$ takes the form of an outer product, which facilitates the verification of Assumption C6**, as shown in the examples.

ASSUMPTION C6**: (i) $d_\beta = 1$ (*i.e.*, β *is a scalar*).
(ii) $\Omega_G(\pi_1, \pi_2; \gamma_0)$ *is positive definite* $\forall \pi_1, \pi_2 \in \Pi$ *with* $\pi_1 \neq \pi_2$ $\forall \gamma_0 \in \Gamma$ *with* $\beta_0 = 0$.

LEMMA 8.5: *Assumption* C6** *implies Assumption* C6*, *which in turn implies Assumption* C6.

[73] Kim and Pollard (1990, Lemma 2.6) provide conditions under which the sample paths of a Gaussian process are maximized at a unique point with probability 1. Here the process of interest is a quadratic function of a Gaussian process.

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

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

$$Q_n(\theta) = n^{-1} \sum_{i=1}^{n} \rho(W_i, \theta). \tag{8.8}$$

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 (8.8), $Q(\theta; \gamma_0) = E_{\gamma_0} \rho(W_i, \theta)$.

8.7.1. *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 w.r.t. ψ, and let $\rho_\theta(W_i, \theta)$ and $\rho_{\theta\theta}(W_i, \theta)$ denote the first-order and second-order partial derivatives w.r.t. 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^* $\forall w \in \mathcal{W}$.

(iii) *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, *for all constants* $\delta_n \to 0$,

$$\sup_{\psi \in \Psi(\pi): \|\psi - \psi_{0,n}\| \le \delta_n} \Big\| n^{-1} \sum_{i=1}^{n} (\rho_{\psi\psi}(W_i, \psi, \pi) - \rho_{\psi\psi}(W_i, \psi_{0,n}, \pi)) \Big\| = o_{p\pi}(1).$$

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

$$\sup_{\theta \in \Theta_n(\delta_n)} \Big\| n^{-1} \sum_{i=1}^{n} B^{-1}(\beta_n)[\rho_{\theta\theta}(W_i, \theta) - \rho_{\theta\theta}(W_i, \theta_n)]B^{-1}(\beta_n) \Big\| = o_p(1),$$

where $\Theta_n(\delta_n) = \{\theta \in \Theta : \|\psi - \psi_n\| \le \delta_n \|\beta_n\|$ *and* $\|\pi - \pi_n\| \le \delta_n\}$.

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 rescaling 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 8.6: *Suppose Assumptions* B1 *and* B2 *hold*.
(a) *Assumption* Q1 *implies that Assumption* C1 *holds with*

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

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

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

8.7.2. *Sufficient Conditions via Stochastic Differentiability*

Next, we provide sufficient conditions for Assumptions C1 and D1 that do not require pointwise 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, and so forth.

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

$$\rho(w,\theta)=\rho(w,\theta_n)+\Delta(w,\theta_n)'(\theta-\theta_n)+r(w,\theta),\tag{8.9}$$

where $\Delta(w,\theta_n)$ is a "stochastic derivative" w.r.t. 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 $r(w,\theta)$ so as 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

$$\rho(w,\psi,\pi)=\rho(w,\psi_{0,n},\pi)+\Delta_\psi(w,\psi_{0,n},\pi)'(\psi-\psi_{0,n})+r_\psi(w,\psi,\pi),\tag{8.10}$$

where $\Delta_\psi(w,\psi_{0,n},\pi)$ is a "stochastic partial derivative" w.r.t. 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 subvector 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 (8.10) are not derivatives of $\Delta(w,\theta_n)$ and $r(w,\theta)$ that appear in (8.9).)

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

$$\nu_n r(\theta) = n^{-1/2} \sum_{i=1}^{n} (r(W_i, \theta) - E_{\gamma_n} r(W_i, \theta)), \tag{8.11}$$

where $r(w, \theta)$ is defined in (8.9). 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 (8.10).

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

$$Q_n^*(\theta) = n^{-1} \sum_{i=1}^{n} E_{\gamma_n} \rho(W_i, \theta). \tag{8.12}$$

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* $\delta_n \to 0$,

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

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

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

where $\Theta_n(\delta_n) = \{\theta \in \Theta : \|\psi - \psi_n\| \le \delta_n \|\beta_n\|$ *and* $\|\pi - \pi_n\| \le \delta_n\}$.

(v) *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, *for all constants* $\delta_n \to 0$,

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

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

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

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 piecewise differentiable in θ and is only nonsmooth in θ on a negligible set of $\{W_i : 1 \le i \le n\}$. Such cases include quantile regression, censored and truncated regression models, and so forth.

Assumption Q2(iii) and (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, Assumption Q2(iii) and (iv) can be verified easily by omitting the "1" summand in the denominators. The verification is similar to that in Lemma 8.6 above.

When $\rho(W_i, \theta)$ is not pointwise smooth, Assumption Q2(iii) and (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 the 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 Hubertype 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

$$\nu_n \Delta(\theta) = n^{-1/2} \sum_{i=1}^{n} (\Delta(W_i, \theta) - E_{\gamma_n} \Delta(W_i, \theta)), \tag{8.13}$$

where $\Delta(w, \theta)$ is defined in (8.9). 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 (8.10).

LEMMA 8.7: *Suppose Assumptions* B1 *and* B2 *hold*.

(a) *Assumption* Q2 *implies that Assumption* C1 *holds with*

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

$$D_{\psi\psi} Q_n(\theta) = \frac{\partial^2}{\partial \psi\, \partial \psi'} Q_n^*(\theta).$$

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

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

COMMENTS: (i) When $Q_n^*(\theta)$ is minimized at θ_n under $\{\gamma_n\} \in \Gamma(\gamma_0)$, $DQ_n(\theta)$ in Lemma 8.7(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 8.7(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.

(ii) In Lemma 8.7, $D_{\psi\psi}Q_n(\theta)$ and $D^2Q_n(\theta)$ are both nonrandom. With identically distributed observations, $D_{\psi\psi}Q_n(\theta)$ and $D^2Q_n(\theta)$ are second-order partial derivatives of $E_{\gamma_n}\rho(W_i, \theta)$ w.r.t. ψ and θ, respectively.

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

$$m(W_i, \theta) = \Delta_\psi(W_i, \theta) - E_{\gamma^*}\Delta_\psi(W_i, \theta) + \frac{\partial}{\partial\psi}E_{\gamma^*}\rho(W_i, \theta). \tag{8.14}$$

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, Assumption C2(ii) and (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 w.r.t. some dominating measure for $\gamma \in \Gamma$. Assumption C6 can be verified using Lemma 8.4 or 8.5. 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 nonrandom form of $D^2Q_n(\theta_n)$ given in Lemma 8.7(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 $DQ_n(\theta_n) = n^{-1/2}\nu_n\Delta_\psi(\theta_n)$.

8.7.3. *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, $Q_n^{\mathrm{IC}}(\theta)$, that is asymptotically negligible in a suitable sense. A similar adjustment is introduced in Andrews (2001).

ASSUMPTION Q3: (i) *For some function* $\rho(w, \theta) \in R$, $Q_n(\theta) = n^{-1}\sum_{i=1}^n \rho(W_i, \theta) + Q_n^{\mathrm{IC}}(\theta)$.

(ii) *Assumption* C1(ii) *holds with* $R_n(\theta)$ *replaced by* $Q_n^{\mathrm{IC}}(\theta) - Q_n^{\mathrm{IC}}(\psi_{0,n}, \pi)$ *and Assumption* D1(ii) *holds with* $R_n^*(\theta)$ *replaced by* $Q_n^{\mathrm{IC}}(\theta) - Q_n^{\mathrm{IC}}(\theta_n)$.

LEMMA 8.8: (a) *Lemma* 8.6 *holds with Assumption* Q1(i) *replaced by Assumption* Q3.

(b) *Lemma* 8.7 *holds with Assumption* Q2(i) *replaced by Assumption* Q3.

8.8. *Bivariate Probit Model With Endogeneity and Reparametrization*

Next, we briefly discuss reparametrization in the simple bivariate probit model with endogeneity considered in Han (2009) and Han and Vytlacil (2009). The model is

$$\begin{aligned} Y_i &= 1(\lambda_1 + D_i\lambda_2 - \varepsilon_i \geq 0), \\ D_i &= 1(\alpha + Z_i\beta - \nu_i \geq 0), \end{aligned} \tag{8.15}$$

where (Y_i, D_i, Z_i) is observed, $Z_i \in R$, and (ε_i, ν_i) has a bivariate normal distribution with means zero, variances normalized to equal 1, and correlation ρ. Han and Vytlacil (2009) show that the parameters are identified under some conditions including $\beta \neq 0$. If $\beta = 0$, then none of the parameters λ_1, λ_2, and ρ is identified, but a two dimensional subspace of the parameter space for these three parameters is identified. Han (2009) introduces a nonlinear transformation of $(\lambda_1, \lambda_2, \rho)$, call it (ζ_1, ζ_2, ρ), such that ρ is not identified if $\beta = 0$, but (ζ_1, ζ_2) are identified. He shows that the assumptions in AC1 hold with $\zeta = (\zeta_1, \zeta_2)$ and $\pi = \rho$. This transformation is not unique. One can create other transformations such that λ_1 is not identified when $\beta = 0$, but the other two transformed parameters are. See Han (2009) for details concerning the reparametrization that he provides.

9. SUPPLEMENTAL APPENDIX B: PROOFS

This appendix contains proofs of the following results given in AC1: (i) the asymptotic size lemma, Lemma 2.1, (ii) the asymptotic distributions of the unrestricted estimator, (iii) the asymptotic distributions of the t statistic, (iv) the asymptotic distributions of the restricted estimator and QLR statistic, and (v) the asymptotic size results for t and QLR CS's.

This appendix also provides proofs of the sufficient conditions given in Supplemental Appendix A.

9.1. *Proof of Lemma 2.1*

PROOF OF LEMMA 2.1: The proof follows the lines of the argument in Andrews and Guggenberger (2010). Define $g_n(\gamma) = (n^{1/2}\|\beta\|, \|\beta\|, \beta/\|\beta\|, \zeta, \pi, \phi)$, where by definition $\beta/\|\beta\| = 1_{d_\beta}/\|1_{d_\beta}\|$ if $\beta = 0$ and $1_{d_\beta} = (1, \ldots, 1)' \in R^{d_\beta}$. Define $G_1 = \{g : g_n(\gamma_n) \to g \text{ for some } \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \text{ with } \|b\| < \infty\}$, $G_2 = \{g : g_n(\gamma_n) \to g \text{ for some } \{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)\}$, and $G = G_1 \cup G_2$.

First, we show $\text{AsySz} \geq \min\{\inf_{h\in H} \text{CP}(h), \text{CP}_\infty\}$. Let $\{\gamma_n \in \Gamma : n \geq 1\}$ be a sequence such that $\liminf_{n\to\infty} \text{CP}_n(\gamma_n) = \liminf_{n\to\infty} \inf_{\gamma\in\Gamma} \text{CP}_n(\gamma)$ $(= \text{AsySz})$. Such a sequence always exists. Let $\{w_n : n \geq 1\}$ be a subsequence of $\{n\}$ such that $\lim_{n\to\infty} \text{CP}_{w_n}(\lambda_{w_n})$ exists and equals AsySz. Such a sequence always exists. Below we show there exists a subsequence $\{p_n\}$ of $\{w_n\}$ such that $\text{CP}_{p_n}(\gamma_{p_n}) \to \text{CP}(h)$ for some $h \in H$ or $\lim_{n\to\infty} \text{CP}_{p_n}(\gamma_{p_n}) \geq \text{CP}_\infty$. In consequence, $\text{AsySz} = \lim_{n\to\infty} \text{CP}_{p_n}(\gamma_{p_n}) \geq \min\{\inf_{h\in H} \text{CP}(h), \text{CP}_\infty\}$.

Now we show that the claim concerning the subsequence $\{p_n\}$ holds. To this end, we show (a) for any sequence $\{\gamma_n \in \Gamma : n \geq 1\}$ and any subsequence $\{w_n\}$ of n, there exists a subsequence $\{p_n\}$ of $\{w_n\}$ such that $g_{p_n}(\gamma_{p_n}) \to g$ for some $g \in G$ and (b) for any subsequence $\{p_n\}$ of $\{n\}$ and any sequence $\{\gamma_{p_n} \in \Gamma : n \geq 1\}$ for which $g_{p_n}(\gamma_{p_n}) \to g$ for some $g \in G$, $\text{CP}_{p_n}(\gamma_{p_n}) \to \text{CP}(h)$ for some $h \in H$ if $g \in G_1$ and $\liminf_{n\to\infty} \text{CP}_{p_n}(\gamma_{p_n}) \geq \text{CP}_\infty$ if $g \in G_2$.

To show (a), let $\beta_{w_n,j}$ denote the jth component of β_{w_n} and let $p_{1,n} = w_n$ $\forall n \geq 1$. For $j = 1$, either (i) $\limsup_{n\to\infty} p_{j,n}^{1/2}\beta_{p_{j,n},j} < \infty$ or (ii) $\limsup_{n\to\infty} p_{j,n}^{1/2} \times \beta_{p_{j,n},j} = \infty$. If (i) holds, then for some subsequence $\{p_{j+1,n}\}$ of $\{p_{j,n}\}$, $p_{j+1,n}^{1/2} \times \beta_{p_{j+1,n},j} \to b_j$ for some $b_j \in R$. If (ii) holds, then for some subsequence $\{p_{j+1,n}\}$ of $\{p_{j,n}\}$, $p_{j+1,n}^{1/2}\beta_{p_{j+1,n},j} \to \infty$ or $-\infty$. Applying the same argument successively for $j = 1, \ldots, d_\beta$ yields a subsequence $\{p_n^*\} = \{p_{d_\beta+1,n}\}$ of $\{w_n\}$ such that $(p_n^*)^{1/2}\beta_{p_n^*} \to b \in R^{d_\beta}$ or $(p_n^*)^{1/2}\|\beta_{p_n^*}\| \to \infty$. Because Γ is compact, there exists a subsequence $\{p_n^{**}\}$ of $\{p_n^*\}$ such that $\gamma_{p_n^{**}} \to \gamma_0 \in \Gamma$. Finally, let $\{p_n\}$ be a subsequence of $\{p_n^{**}\}$ such that $\beta_{p_n}/\|\beta_{p_n}\| \to \omega_0$. By construction, $g_{p_n}(\gamma_{p_n}) \to g = (\|b\|, \|\beta_0\|, \omega_0, \zeta_0, \pi_0, \phi_0)$, where $b \in (R \cup \{\pm\infty\})^{d_\beta}$.

It remains to show that the vector g constructed in the previous paragraph is in G. (This is needed because G is defined by the limits of full sequences rather than subsequences.) To this end, it suffices to show that there exists a sequence $\{\gamma_n^* \in \Gamma : n \geq 1\}$ such that $g_n(\gamma_n^*) \to g$ and $\gamma_{p_n}^* = \gamma_{p_n}$ $\forall n \geq 1$. Such a sequence $\{\gamma_k^* : k \geq 1\}$ can be constructed as follows: (i) $\forall k = p_n$, define $\gamma_k^* = \gamma_{p_n}$ and (ii) $\forall k \in (p_n, p_{n+1})$, define $\beta_k^* = (p_n/k)^{1/2}\beta_{p_n}$ when $\|b\| \in R$ and $\beta_k^* = \beta_{p_n}$ when $\|b\| = \infty$, and (iii) $\zeta_k^* = \zeta_{p_n}$, $\pi_k^* = \pi_{p_n}$, and $\phi_k^* = \phi_{p_n}$ in both cases. Note that when $\|b\| \in R$, $\gamma_k^* = (\beta_k^*, \zeta_k^*, \pi_k^*, \phi_k^*) \in \Gamma$ for k large by Assumption ACP(iv). When $\|b\| \in R$, $g_n(\gamma_n^*) \to g$ because $k^{1/2}\beta_k^* = p_n^{1/2}\beta_{p_n}$ $\forall k \in [p_n, p_{n+1})$, $p_n^{1/2}\beta_{p_n} \to b$ as $n \to \infty$, and $\beta_{p_n}/\|\beta_{p_n}\| \to \omega_0$ as $n \to \infty$ imply that $k^{1/2}\|\beta_k^*\| \to \|b\|$ and $\beta_k^*/\|\beta_k^*\| \to \omega_0$ as $k \to \infty$. When $\|b\| = \infty$, $k^{1/2}\|\beta_k^*\| \geq p_n^{1/2}\|\beta_{p_n}\|$ $\forall k \in [p_n, p_{n+1})$. Thus, $p_n^{1/2}\|\beta_{p_n}\| \to \infty$ as $n \to \infty$ implies $\|k^{1/2}\beta_k^*\| \to \infty$ as $k \to \infty$. In addition, when $\|b\| = \infty$, $\beta_k^*/\|\beta_k^*\| = \beta_{p_n}/\|\beta_{p_n}\|$

$\forall k \in [p_n, p_{n+1})$ and $\beta_{p_n}/\|\beta_{p_n}\| \to \omega_0$ as $n \to \infty$ implies that $\beta_k^*/\|\beta_k^*\| \to \omega_0$ as $k \to \infty$.

To show (b), note that we have shown that for any subsequence $\{p_n\}$ of $\{n\}$ and any sequence $\{\gamma_{p_n} \in \Gamma : n \geq 1\}$ for which $g_{p_n}(\gamma_{p_n}) \to g$ for some $g \in G$, there exists a sequence $\{\gamma_n^* \in \Gamma : n \geq 1\}$ such that $g_n(\gamma_n^*) \to g \in G$ and $\gamma_{p_n}^* = \gamma_{p_n}$ $\forall n \geq 1$. This and Assumption ACP(i) and (ii) imply (b). This completes the proof of AsySz $\geq \min\{\inf_{h \in H} \mathrm{CP}(h), \mathrm{CP}_\infty\}$.

Next, we show AsySz $\leq \min\{\inf_{h \in H} \mathrm{CP}(h), \mathrm{CP}_\infty\}$. First, we show that H equals

$$H^* = \big\{h = (b, \gamma_0) : n^{1/2}\beta_n \to b \in R^{d_\beta}, \gamma_n \to \gamma_0 \text{ for some } \{\gamma_n \in \Gamma : n \geq 1\}\big\}. \tag{9.1}$$

We have $H^* \subset H$ because γ_0 in H^* has $\beta_0 = 0$ since $n^{1/2}\|\beta_n\| \to \|b\| < \infty$. To show $H \subset H^*$, we need to show that for all $b \in R^{d_\beta}$ and $\gamma_0 \in \Gamma$ with $\beta_0 = 0$, there exists a sequence $\{\gamma_n \in \Gamma : n \geq 1\}$ such that $n^{1/2}\beta_n \to b$ and $\gamma_n \to \gamma_0$. Take $\gamma_n = (\beta_n, \zeta_0, \pi_0, \phi_0)$ with $\beta_n = b/n^{1/2}$ for $n \geq 1$. Then $n^{1/2}\beta_n = b$ for all n, $\gamma_n \to \gamma_0$, and $\gamma_n \in \Gamma$ for n sufficiently large that $b/n^{1/2} < \delta$ by Assumption ACP(iv).

Given that $H = H^*$, for any $h \in H$, there exists a sequence $\{\gamma_n \in \Gamma : n \geq 1\}$ such that $\{\gamma_n\} \in \Gamma\{\gamma_0, 0, b)$ by the definition of H^*. Then AsySz $= \liminf_{n\to\infty} \inf_{\gamma \in \Gamma} \mathrm{CP}_n(\gamma) \leq \liminf_{n\to\infty} \mathrm{CP}_n(\gamma_n) = \mathrm{CP}(h)$, where the last equality holds by Assumption ACP(i). There also exists a sequence $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ such that $\mathrm{CP}_n(\gamma_n) \to \mathrm{CP}_\infty$ by Assumption ACP(iii). Thus, AsySz $\leq \liminf_{n\to\infty} \mathrm{CP}_n(\gamma_n) = \mathrm{CP}_\infty$. Hence, AsySz $\leq \min\{\inf_{h \in H} \mathrm{CP}(h), \mathrm{CP}_\infty\}$ as desired. *Q.E.D.*

9.2. *Proofs of Estimation Results*

PROOF OF LEMMA 3.1: The first result of Lemma 3.1(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,

$$\begin{aligned} &P(\widehat{\psi}_n(\pi) \in \Psi(\pi)/\Psi_0 \text{ for some } \pi \in \Pi) \\ &\quad \leq P\big(Q(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0) \geq \varepsilon \text{ for some } \pi \in \Pi\big) \to 0, \end{aligned} \tag{9.2}$$

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

$$\begin{aligned} 0 &\leq \inf_{\pi \in \Pi}\big[Q(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)\big] \\ &\leq \sup_{\pi \in \Pi}\big[Q(\widehat{\psi}_n(\pi), \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)\big] \end{aligned} \tag{9.3}$$

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

where the first inequality holds by Assumption B3(ii) and the fourth inequality holds by the definition of $\widehat{\psi}_n(\pi)$ in (3.2), 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). *Q.E.D.*

The following two lemmas are used in the proofs of Lemma 3.2 and Theorem 3.1.

LEMMA 9.1: *Suppose Assumptions* B1, B2, C2, C3, *and* C5 *hold. Under* $\{\gamma_n\}\in\Gamma(\gamma_0,0,b)$, *there are two alternatives*:

(a) *When* $\|b\|<\infty$, $n^{1/2}D_\psi Q_n(\psi_{0,n},\cdot)\Rightarrow G(\cdot;\gamma_0)+K(\cdot;\gamma_0)b$.

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

Define

$$Z_n(\pi)=-a_n(\gamma_n)(D_{\psi\psi}Q_n(\psi_{0,n},\pi))^{-1}D_\psi Q_n(\psi_{0,n},\pi). \tag{9.4}$$

LEMMA 9.2: *Suppose Assumptions* A, B1–B3, *and* C1–C5 *hold. Under* $\{\gamma_n\}\in\Gamma(\gamma_0,0,b)$, *the following equalities hold*:

(a) $a_n(\gamma_n)(\widehat{\psi}_n(\pi)-\psi_{0,n})=O_{p\pi}(1)$.

(b) $a_n(\gamma_n)(\widehat{\psi}_n(\pi)-\psi_{0,n})=Z_n(\pi)+o_{p\pi}(1)$.

(c)

$$a_n^2(\gamma_n)\big(Q_n(\widehat{\psi}_n(\pi), \pi) - Q_n(\psi_{0,n}, \pi)\big)$$
$$= -\frac{1}{2} Z_n(\pi)' D_{\psi\psi} Q_n(\psi_{0,n}, \pi) Z_n(\pi) + o_{p\pi}(1).$$

COMMENT: When $\|b\| < \infty$, Lemma 9.2(b) is used to derive the asymptotic distribution of $\widehat{\psi}_n$. Lemma 9.2(c) is used in the proof of Lemma 3.2 below.

PROOF OF LEMMA 9.1: First, we decompose $D_\psi Q_n(\psi_{0,n}, \pi)$ as

$$D_\psi Q_n(\psi_{0,n}, \pi) = n^{-1/2} G_n(\pi) + n^{-1} \sum_{i=1}^n E_{\gamma_n} m(W_i, \psi_{0,n}, \pi). \tag{9.5}$$

To analyze $n^{-1} \sum_{i=1}^n E_{\gamma_n} m(W_i, \psi_{0,n}, \pi)$ 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

$$\begin{aligned} &n^{-1} \sum_{i=1}^n E_{\gamma_n} m(W_i, \psi_{0,n}, \pi) \\ &= n^{-1} \sum_{i=1}^n E_{\gamma_{0,n}} m(W_i, \psi_{0,n}, \pi) + K_n(\psi_{0,n}, \pi; \widetilde{\gamma}_n)\beta_n \\ &= K_n(\psi_{0,n}, \pi; \widetilde{\gamma}_n)\beta_n, \end{aligned} \tag{9.6}$$

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 $\gamma_n \to \gamma_0$ for γ_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(ii). Furthermore, $(\psi_{0,n}, \pi, \widetilde{\gamma}_n)$ is in the domain $\Theta_\delta \times \Gamma_0$ of $K_n(\cdot;\cdot)$ by Assumption B2(ii).

By Assumption C5,

$$K_n(\psi_{0,n}, \pi; \widetilde{\gamma}_n) \to_p K(\pi; \gamma_0) \tag{9.7}$$

uniformly over $\pi \in \Pi$. From (9.5)–(9.7), we obtain

$$D_\psi Q_n(\psi_{0,n}, \pi) = n^{-1/2} G_n(\pi) + K(\pi; \gamma_0)\beta_n + o_{p\pi}(\|\beta_n\|). \tag{9.8}$$

In part (a), in which case $n^{1/2}\beta_n \to b$ with $\|b\| < \infty$, (9.8) leads to

$$\begin{aligned} n^{1/2} D_\psi Q_n(\psi_{0,n}, \cdot) &= G_n(\cdot) + K(\cdot; \gamma_0) n^{1/2}\beta_n + o_{p\pi}(1) \\ &\Rightarrow G(\cdot; \gamma_0) + K(\cdot; \gamma_0) b, \end{aligned} \tag{9.9}$$

where the weak-convergence result holds by Assumption C3.

In part (b), in which case $n^{1/2}\|\beta_n\| \to \infty$ and $\beta_n/\|\beta_n\| \to \omega_0$, (9.8) leads to

$$
\begin{aligned}
(9.10)\qquad & \|\beta_n\|^{-1} D_\psi Q_n(\psi_{0,n}, \pi) \\
&= \big(n^{1/2}\|\beta_n\|\big)^{-1} G_n(\pi) + K(\pi; \gamma_0)\beta_n/\|\beta_n\| + o_{p\pi}(1) \\
&\to_p K(\pi; \gamma_0)\omega_0
\end{aligned}
$$

uniformly over $\pi \in \Pi$ using Assumption C3. *Q.E.D.*

PROOF OF LEMMA 9.2: 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

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

where the inequality holds $\forall \pi \in \Pi$ for n large by (3.2) 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 (9.11) holds by Assumption C1(i) with $\psi = \widehat{\psi}_n(\pi)$; the second equality holds by Lemma 3.1(a), Assumptions C1(ii) and C4, and the implication of Lemma 9.1 that $a_n(\gamma_n) D_\psi Q_n(\psi_{0,n}, \pi) = O_{p\pi}(1)$. Rearranging (9.11) 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

$$
(9.12)\qquad (\|\kappa_{n,\pi}\| - \xi_{n,\pi})^2 \leq \xi^2_{n,\pi} + o_{p\pi}(1).
$$

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

$$\Delta_n(\pi) = a_n(\gamma_n)(\widehat{\psi}_n(\pi) - \psi_{0,n}) \quad \text{and} \quad \psi_n^\dagger(\pi) = \psi_{0,n} + a_n^{-1}(\gamma_n) Z_n(\pi). \tag{9.13}$$

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

$$\begin{aligned} &a_n^2(\gamma_n)\big(Q_n(\psi_n^\dagger(\pi), \pi) - Q_n(\psi_{0,n}, \pi)\big) \\ &\quad = -\frac{1}{2} Z_n(\pi)' D_{\psi\psi,n}(\pi) Z_n(\pi) + o_{p\pi}(1), \end{aligned} \tag{9.14}$$

where the $o_{p\pi}(1)$ term is obtained from Assumption C1(ii), Lemma 9.1, 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

$$\begin{aligned} &a_n^2(\gamma_n)\big(Q_n(\widehat{\psi}_n(\pi), \pi) - Q_n(\psi_{0,n}, \pi)\big) \\ &\quad = -Z_n(\pi)' D_{\psi\psi,n}(\pi) \Delta_n(\pi) + \frac{1}{2} \Delta_n(\pi)' D_{\psi\psi,n}(\pi) \Delta_n(\pi) + o_{p\pi}(1) \\ &\quad = \frac{1}{2} (\Delta_n(\pi) - Z_n(\pi))' D_{\psi\psi,n}(\pi) (\Delta_n(\pi) - Z_n(\pi)) \\ &\qquad - \frac{1}{2} Z_n(\pi)' D_{\psi\psi,n}(\pi) Z_n(\pi) + o_{p\pi}(1), \end{aligned} \tag{9.15}$$

where the $o_{p\pi}(1)$ term in the first equality is obtained from Assumption C1(ii) and Lemma 9.2(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 $a_n^{-1}(\gamma_n) \le n^{-1/2} \to 0$. This and Assumption B1(ii) lead to

$$\psi_n^\dagger(\pi) = (0, \zeta_n) + (\beta_n^\dagger(\pi), \zeta_n^{\dagger\dagger}(\pi)) \in \Psi(\pi) \tag{9.16}$$

$\forall \pi \in \Pi$, where $\in \Psi(\pi)$ holds with probability that goes to 1 as $n \to \infty$. Specifically, (9.16) 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 1, 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 (9.16).

Using (9.16) and (3.2), we have

$$Q_n(\widehat{\psi}_n(\pi), \pi) \le Q_n(\psi_n^\dagger(\pi), \pi) + o_{p\pi}(n^{-1}) \tag{9.17}$$

$\forall \pi \in \Pi$. This, (9.14), and (9.15) give

$$\frac{1}{2}(\Delta_n(\pi) - Z_n(\pi))' D_{\psi\psi,n}(\pi)(\Delta_n(\pi) - Z_n(\pi)) \leq o_{p\pi}(1). \tag{9.18}$$

Assumption C4 and (9.18) 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 (r.h.s.) of (9.15) is $o_{p\pi}(1)$ by Lemma 9.2(b) and Assumption C4. *Q.E.D.*

PROOF OF LEMMA 3.2: Lemma 9.1(a) and Assumption C4 yield

$$Z_n(\cdot) \Rightarrow -H^{-1}(\cdot; \gamma_0)(G(\cdot; \gamma_0) + K(\cdot; \gamma_0)b) \tag{9.19}$$

under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ when $\|b\| < \infty$. Lemma 9.1(b) and Assumption C4 yield

$$Z_n(\pi) \to_p -H^{-1}(\pi; \gamma_0)K(\pi; \gamma_0)\omega_0 \tag{9.20}$$

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 9.2(c), (9.19), Assumption C4, and the CMT. Replacing (9.19) with (9.20) gives the result of part (b). *Q.E.D.*

PROOF OF THEOREM 3.1: First we prove part (a). We have $\widehat{\pi}_n \to_d \pi^*(\gamma_0, b)$ by (3.3), Lemma 3.2(a), Assumptions A, B1(iii), C3, C4(i), C5(iii), and C6, and the CMT. For details, see the proof of the arg max/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

$$\begin{aligned} \tau_n(\cdot) &= n^{1/2}(\widehat{\psi}_n(\cdot) - \psi_{0,n}) - n^{1/2}(\psi_n - \psi_{0,n}) \\ &= Z_n(\cdot) - \left(n^{1/2}\beta_n, 0_{d_\zeta}\right) + o_{p\pi}(1) \\ &\Rightarrow -H^{-1}(\cdot; \gamma_0)(G(\cdot; \gamma_0) + K(\cdot; \gamma_0)b) - (b, 0_{d_\zeta}), \end{aligned} \tag{9.21}$$

where the second equality holds by Lemma 9.2(b) and the definition of $\psi_{0,n}$, and the weak-convergence result holds by Lemma 9.1(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 nonrandom.

To prove part (b), we write

$$Q_n(\widehat{\theta}_n) = Q_n(\widehat{\psi}_n(\widehat{\pi}_n), \widehat{\pi}_n) = Q_n^c(\widehat{\pi}_n) = \inf_{\pi \in \Pi} Q_n^c(\pi) + o(n^{-1}), \tag{9.22}$$

where the first equality holds by assumption (see the paragraph following (3.3)), the second equality holds by the definition of $Q_n^c(\pi)$ given just above (3.3), and the third equality holds by (3.3). Part (b) follows from Lemma 3.2(a), (9.22), and the CMT. *Q.E.D.*

PROOF OF LEMMA 3.3: When $\beta_0 = 0$, $\widehat{\pi}_n \to_p \pi_0$ by a standard consistency argument, such as a simplification of the argument given in the proof of Lemma 3.1(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 3.2(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 $\beta_n/\|\beta_n\| \to \omega_0$) 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 3.1(a).

When $\beta_0 \neq 0$, the desired results are given in Lemma 3.1(b). *Q.E.D.*

The following lemma is used in the proof of Lemma 3.4, which is used in the proof of Theorem 3.2 below. 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 9.3: *Suppose Assumptions* C2, C4, C5, *and* C8 *hold. Then* $K(\pi_0; \gamma_0) = -H(\pi_0; \gamma_0)S_\beta'$.

PROOF OF LEMMA 9.3: For notational simplicity, define a function

$$h^n(\gamma^*, \psi) = n^{-1} \sum_{i=1}^{n} E_{\gamma^*} m(W_i, \psi, \pi^*). \tag{9.23}$$

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

$$h^n(\gamma^*, \psi^*) = 0 \quad \forall \gamma^* \in \Gamma. \tag{9.24}$$

In (9.24), ψ^* enters $h^n(\gamma^*, \psi^*)$ through both γ^* and the second argument of $h^n(\cdot, \cdot)$. Taking the derivative of $h^n(\gamma^*, \psi^*)$ w.r.t. ψ^* gives

$$h^n_{\psi^*}(\gamma^*, \psi^*) + h^n_{\psi}(\gamma^*, \psi^*) = 0 \quad \forall \gamma^* \in \Gamma. \tag{9.25}$$

The definition of $h^n(\cdot,\cdot)$ in (9.23) and the equality in (9.25) yield

$$
\begin{aligned}
(9.26)\qquad & n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\psi^{*\prime}}E_{\gamma^*}m(W_i,\psi^*,\pi^*)\\
&= h^n_{\psi^*}(\gamma^*,\psi^*) = -h^n_{\psi}(\gamma^*,\psi^*)\\
&= -n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\psi'}E_{\gamma^*}m(W_i,\psi^*,\pi^*).
\end{aligned}
$$

Postmultiplying both sides of (9.26) by S'_β, which selects the first d_β columns, yields

$$
(9.27)\qquad n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\beta^{*\prime}}E_{\gamma^*}m(W_i,\psi^*,\pi^*) = \left(-n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\psi'}E_{\gamma^*}m(W_i,\psi^*,\pi^*)\right)S'_\beta.
$$

The partial derivative $(\partial/\partial\beta^{*\prime})E_{\gamma^*}m(W_i,\psi^*,\pi^*)$ on the left-hand side (l.h.s.) of (9.27) denotes the partial derivative of $E_{\gamma^*}m(W_i,\psi^*,\pi^*)$ w.r.t. β^*, which is a subvector of the true value γ^*, whereas $(\partial/\partial\psi')E_{\gamma^*}m(W_i,\psi^*,\pi^*)$ on the r.h.s. of (9.27) denotes the partial derivative w.r.t. ψ, which is an argument of the function $m(W_i,\psi,\pi)$.

Under $\{\gamma_n\}\in\Gamma(\gamma_0,\infty,\omega_0)$, (9.27) with γ^* replaced by γ_n becomes

$$
(9.28)\qquad n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\beta^{*\prime}}E_{\gamma_n}m(W_i,\psi_n,\pi_n) = \left(-n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\psi'}E_{\gamma_n}m(W_i,\psi_n,\pi_n)\right)S'_\beta.
$$

Under $\{\gamma_n\}\in\Gamma(\gamma_0,\infty,\omega_0)$ with $\beta_0=0$, the l.h.s. of (9.28) satisfies

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

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 r.h.s. of (9.28) satisfies

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

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

Equations (9.28)–(9.30) yield the desired result. *Q.E.D.*

PROOF OF LEMMA 3.4: From Lemma 9.2(b), we have

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

where the convergence in probability holds by Lemma 9.1(b), Assumption C4, $\widehat{\pi}_n - \pi_n = o_p(1)$ (which holds by Lemma 3.3), and $\pi_n = \pi_0 + o(1)$, and the last equality holds by Lemma 9.3.

Note that

$$
(9.32)\qquad \psi_n = \psi_{0,n} + S'_{\beta}\beta_n
$$

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

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

where the first equality is straightforward, the second equality uses (9.31) and (9.32), and the last equality holds because $\|\beta_n\|^{-1}\beta_n \to \omega_0$. *Q.E.D.*

PROOF OF THEOREM 3.2: 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 9.2. Let $\kappa_n = J_n^{1/2}n^{1/2}B(\beta_n)(\widehat{\theta}_n - \theta_n)$. We have

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

where the inequality holds by (2.1), 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 1 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 3.3 and $\|\beta_n\|^{-1}(\widehat{\psi}_n - \psi_n) = o_p(1)$ by Lemma 3.4 when $\beta_n \to 0$. Rearranging (9.34) 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

$$
(9.35)\qquad (\|\kappa_n\| - \xi_n^*)^2 \leq (\xi_n^*)^2 + o_p(1).
$$

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

$$Z_n^* = -n^{1/2}J_n^{-1}B^{-1}(\beta_n)DQ_n(\theta_n), \quad \Delta_n^* = n^{1/2}B(\beta_n)(\widehat{\theta}_n - \theta_n), \tag{9.36}$$
$$\theta_n^\dagger = \theta_n + n^{-1/2}B^{-1}(\beta_n)Z_n^*.$$

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

$$n(Q_n(\theta_n^\dagger) - Q_n(\theta_n)) = -\frac{1}{2}Z_n^{*\prime}J_nZ_n^* + o_p(1), \tag{9.37}$$

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 1 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 (9.36), the structure of $B(\beta_n)$, $Z_n^* = O_p(1)$, and $n^{1/2}\|\beta_n\| \to \infty$, yield

$$\psi_n^\dagger - \psi_n = n^{-1/2}O_p(1) = o_p(\|\beta_n\|) \quad \text{and} \tag{9.38}$$
$$\pi_n^\dagger - \pi_n = n^{-1/2}\|\beta_n\|^{-1}O_p(1) = o_p(1)$$

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

$$\begin{aligned} n(Q_n(\widehat{\theta}_n) - Q_n(\theta_n)) &= -Z_n^{*\prime}J_n\Delta_n^* + \frac{1}{2}\Delta_n^{*\prime}J_n\Delta_n^* + o_p(1) \\ &= \frac{1}{2}(\Delta_n^* - Z_n^*)'J_n(\Delta_n^* - Z_n^*) - \frac{1}{2}Z_n^{*\prime}J_nZ_n^* + o_p(1), \end{aligned} \tag{9.39}$$

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 1 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 (9.38), $\theta_n \in \Theta^*$, and Assumption B1(i). In consequence,

$$Q_n(\widehat{\theta}_n) \le Q_n(\theta_n^\dagger) + o_p(1) \tag{9.40}$$

using (2.1). This, (9.37), and (9.39) give

$$\frac{1}{2}(\Delta_n^* - Z_n^*)'J_n(\Delta_n^* - Z_n^*) \le o_p(1). \tag{9.41}$$

Assumption D2, (9.39), and (9.41) imply

$$\Delta_n^* = Z_n^* + o_p(1) \quad \text{and} \quad n(Q_n(\widehat{\theta}_n) - Q_n(\theta_n)) = -\frac{1}{2}Z_n^{*\prime}J_nZ_n^* + o_p(1). \tag{9.42}$$

This, combined with Assumptions D2 and D3, gives the desired results. *Q.E.D.*

9.3. *Proofs of t Asymptotic Distributions*

The proof of Theorem 4.1 given below uses the following lemma. Define $\widehat{\omega}_n = \widehat{\beta}_n/\|\widehat{\beta}_n\|$.

LEMMA 9.4: *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 9.4: To prove Lemma 9.4(a), we have

$$\begin{aligned}(9.43)\qquad \widehat{\omega}_n &= n^{1/2}\widehat{\beta}_n/\left\|n^{1/2}\widehat{\beta}_n\right\| \to_d \frac{\tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)}{\|\tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)\|} \\ &= \omega^*(\pi^*(\gamma_0, b); \gamma_0, b)\end{aligned}$$

by the CMT, because $n^{1/2}\widehat{\beta}_n \to_d \tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)$ by Theorem 3.1(a) and Comment (i) to Theorem 3.1, and $P(\tau_\beta(\pi^*; \gamma_0, b) = 0) = 0$ by Assumption V1(iv) (vector β).

Next, we prove that Lemma 9.4(b) holds when $\beta_0 = 0$. By Lemma 3.4, $\|\beta_n\|^{-1}(\widehat{\beta}_n - \beta_n) = o_p(1)$. This implies that $\widehat{\beta}_n = \beta_n + \|\beta_n\| o_p(1)$ and $\|\widehat{\beta}_n\|/\|\beta_n\| = 1 + o_p(1)$. Hence,

$$(9.44)\qquad \widehat{\omega}_n = \frac{\widehat{\beta}_n}{\|\widehat{\beta}_n\|} = \frac{\widehat{\beta}_n - \beta_n}{\|\beta_n\|}\frac{\|\beta_n\|}{\|\widehat{\beta}_n\|} + \frac{\beta_n}{\|\beta_n\|}\frac{\|\beta_n\|}{\|\widehat{\beta}_n\|} \to_p \omega_0.$$

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 3.3. *Q.E.D.*

PROOF OF THEOREM 4.1: Under the null hypothesis $H_0 : r(\theta_n) = v_n$, the t statistic defined in (4.2) with $v = v_n$ becomes

$$(9.45)\qquad T_n = \frac{n^{1/2}(r(\widehat{\theta}_n) - r(\theta_n))}{(r_\theta(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)r_\theta(\widehat{\theta}_n)')^{1/2}}.$$

First, we prove Theorem 4.1(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 R(iii). In consequence, $r_\theta(\theta) = [r_\psi(\theta) : 0]$ and the denominator of the t statistic in (9.45) becomes

$$(9.46)\qquad (r_\theta(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)r_\theta(\widehat{\theta}_n)')^{1/2} = (r_\psi(\widehat{\theta}_n)\widehat{\Sigma}_{\psi\psi,n}r_\psi(\widehat{\theta}_n)')^{1/2}$$

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

$$
\begin{aligned}
(9.47) \qquad r(\widehat{\theta}_n) - r(\theta_n) &= r(\widehat{\psi}_n, \widehat{\pi}_n) - r(\psi_n, \widehat{\pi}_n) + r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n) \\
&= r_\psi(\widetilde{\psi}_n, \widehat{\pi}_n)(\widehat{\psi}_n - \psi_n)
\end{aligned}
$$

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

Under the conditions of Theorem 4.1(a),

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

where the first equality follows from (9.45)–(9.47), 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 (9.21)), Assumptions V1 (scalar β) and R, Theorem 3.1(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, (9.48) holds with $\widehat{\Sigma}_{\psi\psi,n}$ being the $d_\psi \times d_\psi$ upper left submatrix 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 (9.48) 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 9.4(a) and Assumption V1 (vector β). This completes the proof of part (a).

Next, we prove Theorem 4.1(b). Note that

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

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 3.1(a).

By a mean value expansion of $r(\widehat{\theta}_n) = r(\widehat{\psi}_n, \widehat{\pi}_n)$ about $\theta = (\psi_n, \widehat{\pi}_n)$, we obtain

$$\begin{aligned}
&r(\widehat{\theta}_n) = r(\psi_n, \widehat{\pi}_n) + r_\psi(\overline{\psi}_n, \widehat{\pi}_n)(\widehat{\psi}_n - \psi_n) \quad \text{and} \\
&n^{1/2}|\iota(\widehat{\beta}_n)|(r(\widehat{\theta}_n) - r(\theta_n)) \\
&\quad = n^{1/2}|\iota(\widehat{\beta}_n)|(r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n)) \\
&\qquad + |\iota(\widehat{\beta}_n)|r_\psi(\overline{\psi}_n, \widehat{\pi}_n)n^{1/2}(\widehat{\psi}_n - \psi_n) \\
&\quad = |\iota(n^{1/2}\widehat{\beta}_n)|(r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n)) + o_p(1),
\end{aligned}$$

where $\overline{\psi}_n$ lies between $\widehat{\psi}_n$ and ψ_n, and hence, $\overline{\psi}_n \to_p \psi_0$ by Lemma 3.1(a), and the third equality uses $|\iota(\widehat{\beta}_n)| = \|\widehat{\beta}_n\| = o_p(1)$, $n^{1/2}(\widehat{\psi}_n - \psi_n) = O_p(1)$, and $r_\psi(\overline{\psi}_n, \widehat{\pi}_n) = O_p(1)$, which hold by Theorem 3.1(a) and Assumption R(i).

When β is a scalar, in Theorem 4.1(b), the t statistic becomes

$$\begin{aligned}
T_n &= \frac{n^{1/2}|\iota(\widehat{\beta}_n)|(r(\widehat{\theta}_n) - r(\theta_n))}{(r_\pi(\widehat{\theta}_n)\widehat{\Sigma}_{\pi\pi,n}r_\pi(\widehat{\theta}_n)')^{1/2} + o_p(1)} \\
&= \frac{|\iota(n^{1/2}\widehat{\beta}_n)|(r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n))}{(r_\pi(\psi_0, \widehat{\pi}_n)\widehat{\Sigma}_{\pi\pi,n}r_\pi(\psi_0, \widehat{\pi}_n)')^{1/2}} + o_p(1) \\
&= T_{n,\pi}(\widehat{\pi}_n) + o_p(1) \to_d T_\pi(\pi^*; b, \gamma_0),
\end{aligned} \tag{9.50}$$

where the first equality uses (9.45) and (9.49), the second equality uses the previously displayed equation and $\overline{\psi}_n \to_p \psi_0$, the third equality defines $T_{n,\pi}(\beta, \pi)$ implicitly, and the convergence holds by arguments analogous to those used to establish the convergence in (9.48).

In the case of a vector β, (9.50) holds with $\widehat{\Sigma}_{\pi\pi,n}$ being the $d_\pi \times d_\pi$ lower right submatrix 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 (9.50) 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 (9.48) plus Lemma 9.4(a) and Assumption V1 (vector β). This completes the proof of Theorem 4.1(b).

Next, we prove Theorem 4.1(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 (9.48) holds by the same arguments as above. This equality, Assumptions V2 and R, the consistency of $\widehat{\theta}_n$ established in Lemma 3.3, Theorem 3.2(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 $\beta_n \to 0$, (9.49) still holds using $\widehat{\beta}_n \to 0$ by Lemma 3.3(b). Hence, the first equality in (9.50) also holds. In this case, the t statistic becomes

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

where the first equality follows from (9.45), (9.49), and a mean-value expansion of $r(\widehat{\theta}_n)$ w.r.t. 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 3.2(a), (ii) $\beta_n \to 0$ and the consistency of $\widehat{\theta}_n$ in Lemma 3.3, (iii) the continuity of $r_\theta(\theta)$ in Assumption R, 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) $n^{1/2}\iota(\beta_n)(\widehat{\pi}_n - \pi_n) \to_d N(0, \Sigma_{\pi\pi}(\gamma_0))$ by Theorem 3.2(a), where $\Sigma_{\pi\pi}(\gamma_0)$ is the lower right $d_\pi \times d_\pi$ submatrix of $\Sigma(\gamma_0) = J^{-1}(\gamma_0)V(\gamma_0)J^{-1}(\gamma_0)$, (iv) $|\iota(\widehat{\beta}_n)|/|\iota(\beta_n)| = \|\beta_n/\|\beta_n\| + (n^{1/2}(\widehat{\beta}_n - \beta_n)/\|n^{1/2}\beta_n\|)\| = \|\omega_n + o_p(1)\| = 1 + o_p(1)$, where the third equality uses $n^{1/2}(\widehat{\beta}_n - \beta_n) = O_p(1)$, $\|n^{1/2}\beta_n\| \to \infty$, $\omega_n = \beta_n/\|\beta_n\| \to \omega_0$, and $\|\omega_0\| = 1$, (v) if β is a scalar, $|\iota(\beta_n)|/\iota(\beta_n) = \text{sgn}(\beta_n) = 1$ w.p. $\to 1$ or $= -1$ w.p. $\to 1$ because $n^{1/2}\beta_n \to \infty$ or $n^{1/2}\beta_n \to -\infty$, (vi) if β is a vector, $|\iota(\beta_n)|/\iota(\beta_n) = 1$ because $\iota(\beta_n) = \|\beta_n\|$, (vii) Assumption V2, and (viii) the delta method.

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

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

by Theorem 3.2(a) and the delta method. By Assumptions R(i) and V2 and the consistency of $\widehat{\theta}_n$ established in Lemma 3.3,

$$
\begin{aligned}
(9.53) \qquad & r_\theta(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)r_\theta(\widehat{\theta}_n)' \\
& \to_p r_\theta(\theta_0)B^{-1}(\beta_0)\Sigma(\gamma_0)B^{-1}(\beta_0)r_\theta(\theta_0)'.
\end{aligned}
$$

The desired result follows from (9.45), (9.52), and (9.53). *Q.E.D.*

9.4. *Proofs of QLR Asymptotic Distributions and Restricted Estimator Results and Proofs*

In this section, we prove Theorems 4.2 and 4.3 concerning the asymptotic distribution of the QLR statistic. We also state and prove results concerning the asymptotic distribution of the restricted estimator $\widetilde{\theta}_n$. The QLR proofs rely on some of the results for the restricted estimator.

When γ_n is the true value, the set of π values that satisfies the restrictions $r(\theta) = v_n$ is $\Pi_r(v_{n,2})$, defined in (4.10), where $v_n = (v_{n,1}, v_{n,2}) = (r_1(\psi_n), r_2(\pi_n)) = r(\theta_n)$. We let $\Pi_{r,0} = \Pi_r(v_{0,2})$, where $v_{0,2} = \lim v_{n,2} = \lim r_2(\pi_n)$. Throughout this section, we let $o_{p\pi}(1)$ and $O_{p\pi}(1)$ denote quantities that are $o_p(1)$ and $O_p(1)$, respectively, uniformly over $\pi \in \Pi$ (not just over the restricted set $\Pi_r(v_{n,2})$) as $n \to \infty$. Thus, $X_n(\pi) = o_{p\pi}(1)$ means that $\sup_{\pi\in\Pi} \|X_n(\pi)\| = o_p(1)$, where $\|\cdot\|$ denotes the Euclidean norm.

As in AC1, we define

$$a_n(\gamma_n) = \begin{cases} n^{1/2}, & \text{if } \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \text{ and } \|b\| < \infty, \\ \|\beta_n\|^{-1}, & \text{if } \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \text{ and } \|b\| = \infty. \end{cases} \tag{9.54}$$

For notational simplicity, throughout this section we abbreviate $a_n(\gamma_n)$ by a_n and $Q_n(\psi_{0,n}, \pi)$ (which does not depend on π) by $Q_{0,n}$.

9.4.1. *Close to $\beta = 0$ Results*

In this subsection, we provide results for sequences $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for which $\|b\| < \infty$, and $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for which $\|b\| = \infty$ and $\beta_n/\|\beta_n\| \to \omega_0$ for some $\omega_0 \in R^{d_\beta}$ with $\|\omega_0\| = 1$.

The results of this subsection prove Theorem 4.2 and include results that are required for the proof of Theorem 4.3, which is given in Section 9.4.3. The proofs of the results in this subsection are given in Section 9.4.2.

To obtain the asymptotic distribution of the restricted estimators $(\widetilde{\psi}_n, \widetilde{\pi}_n)$ under sequences $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$, we need the following assumption. It is not needed to obtain the asymptotic distribution of the QLR test statistic.

The stochastic process $\{\xi_r(\pi; \gamma_0, b) : \pi \in \Pi\}$ is the limit under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$ of the restricted concentrated criterion function after suitable normalization. It is defined in (4.13).

ASSUMPTION C6r: *Each sample path of the stochastic process $\{\xi_r(\pi; \gamma_0, b) : \pi \in \Pi_{r,0}\}$ in some set $A_r(\gamma_0, b)$ with $P_{\gamma_0}(A_r(\gamma_0, b)) = 1$ is minimized over $\Pi_{r,0}$ at a unique point (which may depend on the sample path), denoted $\pi_r^*(\gamma_0, b)$, $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$, $\forall b$ with $\|b\| < \infty$.*

In Assumption C6r, $\pi_r^*(\gamma_0, b)$ is random. By Assumption C6r,

$$\pi_r^*(\gamma_0, b) = \arg\min_{\pi\in\Pi_{r,0}} \xi_r(\pi; \gamma_0, b). \tag{9.55}$$

The following matrix appears in the asymptotic distribution of the restricted estimators $(\widetilde{\psi}_n, \widetilde{\pi}_n)$:

$$P_\psi^\perp(\pi; \gamma_0) = I_{d_\psi} - P_\psi(\pi; \gamma_0), \tag{9.56}$$

where $P_\psi(\pi;\gamma_0)$ is defined in (4.13). The matrix $P_\psi^\perp(\pi;\gamma_0)$ projects obliquely onto the orthogonal complement of the space spanned by the rows of $r_{1,\psi}(\psi_0)$.

The following result gives the asymptotic distribution of the QLR statistic and the restricted estimators $(\widetilde{\psi}_n, \widetilde{\pi}_n)$ under sequences $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$.

THEOREM 9.1: *Suppose Assumptions* A, B1–B3, C1–C5, *and* RQ1 *hold*. *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ *with* $\|b\| < \infty$, *the following statements hold*:

(a) $n(Q_n(\widetilde{\theta}_n) - Q_{0,n}) \to_d \inf_{\pi \in \Pi_{r,0}} \xi_r(\pi;\gamma_0, b)$.

(b) $\mathrm{QLR}_n \to_d 2(\inf_{\pi\in\Pi_{r,0}} \xi_r(\pi;\gamma_0,b) - \inf_{\pi\in\Pi} \xi(\pi;\gamma_0,b))/s(\gamma_0)$, *provided Assumption* RQ3 *also holds*.

(c)

$$\begin{pmatrix} n^{1/2}(\widetilde{\psi}_n - \psi_n) \\ \widetilde{\pi}_n \end{pmatrix} \to_d \begin{pmatrix} P_\psi^\perp(\pi_r^*(\gamma_0,b);\gamma_0)\tau(\pi_r^*(\gamma_0,b);\gamma_0,b) \\ \pi_r^*(\gamma_0,b) \end{pmatrix}$$

provided Assumption C6r *also holds*.

COMMENTS: (i) Theorem 9.1(b) is the same as Theorem 4.2. Hence, to prove Theorem 4.2, it suffices to prove Theorem 9.1.

(ii) Define the Gaussian process $\{\tau_{r,\beta}(\pi;\gamma_0,b) : \pi \in \Pi\}$ by

$$\tau_{r,\beta}(\pi;\gamma_0,b) = S_\beta P_\psi^\perp(\pi;\gamma_0)\tau(\pi;\gamma_0,b) + b, \tag{9.57}$$

where $S_\beta = [I_{d_\beta} : 0_{d_\beta \times d_\zeta}]$ is the $d_\beta \times d_\psi$ selector matrix that selects β out of ψ. The asymptotic distribution of $n^{1/2}\widetilde{\beta}_n$ (without centering at β_n) under $\Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$ is given by $\tau_{r,\beta}(\pi_r^*(\gamma_0,b);\gamma_0,b)$. This quantity appears in the NI-ICS statistic $A_n(v_n)$ defined in Section 5.2 of AC1.

(iii) Suppose the assumptions of Theorem 9.1(c) hold, and Assumptions V1 and V2 hold with $\widetilde{J}_n$ and $\widetilde{V}_n$ in place of $\widehat{J}_n$ and $\widehat{V}_n$, respectively. Then in the scalar β case, the NI-ICS statistic $A_n(v_n)$ satisfies

$$\begin{aligned} &A_n(v_n) \to_d A(h, v_0) \quad \text{under} \quad \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \\ &\quad \text{with} \quad \|b\| < \infty, \quad \text{where} \\ &A(h, v_0) = (\tau_{r,\beta}(\pi_r^*;\gamma_0,b)'\Sigma_{r,\beta\beta}^{-1}(\pi_r^*;\gamma_0)\tau_{r,\beta}(\pi_r^*;\gamma_0,b)/d_\beta)^{1/2}, \end{aligned} \tag{9.58}$$

$v_0 = r(\theta_0)$, π_r^* abbreviates $\pi_r^*(\gamma_0, b)$, and $\Sigma_{r,\beta\beta}(\pi;\gamma_0)$ is the upper left $d_\beta \times d_\beta$ submatrix of $\Sigma_r(\pi;\gamma_0)$. The matrix $\Sigma_r(\pi;\gamma_0)$ is defined by

$$\begin{aligned} &\Sigma_r(\pi;\gamma_0) = \Sigma_r(\psi_0, \pi;\gamma_0), \\ &\Sigma_r(\theta;\gamma_0) = P_\theta^\perp(\gamma_0)J^{-1}(\theta;\gamma_0)V(\theta;\gamma_0)J^{-1}(\theta;\gamma_0)P_\theta^\perp(\gamma_0)', \\ &P_\theta^\perp(\gamma_0) = I_{d_\theta} - P_\theta(\gamma_0), \\ &P_\theta(\gamma_0) = J^{-1}(\gamma_0)r_\theta(\theta_0)'(r_\theta(\theta_0)J^{-1}(\gamma_0)r_\theta(\theta_0)')^{-1}r_\theta(\theta_0). \end{aligned} \tag{9.59}$$

In the vector β case, β is reparametrized as $(\|\beta\|, \omega)$, as in Section 8.2 in Supplemental Appendix A. Correspondingly, θ is reparametrized as $\theta^+ = (\|\beta\|, \omega, \zeta, \pi)$. In the vector β case, $\Sigma_{r,\beta\beta}(\pi; \gamma_0)$ is replaced in (9.58) by $\Sigma_{r,\beta\beta}(\pi, \omega_r^*(\pi; \gamma_0, b); \gamma_0)$, where $\omega_r^*(\pi; \gamma_0, b) = \tau_{r,\beta}(\pi; \gamma_0, b)/\|\tau_{r,\beta}(\pi; \gamma_0, b)\|$ (defined analogously to $\omega^*(\pi; \gamma_0, b)$ in (8.2) in Supplemental Appendix A) and $\Sigma_{r,\beta\beta}(\pi, \omega; \gamma_0)$ is the upper left $d_\beta \times d_\beta$ submatrix of $\Sigma_r(\pi, \omega; \gamma_0)$. The matrix $\Sigma_r(\pi, \omega; \gamma_0)$ is defined by

$$\begin{aligned}(9.60)\qquad &\Sigma_r(\pi, \omega; \gamma_0) = \Sigma_r(\|\beta_0\|, \omega, \zeta_0, \pi; \gamma_0),\\ &\Sigma_r(\theta^+; \gamma_0) = P_\theta^\perp(\gamma_0) J^{-1}(\theta^+; \gamma_0) V(\theta^+; \gamma_0) J^{-1}(\theta^+; \gamma_0) P_\theta^\perp(\gamma_0)'\end{aligned}$$

(analogously to the definitions in (8.1)), where $J(\theta^+; \gamma_0)$ and $V(\theta^+; \gamma_0)$ are the nonstochastic $d_\theta \times d_\theta$ matrix-valued functions that appear in Assumption V1 (vector β) in Section 8.2 in Supplemental Appendix A and are such that $J(\theta_0^+; \gamma_0) = J(\gamma_0)$ and $V(\theta_0^+; \gamma_0) = V(\gamma_0)$.

Note that when the type 2 robust critical value is considered in the vector β case, h is defined to include $\omega_0 \in R^{d_\beta}$ with $\|\omega_0\| = 1$ as an element, that is, $h = (b, \gamma_0, \omega_0)$ and $H(v) = \{h = (b, \gamma_0, \omega_0) : \|b\| < \infty, \gamma_0 \in \Gamma$ with $\beta_0 = 0, \|\omega_0\| = 1, r(\theta_0) = v\}$.

To prove Theorem 9.1, we start by defining a concentrated restricted estimator $\widetilde{\psi}_n(\pi, v_1)$ of ψ. This estimator is restricted only by the restrictions on ψ. It is defined for all $\pi \in \Pi$, not just for those π that satisfy the restrictions $r_2(\pi) = v_{n,2}$, that is, $\pi \in \Pi_r(v_{n,2})$. This is important for the use of the extended CMT and the extended arg max/min theorems below. For given $\pi \in \Pi$ and $v = (v_1, v_2) \in r(\Theta)$, let

$$\begin{aligned}(9.61)\qquad &\widetilde{\psi}_n(\pi, v_1) \in \Psi_r(\pi, v_1) \quad \text{and}\\ &Q_n(\widetilde{\psi}_n(\pi, v_1), \pi) = \inf_{\psi \in \Psi_r(\pi, v_1)} Q_n(\psi, \pi) + o(n^{-1}), \quad \text{where}\\ &\Psi_r(\pi, v_1) = \{\psi : (\psi, \pi) \in \Theta, r_1(\psi) = v_1\}\end{aligned}$$

and the $o(n^{-1})$ term does not depend on π.

Let $Q_n^{\mathrm{cr}}(\pi, v_1)$ denote the concentrated restricted criterion function $Q_n(\widetilde{\psi}_n(\pi, v_1), \pi)$ for $\pi \in \Pi$. Define a restricted extremum estimator $\widetilde{\pi}_n(v) \in \Pi_r(v_2)$ by

$$(9.62)\qquad Q_n^{\mathrm{cr}}(\widetilde{\pi}_n(v), v_1) = \inf_{\pi \in \Pi_r(v_2)} Q_n^{\mathrm{cr}}(\pi, v_1) + o(n^{-1}).$$

Analogously to $\widehat{\theta}_n$, we assume $\widetilde{\theta}_n(v)$ can be written as

$$(9.63)\qquad \widetilde{\theta}_n(v) = \big(\widetilde{\psi}_n(\widetilde{\pi}_n(v), v_1), \widetilde{\pi}_n(v)\big).$$

In this section, we use the notational simplifications

$$\begin{aligned}(9.64)\quad &\mathrm{QLR}_n = \mathrm{QLR}_n(v_n), \quad \widetilde{\theta}_n = \widetilde{\theta}_n(v_n), \quad \widetilde{\psi}_n(\pi) = \widetilde{\psi}_n(\pi, v_{n,1}),\\ &\widetilde{\pi}_n = \widetilde{\pi}_n(v_n), \quad \text{where}\\ &v_n = (v_{n,1}, v_{n,2}) = r(\theta_n) \quad \text{and} \quad \gamma_n = (\theta_n, \phi_n).\end{aligned}$$

Thus, the asymptotic results given below are results that hold when the restrictions are true.

The first result is a uniform consistency result for the concentrated estimator $\widetilde{\psi}_n(\pi)$.

LEMMA 9.5: *Suppose Assumptions* A, B3, *and* RQ1 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0)$, *where* $\gamma_0 = (\beta_0, \zeta_0, \pi_0, \phi_0)$ *and* $\beta_0 = 0$, $\sup_{\pi\in\Pi} \|\widetilde{\psi}_n(\pi) - \psi_n\| \to_p 0$.

COMMENT: Assumption RQ1(v), defined in Section 4.5, is used in the proof of this lemma and nowhere else. Assumption RQ1(vi) is used in the proof of Lemma 9.11 below and nowhere else.

The second result is a uniform rate of convergence result for $\widetilde{\psi}_n(\pi)$.

LEMMA 9.6: *Suppose Assumptions* A, B1–B3, C1–C5, *and* RQ1 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $\forall \pi \in \Pi$, *the following results hold*:

(a) $a_n(\widetilde{\psi}_n(\pi) - \psi_{0,n}) = O_{p\pi}(1)$.

(b) $a_n(\widetilde{\psi}_n(\pi) - \psi_n) = O_{p\pi}(1)$.

Let $D_{\psi\psi,n}(\pi)$ abbreviate $D_{\psi\psi}Q_n(\psi_{0,n}, \pi)$. The key to the results that follow is to rewrite the quadratic approximation in Assumption C1 as follows: For $\pi \in \Pi$,

$$\begin{aligned}(9.65)\quad &a_n^2(Q_n(\psi, \pi) - Q_{0,n})\\ &\quad = a_n D_\psi Q_n(\psi_{0,n}, \pi)' a_n(\psi - \psi_{0,n})\\ &\qquad + \frac{1}{2} a_n(\psi - \psi_{0,n})' D_{\psi\psi,n}(\pi) a_n(\psi - \psi_{0,n}) + a_n^2 R_n(\psi, \pi)\\ &\quad = -\frac{1}{2} Z_n(\pi)' D_{\psi\psi,n}(\pi) Z_n(\pi) + \frac{1}{2} q_n(a_n(\psi - \psi_n), \pi) + a_n^2 R_n(\psi, \pi),\end{aligned}$$

where

$$\begin{aligned}(9.66)\quad &Z_n(\pi) = -a_n D_{\psi\psi,n}^{-1}(\pi) D_\psi Q_n(\psi_{0,n}, \pi),\\ &q_n(\lambda, \pi) = (\lambda - \tau_n(\pi; \gamma_n))' D_{\psi\psi,n}(\pi)(\lambda - \tau_n(\pi; \gamma_n)),\\ &\tau_n(\pi; \gamma_n) = Z_n(\pi) + a_n(\psi_{0,n} - \psi_n)\\ &\qquad\qquad = -a_n D_{\psi\psi,n}^{-1}(\pi) D_\psi Q_n(\psi_{0,n}, \pi) - (a_n\beta_n, 0_{d_\zeta}).\end{aligned}$$

Now we define the limits of $Z_n(\pi)$, $\tau_n(\pi, \gamma_n)$, and $q_n(\lambda, \pi)$. For $\pi \in \Pi$, let

$$Z(\pi; \gamma_0) = \begin{cases} -H^{-1}(\pi; \gamma_0)(G(\pi; \gamma_0) + K(\pi; \gamma_0)b), \\ \quad \text{if } \|b\| < \infty, \\ -H^{-1}(\pi; \gamma_0)K(\pi; \gamma_0)\omega_0, \\ \quad \text{if } \|b\| = \infty \ \& \ \beta_n/\|\beta_n\| \to \omega_0. \end{cases} \tag{9.67}$$

The split definition of $Z(\pi; \gamma_0)$ appears here because, by the definition of a_n in (3.4), $a_n\beta_n = n^{1/2}\beta_n \to b$ if $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ and $\|b\| < \infty$, whereas $a_n\beta_n = \beta_n/\|\beta_n\| \to \omega_0$ if $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $\|b\| = \infty$, and $\beta_n/\|\beta_n\| \to \omega_0$. Note that $Z(\pi; \gamma_0)$ is stochastic if $\|b\| < \infty$ because $G(\pi; \gamma_0)$ is stochastic, whereas $Z(\pi; \gamma_0)$ is nonstochastic if $\|b\| = \infty$.

For $\pi \in \Pi$, define

$$\begin{aligned} \tau(\pi; \gamma_0) &= \begin{cases} Z(\pi; \gamma_0) - (b, 0_{d_\zeta}), & \text{if } \|b\| < \infty, \\ Z(\pi; \gamma_0) - (\omega_0, 0_{d_\zeta}), & \text{if } \|b\| = \infty \ \& \ \beta_n/\|\beta_n\| \to \omega_0, \end{cases} \\ &= \begin{cases} -H^{-1}(\pi; \gamma_0)(G(\pi; \gamma_0) + K(\pi; \gamma_0)b) - (b, 0_{d_\zeta}), \\ \quad \text{if } \|b\| < \infty, \\ -H^{-1}(\pi; \gamma_0)K(\pi; \gamma_0)\omega_0 - (\omega_0, 0_{d_\zeta}), \\ \quad \text{if } \|b\| = \infty \ \& \ \beta_n/\|\beta_n\| \to \omega_0. \end{cases} \end{aligned} \tag{9.68}$$

Note that $\tau(\pi; \gamma_0) = Z(\pi; \gamma_0) + \lim_{n\to\infty} a_n(\psi_{0,n} - \psi_n)$. The difference between $\tau(\pi; \gamma_0)$ and $Z(\pi; \gamma_0)$ is due to the quadratic expansion in Assumption C1 being around $\psi_{0,n}$, rather than around the true value ψ_n. Also note that if $\|b\| < \infty$, then $\tau(\pi; \gamma_0) = \tau(\pi; \gamma_0, b)$, where $\tau(\pi; \gamma_0, b)$ is defined in (3.9).

For $\pi \in \Pi$, define

$$q(\lambda, \pi) = (\lambda - \tau(\pi; \gamma_0))'H(\pi; \gamma_0)(\lambda - \tau(\pi; \gamma_0)). \tag{9.69}$$

Next, we define a minimizer, $\widetilde{\psi}_{n,q}(\pi)$, of the concentrated quadratic approximation to $Q_n(\psi, \pi)$ (which is given by the right-hand side of (9.65) with $a_n^2 R_n(\psi, \pi)$ omitted). By definition, for $\pi \in \Pi$, $\widetilde{\psi}_{n,q}(\pi)$ satisfies $\widetilde{\psi}_{n,q}(\pi) \in \Psi_r(\pi, v_{n,1})$ and

$$q_n(a_n(\widetilde{\psi}_{n,q}(\pi) - \psi_n), \pi) = \inf_{\psi \in \Psi_r(\pi, v_{n,1})} q_n(a_n(\psi - \psi_n), \pi) + o_{p\pi}(1). \tag{9.70}$$

Note that

$$\begin{aligned} &\inf_{\psi \in \Psi_r(\pi, v_{n,1})} q_n(a_n(\psi - \psi_n), \pi) = \inf_{\lambda \in a_n(\Psi_r(\pi, v_{n,1}) - \psi_n)} q_n(\lambda, \pi), \quad \text{where} \\ &a_n(\Psi_r(\pi, v_{n,1}) - \psi_n) \\ &\quad = \{\lambda \in R^{d_\psi} : \lambda = a_n(\psi - \psi_n) \text{ for some } \psi \in \Psi_r(\pi, v_{n,1})\}. \end{aligned} \tag{9.71}$$

The restricted concentrated estimators $\widetilde{\psi}_n(\pi)$ and $\widetilde{\psi}_{n,q}(\pi)$ and the criterion function $Q_n(\psi, \pi)$ evaluated at these estimators satisfy the following properties.

LEMMA 9.7: *Suppose Assumptions* A, B1–B3, C1–C5, *and* RQ1 *hold*. *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $\forall \pi \in \Pi$, *the following results hold*:

$$\text{(a)} \quad a_n(\widetilde{\psi}_{n,q}(\pi) - \psi_n) = O_{p\pi}(1).$$

$$\text{(b)} \quad a_n^2\big(Q_n(\widetilde{\psi}_n(\pi), \pi) - Q_{0,n}\big) = -\frac{1}{2} Z_n(\pi)' D_{\psi\psi,n}(\pi) Z_n(\pi) + \frac{1}{2} q_n\big(a_n(\widetilde{\psi}_n(\pi) - \psi_n), \pi\big) + o_{p\pi}(1).$$

$$\text{(c)} \quad a_n^2\big(Q_n(\widetilde{\psi}_{n,q}(\pi), \pi) - Q_{0,n}\big) = -\frac{1}{2} Z_n(\pi)' D_{\psi\psi,n}(\pi) Z_n(\pi) + \frac{1}{2} q_n\big(a_n(\widetilde{\psi}_{n,q}(\pi) - \psi_n), \pi\big) + o_{p\pi}(1).$$

$$\text{(d)} \quad a_n^2\big(Q_n(\widetilde{\psi}_n(\pi), \pi) - Q_n(\widetilde{\psi}_{n,q}(\pi), \pi)\big) = o_{p\pi}(1).$$

$$\text{(e)} \quad q_n\big(a_n(\widetilde{\psi}_n(\pi) - \psi_n), \pi\big) = q_n(a_n(\widetilde{\psi}_{n,q}(\pi) - \psi_n), \pi) + o_{p\pi}(1).$$

$$\text{(f)} \quad a_n^2\big(Q_n(\widetilde{\psi}_n(\pi), \pi) - Q_{0,n}\big) = -\frac{1}{2} Z_n(\pi)' D_{\psi\psi,n}(\pi) Z_n(\pi) + \frac{1}{2} q_n\big(a_n(\widetilde{\psi}_{n,q}(\pi) - \psi_n), \pi\big) + o_{p\pi}(1).$$

We approximate the sequence of sets $\{\Psi_r(\pi, v_{n,1}) - \psi_n : n \geq 1\}$ by the linear subspace Λ of R^{d_ψ} defined by

$$\Lambda = \{\lambda \in R^{d_\psi} : r_{1,\psi}(\psi_0)\lambda = 0\}. \tag{9.72}$$

The approximation is in the sense of Chernoff (1954), as modified in Andrews (1999) to cover drifting sequences of sets and as modified here to cover uniformity over $\pi \in \Pi$. We say that a sequence of sets indexed by $\pi \in \Pi$, $\{A_n(\pi) : n \geq 1\}$, is *locally approximated* (at the origin) by a cone $\Lambda_s \subset R^s$ uniformly over $\pi \in \Pi$ if

$$\sup_{\pi\in\Pi} \operatorname{dist}(\alpha_n(\pi), \Lambda_s) = o\Big(\sup_{\pi\in\Pi} \|\alpha_n(\pi)\|\Big) \quad \forall \{\alpha_n(\pi) \in A_n(\pi) : n \geq 1\} \text{ such that } \sup_{\pi\in\Pi} \|\alpha_n(\pi)\| \to 0, \tag{9.73}$$

$$\sup_{\pi\in\Pi} \operatorname{dist}(\lambda_n(\pi), A_n(\pi)) = o\Big(\sup_{\pi\in\Pi} \|\lambda_n(\pi)\|\Big) \quad \forall\{\lambda_n(\pi) \in \Lambda_s : n \geq 1\}$$
$$\text{such that} \quad \sup_{\pi\in\Pi} \|\lambda_n(\pi)\| \to 0.$$

LEMMA 9.8: *Suppose Assumptions* B1 *and* RQ1 *hold. Then the sequence of sets* $\{\Psi_r(\pi, v_{n,1}) - \psi_n : n \geq 1\}$ *is locally approximated* (*at the origin*) *by the cone* Λ *uniformly over* $\pi \in \Pi$.

The following result is analogous to Lemma 2 in Andrews (1999). Lemma 9.8 is used in its proof.

LEMMA 9.9: *Suppose Assumptions* A, B1–B3, C1–C5, *and* RQ1 *hold. Then, under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $\forall \pi \in \Pi$, *the following results hold*:
(a) $\inf_{\lambda\in\Lambda} q_n(\lambda, \pi) = \inf_{\lambda\in a_n(\Psi_r(\pi,v_{n,1})-\psi_n)} q_n(\lambda, \pi) + o_{p\pi}(1)$.
(b) $a_n^2(Q_n(\widetilde{\psi}_n(\pi), \pi) - Q_{0,n}) = -\frac{1}{2} Z_n(\pi)' D_{\psi\psi,n}(\pi) Z_n(\pi) + \frac{1}{2} \inf_{\lambda\in\Lambda} q_n(\lambda, \pi) + o_{p\pi}(1)$.

Let $\widetilde{\lambda}_n(\pi) \in \Lambda$ be the unique random vector that minimizes $q_n(\lambda, \pi)$ over $\lambda \in \Lambda$; that is,

$$q_n(\widetilde{\lambda}_n(\pi), \pi) = \inf_{\lambda\in\Lambda} q_n(\lambda, \pi) \quad \forall \pi \in \Pi. \tag{9.74}$$

Correspondingly, let $\widetilde{\lambda}(\pi) \in \Lambda$ be the unique random vector that minimizes $q(\lambda, \pi)$, the asymptotic analogue of $q_n(\lambda, \pi)$, over $\lambda \in \Lambda$. Specifically, define $\widetilde{\lambda}(\pi) \in \Lambda$ to be such that

$$q(\widetilde{\lambda}(\pi), \pi) = \inf_{\lambda\in\Lambda} q(\lambda, \pi) \quad \forall \pi \in \Pi. \tag{9.75}$$

Standard Lagrangean calculations for the minimum of a quadratic form subject to linear constraints yield a closed form expression for $\widetilde{\lambda}(\pi)$: For $\pi \in \Pi$,

$$\widetilde{\lambda}(\pi) = P_{\psi}^{\perp}(\pi; \gamma_0)\tau(\pi; \gamma_0), \tag{9.76}$$

where $P_{\psi}^{\perp}(\pi; \gamma_0)$ is defined in (9.56) (e.g., see Andrews (1999, p. 1361)).

Now we define the limit, $\overline{\xi}_r(\pi; \gamma_0)$, of the normalized restricted concentrated criterion function, $a_n^2(Q_n(\widetilde{\psi}_n(\pi), \pi) - Q_{0,n})$: For $\pi \in \Pi$,

$$\begin{aligned} \overline{\xi}_r(\pi; \gamma_0) &= -\frac{1}{2} Z(\pi; \gamma_0)' H(\pi; \gamma_0) Z(\pi; \gamma_0) + \frac{1}{2} \inf_{\lambda\in\Lambda} q(\lambda, \pi) \\ &= -\frac{1}{2} Z(\pi; \gamma_0)' H(\pi; \gamma_0) Z(\pi; \gamma_0) + \frac{1}{2} q(\widetilde{\lambda}(\pi), \pi) \end{aligned} \tag{9.77}$$

$$= -\frac{1}{2} Z(\pi; \gamma_0)' H(\pi; \gamma_0) Z(\pi; \gamma_0)$$
$$+ \frac{1}{2} \tau(\pi; \gamma_0)' P_\psi(\pi; \gamma_0)' H(\pi; \gamma_0) P_\psi(\pi; \gamma_0) \tau(\pi; \gamma_0).$$

As defined,

$$\overline{\xi}_r(\pi; \gamma_0) = \begin{cases} \xi_r(\pi; \gamma_0, b) = \xi(\pi; \gamma_0, b) + \frac{1}{2} \inf_{\lambda \in \Lambda} q(\lambda, \pi), \\ \quad \text{if } \|b\| < \infty, \\ \eta(\pi; \gamma_0, \omega_0) + \frac{1}{2} \inf_{\lambda \in \Lambda} q(\lambda, \pi), \\ \quad \text{if } \|b\| = \infty \text{ and } \beta_n / \|\beta_n\| \to \omega_0, \end{cases} \tag{9.78}$$

where $\xi_r(\pi; \gamma_0, b)$ is defined in (4.13), $\xi(\pi; \gamma_0, b)$ is defined in (3.8), $\eta(\pi; \gamma_0, \omega_0)$ is defined in (3.8), and the equality for $\|b\| < \infty$ holds because $\xi(\pi; \gamma_0, b) = -(1/2) Z(\pi; \gamma_0)' H(\pi; \gamma_0) Z(\pi; \gamma_0)$.

Note that if $\Lambda = R^{d_\psi}$, which corresponds to the case where there are no restrictions on ψ, then $\inf_{\lambda \in \Lambda} q(\lambda, \pi) = 0$, $\overline{\xi}_r(\pi; \gamma_0) = \xi(\pi; \gamma_0, b)$ when $\|b\| < \infty$, and $\overline{\xi}_r(\pi; \gamma_0) = \eta(\pi; \gamma_0, \omega_0)$ when $\|b\| = \infty$ and $\beta_n / \|\beta_n\| \to \omega_0$.

When $\|b\| < \infty$ and Assumption C6r holds or if $\|b\| = \infty$, $\beta_n / \|\beta_n\| \to \omega_0$, and Assumption C7 holds, we define the unique minimizer of $\overline{\xi}_r(\pi; \gamma_0)$ over the restricted set $\Pi_{r,0}$ to be

$$\pi_r^*(\gamma_0) = \arg\min_{\pi \in \Pi_{r,0}} \overline{\xi}_r(\pi; \gamma_0). \tag{9.79}$$

When $\|b\| < \infty$ and Assumption C6r holds, $\pi_r^*(\gamma_0) = \pi_r^*(\gamma_0, b) = \arg\min_{\pi \in \Pi_{r,0}} \xi_r(\pi; \gamma_0, b)$, where $\pi_r^*(\gamma_0, b)$ is defined in (9.55) and $\pi_r^*(\gamma_0)$ is random.

When $\|b\| = \infty$, $\beta_n / \|\beta_n\| \to \omega_0$, and Assumption C7 holds, $\overline{\xi}_r(\pi; \gamma_0)$ is uniquely minimized over $\pi \in \Pi_{r,0}$ by $\pi = \pi_0$, that is, $\pi_r^*(\gamma_0) = \pi_0$, because (i) (as shown below) $\tau(\pi_0; \gamma_0) = 0$, which implies that $\inf_{\lambda \in \Lambda} q(\lambda, \pi_0) = q(0_{d_\psi}, \pi_0) = 0$, and (ii) $\eta(\pi; \gamma_0, \omega_0)$ is uniquely minimized over $\pi \in \Pi_{r,0} \subset \Pi$ by $\pi = \pi_0$ by Assumption C7 $\forall \gamma_0 \in \Gamma$ with $\beta_0 = 0$. Hence, in this case, we have

$$\inf_{\pi \in \Pi_{r,0}} \overline{\xi}_r(\pi; \gamma_0) = \eta(\pi_0; \gamma_0, \omega_0) \tag{9.80}$$
$$= \omega_0' K(\pi_0; \gamma_0) H^{-1}(\pi_0; \gamma_0) K(\pi_0; \gamma_0) \omega_0.$$

Next we state a result that, in conjunction with Theorem 3.1(b), establishes Theorem 9.1. It also establishes some key results that are used in the proof of Theorem 4.3 in Section 9.4.3.

THEOREM 9.2: *Suppose Assumptions* A, B1–B3, C1–C5, *and* RQ1 *hold. Then, under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ *with* $\|b\| < \infty$ *and under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ *with* $\|b\| = \infty$ *and* $\beta_n/\|\beta_n\| \to \omega_0$, *the following statements hold*:

(a) $a_n(\widetilde{\psi}_n(\pi) - \psi_n) = \widetilde{\lambda}_n(\pi) + o_{p\pi}(1)$.

(b) $Z_n(\cdot) \Rightarrow Z(\cdot; \gamma_0)$ *and* $\tau_n(\cdot; \gamma_n) \Rightarrow \tau(\cdot; \gamma_0)$.

(c) $\widetilde{\lambda}_n(\cdot) \Rightarrow \widetilde{\lambda}(\cdot)$ *and* $a_n(\widetilde{\psi}_n(\cdot) - \psi_n) \Rightarrow \widetilde{\lambda}(\cdot)$.

(d) $a_n^2(Q_n(\widetilde{\psi}_n(\cdot), \cdot) - Q_{0,n}) \Rightarrow \overline{\xi}_r(\cdot; \gamma_0)$.

(e) $a_n^2(Q_n(\widetilde{\theta}_n) - Q_{0,n}) \to_d \inf_{\pi \in \Pi_{r,0}} \overline{\xi}_r(\pi; \gamma_0)$.

(f) $(a_n(\widetilde{\psi}_n - \psi_n), \widetilde{\pi}_n) \to_d (\tau(\pi_r^*(\gamma_0); \gamma_0), \pi_r^*(\gamma_0))$ *provided Assumption* C6r *also holds when* $\|b\| < \infty$ *and provided Assumption* C7 *also holds when* $\|b\| = \infty$.

(g) $\tau(\pi_0; \gamma_0) = 0$, $\pi_r^*(\gamma_0) = \pi_0$, $\widetilde{\pi}_n \to_p \pi_0$, *and* $\|\beta_n\|^{-1}(\widetilde{\psi}_n - \psi_n) = o_p(1)$ *when* $\|b\| = \infty$ *and* $\beta_n/\|\beta_n\| \to \omega_0$ *provided Assumptions* C7 *and* C8 *also hold*.

COMMENTS: (i) The results in Theorem 9.2(a)–(d) are for processes indexed by $\pi \in \Pi$.

(ii) Theorem 9.2(e) for the case $\|b\| < \infty$ establishes Theorem 9.1(a). Theorem 9.2(e) for the case $\|b\| < \infty$, combined with Theorem 3.1(b) and Assumption RQ3, establish Theorem 9.1(b) and hence Theorem 4.2. Theorem 9.2(f) for the case $\|b\| < \infty$ establishes Theorem 9.1(c).

(iii) Theorem 9.2(g) for the case where $\|b\| = \infty$ and $\beta_n/\|\beta_n\| \to \omega_0$ is used below in the proofs of Theorems 4.2 and 9.3.

The proof of Theorem 9.2(f) requires the following "extended" arg max/min lemma, which is analogous to the arg max Lemma 3.2.1 of van der Vaart and Wellner (1996, p. 286), but allows the set over which the max/min is taken to depend on n.

LEMMA 9.10: *Let* $\mathbb{M}_n$ *and* $\mathbb{M}$ *be stochastic processes indexed by a metric space* H. *Let* $A_n \subset H$ *and* $A_0 \subset H$ *be such that* $d_H(A_n, A_0) \to 0$, *where* d_H *denotes the Hausdorff metric. Suppose* $\mathbb{M}$ *is continuous on* H *almost surely. Suppose there exists a random element* $\widehat{h} \in A_0$ *such that almost surely* $\mathbb{M}(\widehat{h}) > \sup_{h \notin G, h \in A_0} \mathbb{M}(h)$ *for every open set* $G \subset A_0$ *that contains* $\widehat{h}$. *Suppose the sequence* $\{\widehat{h}_n \in A_n : n \geq 1\}$ *satisfies* $\mathbb{M}_n(\widehat{h}_n) \geq \sup_{h \in A_n} \mathbb{M}_n(h) + o_p(1)$. *If* $\mathbb{M}_n \Rightarrow \mathbb{M}$, *then* $\widehat{h}_n \to_d \widehat{h}$.

COMMENTS: (i) The condition on $\widehat{h}$ is satisfied if $\widehat{h}$ uniquely maximizes $\mathbb{M}(h)$ over A_0 a.s., A_0 is compact, and $\mathbb{M}$ is continuous on A_0 a.s.

(ii) $\mathbb{M}_n \Rightarrow \mathbb{M}$ means $\mathbb{M}_n \rightsquigarrow \mathbb{M}$ in $\ell^\infty(H)$ in the terminology and notation of van der Vaart and Wellner (1996).

9.4.2. *Proofs of Close to* $\beta = 0$ *Results*

PROOF OF LEMMA 9.5: The proof is the same as that for Lemma 3.1(a) with $\widetilde{\psi}_n(\pi)$ in place of $\widehat{\psi}_n(\pi)$ except that (9.3) needs to be altered because

ψ_0 does not necessarily satisfy the restriction $r_1(\psi_0) = v_{n,1}$ $(= r_1(\psi_n))$, which invalidates the fourth inequality in (9.3). However, the fourth inequality holds with $Q_n(\psi_n, \pi; \gamma_0)$ in place of $Q_n(\psi_0, \pi; \gamma_0)$ in the second summand on the right-hand side of the fourth inequality because the true value ψ_n satisfies the restriction $r_1(\psi_n) = v_{n,1}$. With this change, the fifth inequality in (9.3) has the additional term $\sup_{\pi \in \Pi} |Q(\psi_n, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)|$ on the r.h.s., which is $o(1)$ by Assumption RQ1(v). This completes the proof. *Q.E.D.*

PROOF OF LEMMA 9.6: The proof of part (a) is the same as that of Lemma 9.2(a) with $\widetilde{\psi}_n(\pi)$ in place of $\widehat{\psi}_n(\pi)$ and with Lemma 9.5 employed in place of Lemma 3.1(a), except that the inequality in (9.11) does not hold by the argument given, because (3.2) may not hold with the restricted estimator $\widetilde{\psi}_n(\pi)$ in place of $\widehat{\psi}_n(\pi)$ and (9.61) cannot be substituted in the proof for (3.2) because $\psi_{0,n}$ may not lie in the restricted set $\Psi_r(\pi, v_{n,1})$.

Instead of the inequality in (9.11), we establish the inequality

$$O_{p\pi}(1) \geq a_n^2\big(Q_n(\widetilde{\psi}_n(\pi), \pi) - Q_{0,n}\big). \tag{9.81}$$

Although the left-hand side of (9.81) is $O_{p\pi}(1)$ whereas that of (9.11) is $o_{p\pi}(1)$, (9.81) is enough for the remainder of the argument in the proof of Lemma 9.2(a) to go through.

We prove (9.81) by showing

$$\text{(i)} \quad o(1) \geq a_n^2 \sup_{\pi \in \Pi}\big(Q_n(\widetilde{\psi}_n(\pi), \pi) - Q_n(\psi_n, \pi)\big), \tag{9.82}$$

$$\text{(ii)} \quad a_n^2(Q_n(\psi_n, \pi) - Q_{0,n}) = O_{p\pi}(1).$$

Condition (i) holds because $r_1(\psi_n) = v_{n,1}$, which implies that $\psi_n \in \Psi_r(\pi, v_{n,1})$, $\widetilde{\psi}_n(\pi)$ minimizes (up to an $o(n^{-1})$ term) $Q_n(\psi, \pi)$ over $\psi_n \in \Psi_r(\pi, v_{n,1})$, and $a_n^2 \leq n^{-1}$.

To show condition (ii), we apply the quadratic approximation in Assumption C1(i) with $\psi = \psi_n$ to obtain, for $\pi \in \Pi$,

$$\begin{aligned} a_n^2(Q_n(\psi_n, \pi) - Q_{0,n}) &= a_n D_\psi Q_n(\psi_{0,n}, \pi)' a_n(\psi_n - \psi_{0,n}) \\ &\quad + a_n(\psi_n - \psi_{0,n})' D_{\psi\psi} Q_n(\psi_{0,n}, \pi) a_n(\psi_n - \psi_{0,n}) + a_n^2 R_n(\psi_n, \pi) \\ &= O_{p\pi}(1), \end{aligned} \tag{9.83}$$

where the last equality holds because (i) $a_n(\psi_n - \psi_{0,n}) = (a_n\beta_n, 0_{d_\zeta})$, $a_n\beta_n = n^{1/2}\beta_n = O(1)$ if $\|b\| < \infty$, and $a_n\beta_n = \beta_n/\|\beta_n\| = O(1)$ if $\|b\| = \infty$, (ii) $D_{\psi\psi}Q_n(\psi_{0,n}, \pi) = O_{p\pi}(1)$ by Assumption C4, (iii) $a_n D_\psi Q_n(\psi_{0,n}, \pi) = O_{p\pi}(1)$ by Lemma 9.1 (see the Comment following Lemma 9.1), and (iv) $a_n^2 \times R_n(\psi_n, \pi) = o_{p\pi}(1)$ by Assumption C1(ii) because $\|\psi_n - \psi_{0,n}\| = \|(\beta_n, 0_{d_\zeta})\| = \|\beta_n\| \to 0$ since $\beta_0 = 0$.

Part (b) follows from part (a) and the definitions of $\psi_{0,n}$ and a_n. *Q.E.D.*

PROOF OF LEMMA 9.7: The proof is analogous to the proof of Theorem 2 in Andrews (1999). To prove part (a), let $\kappa_{n,q}(\pi) = D^{1/2}_{\psi\psi,n}(\pi)a_n(\widetilde{\psi}_{n,q}(\pi) - \psi_n)$ $\forall \pi \in \Pi$. We have

$$
\begin{aligned}
(9.84)\qquad & \left\| \kappa_{n,q}(\pi) - D^{1/2}_{\psi\psi,n}(\pi)(a_n(\psi_{0,n} - \psi_n) + Z_n(\pi)) \right\|^2 \\
&= q_n\big(a_n(\widetilde{\psi}_{n,q}(\pi) - \psi_n), \pi\big) \\
&\le q_n(0, \pi) + o_{p\pi}(1) \\
&= \left\| D^{1/2}_{\psi\psi,n}(\pi)(a_n(\psi_{0,n} - \psi_n) + Z_n(\pi)) \right\|^2 + o_{p\pi}(1) = O_{p\pi}(1),
\end{aligned}
$$

where the inequality holds by (9.70) because the true value ψ_n is in $\Psi_r(\pi, v_{n,1})$ and the last equality holds by Assumption C4, Lemma 9.1, and $\|a_n\beta_n\| = O(1)$. Hence, $\kappa_{n,q}(\pi) = D^{1/2}_{\psi\psi,n}(\pi)(a_n(\psi_{0,n} - \psi_n) + Z_n(\pi)) + O_{p\pi}(1) = O_{p\pi}(1)$.

Parts (b) and (c) hold by (9.65), Assumption C1, Lemma 9.6, and part (a), using the fact that part (a) implies that $a_n(\widetilde{\psi}_{n,q}(\pi) - \psi_{0,n}) = O_{p\pi}(1)$.

Parts (d) and (e) hold by parts (b) and (c), (9.61), and (9.70):

$$
\begin{aligned}
(9.85)\qquad o(1) &\ge a_n^2 \sup_{\pi\in\Pi}\big(Q_n(\widetilde{\psi}_n(\pi), \pi) - Q_n(\widetilde{\psi}_{n,q}(\pi), \pi)\big) \\
&\ge \inf_{\pi\in\Pi} \frac{1}{2}\big(q_n\big(a_n(\widetilde{\psi}_n(\pi) - \psi_n), \pi\big) - q_n\big(a_n(\widetilde{\psi}_{n,q}(\pi) - \psi_n), \pi\big)\big) \\
&\quad + o_{p\pi}(1) \\
&\ge o_{p\pi}(1).
\end{aligned}
$$

Part (f) holds by parts (b) and (e). *Q.E.D.*

PROOF OF LEMMA 9.8: The proof is similar to the proof of Lemma 4 in Andrews (2002). Let $A_n(\pi) = \Psi_r(\pi, v_{n,1}) - \psi_n$ and $m_n(\psi) = r_1(\psi) - v_{n,1}$. By assumption, $m_n(\psi_n) = 0$ $\forall n \ge 1$, where $\gamma_n = (\psi_n, \pi_n, \phi_n)$. Let $\Gamma_a = r_{1,\psi}(\psi_0)$ $(= (\partial/\partial\psi')r_1(\psi_0))$. Define

$$
(9.86)\qquad \Gamma_* = \begin{bmatrix} \Gamma_a \\ \Gamma_b \end{bmatrix} \quad \text{and} \quad m_n^+(\psi) = \begin{pmatrix} m_n(\psi) \\ \Gamma_b(\psi - \psi_n) \end{pmatrix},
$$

where $\Gamma_b \in R^{(d_\psi - d_{r_1})\times d_\psi}$ is chosen such that $\Gamma_* \in R^{d_\psi \times d_\psi}$ is nonsingular.

Given $\{\alpha_n(\pi) \in A_n(\pi) : n \ge 1\}$ with $\sup_{\pi\in\Pi} \|\alpha_n(\pi)\| \to 0$, define

$$
(9.87)\qquad \lambda_n^*(\pi) = \Gamma_*^{-1} m_n^+(\psi_n + \alpha_n(\pi)).
$$

Then $\Gamma_*\lambda_n^*(\pi) = m_n^+(\psi_n + \alpha_n(\pi))$ and $\Gamma_a\lambda_n^*(\pi) = m_n(\psi_n + \alpha_n(\pi)) = r_1(\psi_n + \alpha_n(\pi)) - v_{n,1} = 0$, where the last equality holds because $\psi_n + \alpha_n(\pi) \in$

$\Psi_r(\pi, v_{n,1})$ since $\alpha_n(\pi) \in A_n(\pi)$. Hence, $\lambda_n^*(\pi) \in \Lambda\ \forall \pi \in \Pi$, by the definition of Λ in (9.72).

Element-by-element mean-value expansions yield

$$\begin{aligned}(9.88)\qquad \lambda_n^*(\pi) &= \Gamma_*^{-1} m_n^+(\psi_n + \alpha_n(\pi)) \\ &= \Gamma_*^{-1} m_n^+(\psi_n) + \Gamma_*^{-1} \frac{\partial}{\partial \psi'} m_n^+(\psi_n)\alpha_n(\pi) + o\big(\|\alpha_n(\pi)\|\big) \\ &= 0 + \alpha_n(\pi) + o\big(\|\alpha_n(\pi)\|\big),\end{aligned}$$

where the last equality uses the continuity of $r_{1,\psi}(\psi)$ at ψ_0 and $\psi_n \to \psi_0$ to give $\Gamma_* - (\partial/\partial\psi')m_n^+(\psi_n) \to 0$. Using (9.88), we conclude that

$$(9.89)\qquad \sup_{\pi\in\Pi} \operatorname{dist}(\alpha_n(\pi), \Lambda) \le \sup_{\pi\in\Pi} \|\alpha_n(\pi) - \lambda_n^*(\pi)\| = o\Big(\sup_{\pi\in\Pi} \|\alpha_n(\pi)\|\Big),$$

which verifies the first condition in (9.73), as desired.

Next, the function $\widetilde{m}_n(\alpha) = m_n^+(\psi_n + \alpha)$ for α in a neighborhood N_0 of 0 ($\in R^{d_\psi}$) is continuously differentiable on a neighborhood N_1 ($\subset N_0$) of 0 with nonsingular Jacobian matrix at 0 and $\widetilde{m}_n(0) = 0$. Hence, by the inverse function theorem, there exists an R^{d_ψ}-valued function $\widetilde{m}_n^{-1}(\alpha)$ for α in a neighborhood N_2 of 0 ($\in R^{d_\psi}$) that satisfies $\widetilde{m}_n^{-1}(\alpha)$ is continuously differentiable on N_2, $\widetilde{m}_n(\widetilde{m}_n^{-1}(\alpha)) = \alpha$ for all $\alpha \in N_2$, $\widetilde{m}_n^{-1}(0) = 0$, and

$$(9.90)\qquad \frac{\partial}{\partial\alpha'}\widetilde{m}_n^{-1}(0) = \left[\frac{\partial}{\partial\alpha'}\widetilde{m}_n(0)\right]^{-1} = \left[\frac{\partial}{\partial\psi'}m_n^+(\psi_n)\right]^{-1} = \Gamma_*^{-1} + o(1).$$

Given any $\{\lambda_n(\pi) \in \Lambda : n \ge 1\}$ with $\sup_{\pi\in\Pi} \|\lambda_n(\pi)\| \to 0$, define

$$(9.91)\qquad \alpha_n^*(\pi) = \widetilde{m}_n^{-1}(\Gamma_*\lambda_n(\pi)).$$

We have $m_n^+(\psi_n + \alpha_n^*(\pi)) = \widetilde{m}_n(\alpha_n^*(\pi)) = \widetilde{m}_n(\widetilde{m}_n^{-1}(\Gamma_*\lambda_n(\pi))) = \Gamma_*\lambda_n(\pi)$, which implies that $m_n(\psi_n + \alpha_n^*(\pi)) = \Gamma_a\lambda_n(\pi) = 0$, where the last equality holds for $\lambda_n(\pi) \in \Lambda$ by the definition of Λ in (9.72); that is, $r_1(\psi_n + \alpha_n^*(\pi)) = v_{n,1}\ \forall \pi \in \Pi$. In addition, $\sup_{\pi\in\Pi} \|\alpha_n^*(\pi)\| \to 0$ and Assumption B1(ii) yield $(\psi_n + \alpha_n^*(\pi), \pi) \in \Theta\ \forall \pi \in \Pi$ for n large. These results combine to give $\alpha_n^*(\pi) \in A_n(\pi)\ \forall \pi \in \Pi$ for n large.

Element-by-element mean-value expansions yield

$$\begin{aligned}(9.92)\qquad \alpha_n^*(\pi) &= \widetilde{m}_n^{-1}(\Gamma_*\lambda_n(\pi)) \\ &= \widetilde{m}_n^{-1}(0) + \frac{\partial}{\partial\alpha'}\widetilde{m}_n^{-1}(0)\Gamma_*\lambda_n(\pi) + o\big(\|\lambda_n(\pi)\|\big) \\ &= 0 + \lambda_n(\pi) + o\big(\|\lambda_n(\pi)\|\big),\end{aligned}$$

where the last equality uses (9.90). Hence,

$$\sup_{\pi\in\Pi} \operatorname{dist}(\lambda_n(\pi), A_n(\pi)) \le \sup_{\pi\in\Pi} \|\lambda_n(\pi) - \alpha_n^*(\pi)\| = o\Big(\sup_{\pi\in\Pi} \|\lambda_n(\pi)\|\Big), \tag{9.93}$$

which verifies the second condition in (9.73) and completes the proof. *Q.E.D.*

PROOF OF LEMMA 9.9: The proof of part (a) is analogous to the proof of Lemma 2 of Andrews (1999) with (i) $q_n(\lambda, \pi)$ in place of $q_T(\lambda)$, (ii) $a_n(\psi_{0,n} - \psi_n) + Z_n(\pi)$ in place Z_T, and (iii) $D_{\psi\psi,n}(\pi)$ in place of $\mathcal{J}_T$, provided $\{\Psi_r(\pi, v_{n,1}) - \psi_n : n \ge 1\}$ is locally approximated by the cone (in this case, linear subspace) Λ defined in (9.72) uniformly over $\pi \in \Pi$. The latter holds by Lemma 9.8. The quantities $a_n(\Psi_r(\pi, v_{n,1}) - \psi_n)$, $a_n I_{d_\psi}$ and a_n play the roles of $B_T(\Theta - \theta_0)$, B_T, and b_T, respectively, that appear in Assumption 5 of Andrews (1999), which is used in the proof of Lemma 2 of Andrews (1999).

Part (b) holds by part (a), Lemma 9.7(f), (9.70), and (9.71). *Q.E.D.*

PROOF OF THEOREM 9.2: The proof of part (a) holds by an argument that is analogous to the argument given in the proof of Theorem 3(a) of Andrews (1999) with (i) $a_n(\psi_{0,n} - \psi_n) + Z_n(\pi)$ in place Z_T, (ii) $D_{\psi\psi,n}(\pi)$ in place of $\mathcal{J}_T$, and (iii) indexing of the quantities by $\pi \in \Pi$, which does not create any difficulty. Theorem 3(a) of Andrews (1999) relies on Assumptions 4–6 of that paper. The analogue of Assumption 4 in the present paper is $a_n(\widetilde{\psi}_n(\pi) - \psi_n) = O_{p\pi}(1)$, which holds by Lemma 9.6(b). The analogue of Assumption 5 is the local approximation of $\{\Psi_r(\pi, v_{n,1}) - \psi_n : n \ge 1\}$ by the cone Λ uniformly over $\pi \in \Pi$, which holds by Lemma 9.8. Assumption 6 holds because Λ is a convex cone. Lemma 9.9(a) of this paper is used in the proof of part (a) because the proof of Theorem 3(a) of Andrews (1999) makes use of Lemma 2 of Andrews (1999) and Lemma 9.9(a) of this paper is the analogue of the latter.

The first result of part (b) holds by (9.19) and (9.20). The second result of part (b) holds by the first result, the fact that $\tau_n(\pi; \gamma_n) = Z_n(\pi) + a_n(\psi_{0,n} - \psi_n)$ by (9.66), $a_n(\psi_{0,n} - \psi_n) \to (-b, 0_{d_\zeta})$ if $\|b\| < \infty$, $a_n(\psi_{0,n} - \psi_n) \to (-\omega_0, 0_{d_\zeta})$ if $\|b\| = \infty$ and $\beta_n/\|\beta_n\| \to \omega_0$, and the definition of $\tau(\pi; \gamma_0)$ in (9.68).

The first result of part (c) holds by the CMT because $\widetilde{\lambda}_n(\cdot)$ is a continuous function of $(\tau_n(\cdot; \gamma_n), D_{\psi\psi,n}(\cdot))$ and $(\tau_n(\cdot; \gamma_n), D_{\psi\psi,n}(\cdot)) \Rightarrow (\tau(\cdot; \gamma_0), H(\cdot; \gamma_0))$ by part (b) and Assumption C4. Continuity holds because the oblique projection onto a convex cone Λ is both unique and continuous provided the weighting matrix $H(\pi; \gamma_0)$ for the oblique projection is nonsingular, which holds because $\inf_{\pi\in\Pi} \lambda_{\min}(H(\pi; \gamma_0)) > 0$ by Assumption C4. The second result of part (c) holds by the first result of part (c) and part (a).

Part (d) holds by the CMT using Lemma 9.9(b), part (b) of the theorem, Assumption C4, and (9.77).

To prove part (e), we use the result of part (d), that is, $a_n^2(Q_n(\widetilde{\psi}_n(\cdot),\cdot) - Q_{0,n}) \Rightarrow \overline{\xi}_r(\cdot;\gamma_0)$, and the extended CMT (see van der Vaart and Wellner (1996, Theorem 1.11.1, p. 67)) applied to the right-hand side of the equation

$$a_n^2(Q_n(\widetilde{\theta}_n) - Q_{0,n}) = \inf_{\pi\in\Pi_r(v_{n,2})} a_n^2\big(Q_n(\widetilde{\psi}_n(\pi),\pi) - Q_{0,n}\big), \tag{9.94}$$

which holds by (9.61)–(9.63) with $v = v_n$. The extended CMT is a generalization of the CMT that allows the continuous map to depend on n. The extended CMT is applied here with the functions $g_n(x) = \inf_{\pi\in\Pi_r(v_{n,2})} x(\pi)$ $\forall n \geq 1$ and $g(x) = \inf_{\pi\in\Pi_{r,0}} x(\pi)$, where $x = x(\pi)$ is a real-valued function on Π. The extended CMT is required here because the restricted sets $\Pi_r(v_{n,2})$ depend on n. For the extended CMT to apply, we need to show that whenever $x_n \to x$ (i.e., $\sup_{\pi\in\Pi} \|x_n(\pi) - x(\pi)\| \to 0$), where x_n and x are real-valued functions on Π with x continuous on Π, we have $g_n(x_n) \to g(x)$. (Continuity of x on Π can be assumed because the limit process $\overline{\xi}_r(\cdot;\gamma_0)$ in our application has continuous sample paths a.s.) Suppose $x_n \to x$. Then we have

$$\begin{aligned} |g_n(x_n) - g_n(x)| &= \Big|\inf_{\pi\in\Pi_r(v_{n,2})} x_n(\pi) - \inf_{\pi\in\Pi_r(v_{n,2})} x(\pi)\Big| \\ &\leq \inf_{\pi\in\Pi} |x_n(\pi) - x(\pi)| \to 0. \end{aligned} \tag{9.95}$$

In addition, by standard arguments, $g_n(x) \to g(x)$ because x is continuous on Π and $d_H(\Pi_r(v_{n,2}),\Pi_{r,0}) \to 0$ by Assumption RQ1(iv). Hence, we obtain the desired result $g_n(x_n) \to g(x)$ and the proof of part (e) is complete.

Now we establish part (f). First, we show $\widetilde{\pi}_n \to_d \pi_r^*(\gamma_0)$. We use the extended arg max lemma, Lemma 9.10, with $H = \Pi$, $h = \pi$, $\mathbb{M}_n(h) = -a_n^2(Q_n(\widetilde{\psi}_n(\pi),\pi) - Q_{0,n})$, $\mathbb{M}(h) = -\overline{\xi}_r(\pi;\gamma_0)$, $A_n = \Pi_r(v_{n,2})$, $A_0 = \Pi_{r,0}$, $\widehat{h}_n = \widetilde{\pi}_n$, and $\widehat{h} = \pi_r^*(\gamma_0)$. (The minus signs in $\mathbb{M}_n(h)$ and $\mathbb{M}(h)$ convert the minimization problem to a maximization problem.) The conditions of Lemma 9.10 hold because (i) $a_n^2(Q_n(\widetilde{\psi}_n(\pi),\pi) - Q_{0,n}) \Rightarrow \overline{\xi}_r(\cdot;\gamma_0)$ by part (d); (ii) $\overline{\xi}_r(\pi;\gamma_0)$ is continuous on Π a.s. by Assumptions C3–C5, RQ1(i), and RQ1(ii); (iii) $d_H(\Pi_r(v_{n,2}),\Pi_{r,0}) \to 0$ by Assumption RQ1(iv); (iv) $\pi_r^*(\gamma_0)$ satisfies the condition on $\widehat{h}$ using Comment (i) to Lemma 9.10 because $\pi_r^*(\gamma_0)$ uniquely maximizes $-\overline{\xi}_r(\pi;\gamma_0)$ over $\Pi_{r,0}$ by Assumption C6r when $\|b\| < \infty$ and by Assumption C7 when $\|b\| = \infty$, $\Pi_{r,0}$ is compact by the compactness of Π using Assumption B1(iii) and the continuity of $r_2(\pi)$ on Π using Assumption RQ1(i), and $\overline{\xi}_r(\pi;\gamma_0)$ is continuous on $\Pi_{r,0}$ a.s.; and (v) $\widetilde{\pi}_n$ satisfies the conditions on $\widehat{h}_n$ because $\widetilde{\pi}_n$ maximizes $-a_n^2(Q_n(\widetilde{\psi}_n(\pi),\pi) - Q_{0,n})$ over $\pi \in \Pi_r(v_{n,2})$ up to $o(n^{-1})$ by (9.62) and (9.64). The result of Lemma 9.10 is $\widetilde{\pi}_n \to_d \pi_r^*(\gamma_0)$.

Using $\widetilde{\pi}_n \to_d \pi_r^*(\gamma_0)$, we complete the proof of part (f). By (9.63) and (9.64), $a_n(\widetilde{\psi}_n - \psi_n) = a_n(\widetilde{\psi}_n(\widetilde{\pi}_n) - \psi_n)$. We have (i) $(a_n(\widetilde{\psi}_n(\cdot) - \psi_n), \widetilde{\pi}_n) \Rightarrow (\widetilde{\lambda}(\cdot), \pi_r^*(\gamma_0))$ as processes on Π by part (c) and $\widetilde{\pi}_n \to_d \pi_r^*(\gamma_0)$, (ii) $\widetilde{\lambda}(\pi) = P_\psi^\perp(\pi;\gamma_0)\tau(\pi;\gamma_0)$ by (9.76), and (iii) $P_\psi^\perp(\pi;\gamma_0)\tau(\pi;\gamma_0)$ is a continuous func-

tion of π on Π a.s. by Assumptions RQ1(i) and C3–C5. Hence, by the CMT, $a_n(\widetilde{\psi}_n(\widetilde{\pi}_n) - \psi_n) \to_d \tau(\pi_r^*(\gamma_0); \gamma_0)$ and the convergence is joint with $\widetilde{\pi}_n \to_d \pi_r^*(\gamma_0)$. This completes the proof of part (f).

The first result of part (g) holds because

$$\begin{aligned}(9.96)\qquad \tau(\pi_0; \gamma_0) &= -H^{-1}(\pi_0; \gamma_0)K(\pi_0; \gamma_0)\omega_0 - \big(\omega_0, 0_{d_\zeta}\big) \\ &= S'_\beta \omega_0 - \big(\omega_0, 0_{d_\zeta}\big) = 0,\end{aligned}$$

where the second equality holds by Lemma 9.3, which employs Assumption C8.

The second result of part (g) holds because (i) when $\|b\| = \infty$, $\pi_r^*(\gamma_0)$ minimizes

$$\begin{aligned}(9.97)\qquad \overline{\xi}_r(\pi; \gamma_0) &= \eta(\pi; \gamma_0, \omega_0) \\ &\quad + \frac{1}{2}\tau(\pi; \gamma_0)' P_\psi(\pi; \gamma_0)' H(\pi; \gamma_0) P_\psi(\pi; \gamma_0)\tau(\pi; \gamma_0)\end{aligned}$$

over $\Pi_{r,0}$ by (9.77), (ii) the first summand on the r.h.s. of (9.97) is uniquely minimized over $\Pi_{r,0}$ by π_0 by Assumption C7, and (iii) the second summand on the r.h.s. of (9.97) is minimized over $\Pi_{r,0}$ by π_0 by the first result of part (g) and the positive semidefiniteness of $P_\psi(\pi; \gamma_0)' H(\pi; \gamma_0) P_\psi(\pi; \gamma_0)$.

The third and fourth results of part (g) hold by part (f) and the first two results of part (g). *Q.E.D.*

PROOF OF LEMMA 9.10: The proof is a variation of the proof of Lemma 3.2.1 of van der Vaart and Wellner (1996, p. 286). First, by the extended CMT (see van der Vaart and Wellner (1996, Theorem 1.11.1, p. 67)), we have

$$(9.98)\qquad \sup_{h\in F\cap A_n} \mathbb{M}_n(h) - \sup_{h\in A_n} \mathbb{M}_n(h) \to_d \sup_{h\in F\cap A_0} \mathbb{M}(h) - \sup_{h\in A_0} \mathbb{M}(h).$$

The verification of the condition required by the extended CMT, that $x_n \to x$ implies $g_n(x_n) \to g(x)$, is essentially the same as that given in the paragraph containing (9.95). In the present case, $g_n(x) = \sup_{h\in F\cap A_n} x(h) - \sup_{h\in A_n} x(h)$, where x is a real-valued function on H.

Now, for all closed sets $F \subset H$,

$$\begin{aligned}(9.99)\qquad &\limsup_{n\to\infty} P^*(\widehat{h}_n \in F) \\ &\le \limsup_{n\to\infty} P^*\Big(\sup_{h\in F\cap A_n} \mathbb{M}_n(h) \ge \sup_{h\in A_n} \mathbb{M}_n(h) + o_p(1)\Big) \\ &\le P\Big(\sup_{h\in F\cap A_0} \mathbb{M}(h) \ge \sup_{h\in A_0} \mathbb{M}(h)\Big) \\ &\le P\Big(\sup_{h\in F\cap A_0} \mathbb{M}(h) \ge \sup_{h\in F^c\cap A_0} \mathbb{M}(h)\Big) \le P(\widehat{h} \in F),\end{aligned}$$

where P^* denotes outer probability, the first inequality holds by the definition of $\widehat{h}_n$, the second inequality holds by (9.98) and the portmanteau theorem (see Theorem 1.3.4 of van der Vaart and Wellner (1996, p. 18)), the third inequality holds because $F^c \cap A_0 \subset A_0$, and the last inequality holds by the argument in the following paragraph. Equation (9.99) and the portmanteau theorem give the result that $\widehat{h}_n \to_d \widehat{h}$.

Suppose $\widehat{h} \in F^c$. Then, by the assumption on $\widehat{h}$,

$$\mathbb{M}(\widehat{h}) > \sup_{h \notin F^c, h \in A_0} \mathbb{M}(h) = \sup_{h \in F \cap A_0} \mathbb{M}(h) \tag{9.100}$$

because F^c is open. Thus, $\widehat{h} \in F^c$ implies that

$$\sup_{h \in F^c \cap A_0} \mathbb{M}(h) > \sup_{h \in F \cap A_0} \mathbb{M}(h). \tag{9.101}$$

The contrapositive is $\sup_{h \in F^c \cap A_0} \mathbb{M}(h) \leq \sup_{h \in F \cap A_0} \mathbb{M}(h)$ implies $\widehat{h} \in F$, which verifies the last inequality in (9.99). *Q.E.D.*

9.4.3. *Distant From* $\beta = 0$ *Case*

Next, we provide results under sequences $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$. We prove Theorem 4.3. We also state and prove results concerning the asymptotic distribution of the restricted estimator $\widetilde{\theta}_n$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$.

Let $P_\theta^\perp(\gamma_0)$ denote a $d_\theta \times d_\theta$ oblique projection matrix that projects onto the orthogonal complement of the space spanned by the rows of $r_\theta(\theta_0)$:

$$P_\theta^\perp(\gamma_0) = I_{d_\theta} - P_\theta(\gamma_0), \tag{9.102}$$

where $P_\theta(\gamma_0)$ is defined in (4.14).

The following theorem shows that the normalized restricted criterion function, $n(Q_n(\widetilde{\theta}_n) - Q_n(\theta_n))$, converges in distribution under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ to $\xi_r^*(\gamma_0)$ and the QLR statistic converges in distribution to $\lambda_{\mathrm{QLR}}(\gamma_0)/s(\gamma_0)$, which are defined by

$$\begin{aligned} \xi_r^*(\gamma_0) &= \xi^*(\gamma_0) + \frac{1}{2}\lambda_{\mathrm{QLR}}(\gamma_0) \\ &= -\frac{1}{2}G^*(\gamma_0)'J^{-1}(\gamma_0)P_\theta^\perp(\gamma_0)'J(\gamma_0)P_\theta^\perp(\gamma_0)J^{-1}(\gamma_0)G^*(\gamma_0), \quad \text{where} \\ \xi^*(\gamma_0) &= -\frac{1}{2}G^*(\gamma_0)'J^{-1}(\gamma_0)G^*(\gamma_0), \\ \lambda_{\mathrm{QLR}}(\gamma_0) &= G^*(\gamma_0)'J^{-1}(\gamma_0)P_\theta(\gamma_0)'J(\gamma_0)P_\theta(\gamma_0)J^{-1}(\gamma_0)G^*(\gamma_0), \end{aligned} \tag{9.103}$$

where $J(\gamma_0)$ and $G^*(\gamma_0)$ are defined in Assumptions D2 and D3. Note that the normalized *unrestricted* criterion function, $n(Q_n(\widehat{\theta}_n) - Q_n(\theta_n))$, converges in distribution to $\xi^*(\gamma_0)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ by Theorem 3.2(b).

The following theorem also shows that the normalized restricted estimator, $n^{1/2}B(\beta_n)(\widetilde{\theta}_n - \theta_n)$, is asymptotically normal under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$.

THEOREM 9.3: *Suppose Assumptions* A, B1–B3, C1–C5, C7, C8, D1–D3, *and* RQ1 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, *the following statements hold*:
(a) $n(Q_n(\widetilde{\theta}_n) - Q_n(\theta_n)) \to_d \xi_r^*(\gamma_0)$.
(b) $\mathrm{QLR}_n \to_d \lambda_{\mathrm{QLR}}(\gamma_0)/s(\gamma_0)$, *provided Assumption* RQ3 *also holds*.
(c)

$$n^{1/2}B(\beta_n)(\widetilde{\theta}_n - \theta_n) \to_d -P_\theta^\perp(\gamma_0)J^{-1}(\gamma_0)G^*(\gamma_0) \sim N\big(0_{d_\theta}, P_\theta^\perp(\gamma_0)J^{-1}(\gamma_0)V(\gamma_0)J^{-1}(\gamma_0)P_\theta^\perp(\gamma_0)'\big).$$

COMMENT: Theorem 9.3(b) is the same as Theorem 4.3. Hence, to prove Theorem 4.3, it suffices to prove Theorem 9.3.

The proof of Theorem 9.3 uses the following preliminary results. The first result establishes the consistency of $\widetilde{\theta}_n$.

LEMMA 9.11: *Suppose Assumptions* A, B1–B3, C1–C5, C7, C8, *and* RQ1 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $\widetilde{\theta}_n - \theta_n \to_p 0$.

Next, by Theorem 9.2(g), we have the following "intermediate" rate of convergence result for $\widetilde{\psi}_n$ for sequences $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_0 = 0$ (which are also in $\Gamma(\gamma_0, 0, b)$ when $\|b\| = \infty$ and $\beta_n/\|\beta_n\| \to \omega_0$):

$$\|\beta_n\|^{-1}(\widetilde{\psi}_n - \psi_n) = o_p(1). \tag{9.104}$$

Using this intermediate rate result and Lemma 9.11, we obtain the sharp rate of convergence for $\widetilde{\theta}_n$ in the following lemma.

LEMMA 9.12: *Suppose Assumptions* A, B1–B3, C1–C5, C7, C8, D1–D3, *and* RQ1 *hold. Then* $n^{1/2}B(\beta_n)(\widetilde{\theta}_n - \theta_n) = O_p(1)$.

We now prove Theorem 9.3 using Lemma 9.12.

PROOF OF THEOREM 9.3: First, we rewrite the quadratic approximation in Assumption D1 as

$$\begin{aligned} &n(Q_n(\theta) - Q(\theta_n)) \\ &\quad = \big(n^{1/2}B^{-1}(\beta_n)DQ_n(\theta_n)\big)'n^{1/2}B(\beta_n)(\theta - \theta_n) \\ &\qquad + \frac{1}{2}\big(n^{1/2}B(\beta_n)(\theta - \theta_n)\big)'J_n n^{1/2}B(\beta_n)(\theta - \theta_n) + n^2R_n^*(\theta) \\ &\quad = -\frac{1}{2}Z_n^{*\prime}J_nZ_n^* + \frac{1}{2}q_n^*\big(n^{1/2}B(\beta_n)(\theta - \theta_n)\big) + n^2R_n^*(\theta), \end{aligned} \tag{9.105}$$

where

$$
\begin{aligned}
(9.106)\quad & Z_n^* = -n^{1/2} J_n^{-1} B^{-1}(\beta_n) D Q_n(\theta_n), \\
& J_n = B^{-1}(\beta_n) D^2 Q_n(\theta_n) B^{-1}(\beta_n), \\
& q_n^*(\lambda) = (\lambda - Z_n^*)' J_n (\lambda - Z_n^*).
\end{aligned}
$$

Now, the proof of the theorem is analogous to the proof of Theorem 9.1 using (9.105) in place of (9.65). The proof of Theorem 9.1 uses Lemma 9.7, Lemma 9.9, and Theorem 9.2. The main changes to the proof of Theorem 9.1 and the accompanying lemmas and theorem are the following:

(i) The dependence of various quantities on π is deleted.

(ii) The quantities $Z_n(\pi)$, $\tau_n(\pi; \gamma_n)$, $D_{\psi\psi,n}(\pi)$, $q_n(\lambda, \pi)$, and $a_n(\widetilde{\psi}_n(\pi) - \psi_n)$ are replaced by Z_n^*, Z_n^*, J_n, $q_n^*(\lambda)$, and $n^{1/2} B(\beta_n)(\widetilde{\theta}_n - \theta_n)$, respectively.[74]

(iii) The limit quantities $Z(\pi; \gamma_0)$, $\tau(\pi; \gamma_0)$, $H(\pi; \gamma_0)$, $q(\lambda, \pi)$, and $\overline{\xi}_r(\pi; \gamma_0)$ are replaced by Z^*, Z^*, $J(\gamma_0)$, $q^*(\lambda)$, and $\xi_r^*(\gamma_0)$, respectively, where

$$
(9.107)\quad Z^* = -J^{-1}(\gamma_0) G^*(\gamma_0) \quad \text{and} \quad q^*(\lambda) = (\lambda - Z^*)' J(\gamma_0)(\lambda - Z^*).
$$

(iv) The normalized parameter space $a_n(\Psi_r(\pi, v_{n,1}) - \psi_n)$ is replaced by $n^{1/2} B(\beta_n)(\Theta_r(v_n) - \theta_n)$, where

$$
(9.108)\quad \Theta_r(v) = \{\theta = (\psi, \pi) \in \Theta, r_1(\psi) = v_1, \ \& \ r_2(\pi) = v_2\} \quad \text{for} \quad v = (v_1, v_2).
$$

(v) Lemma 9.12 is employed in place of Lemma 9.6.

(vi) The quantity $\widetilde{\psi}_{n,q}(\pi)$ is replaced by $\widetilde{\theta}_{n,q}$, where $\widetilde{\theta}_{n,q} \in \Theta_r(v_n)$ is defined to satisfy

$$
(9.109)\quad q_n^*\big(n^{1/2} B(\beta_n)(\widetilde{\theta}_{n,q} - \theta_n)\big) = \inf_{\theta \in \Theta_r(v_n)} q_n^*\big(n^{1/2} B(\beta_n)(\theta - \theta_n)\big) + o_p(1).
$$

(vii) The definition of Λ is changed to

$$
(9.110)\quad \Lambda = \{\lambda \in R^{d_\theta} : r_\theta(\theta_0)\lambda = 0\}.
$$

(viii) The quantities $P_\psi(\pi; \gamma_0)$, $P_\psi^\perp(\pi; \gamma_0)$, and $\widetilde{\lambda}(\pi)$ are replaced by $P_\theta(\gamma_0)$, $P_\theta^\perp(\gamma_0)$, and $\widetilde{\lambda}$, respectively, where $\widetilde{\lambda} \in \Lambda$ is defined to minimize $q^*(\lambda)$ over $\lambda \in \Lambda$ and

$$
(9.111)\quad \widetilde{\lambda} = P_\theta^\perp(\gamma_0) Z^* = -P_\theta^\perp(\gamma_0) J^{-1}(\gamma_0) G^*(\gamma_0),
$$

[74]The quantities $Z_n(\pi)$ and $\tau_n(\pi; \gamma_n)$ differ by the amount $a_n(\gamma_n)(\psi_{0,n} - \psi_n)$ because the quadratic expansion in Assumption C1 is around $\psi_{0,n}$, rather than the true value ψ_n. In contrast, the quadratic expansion in Assumption D1 is around the true value θ_n. In consequence, the same quantity Z_n^* replaces both $Z_n(\pi)$ and $\tau_n(\pi; \gamma_n)$ in the proof of Theorem 9.3.

where the closed form expression for $\widetilde{\lambda}$ is as in Andrews (1999, p. 1361).

With these changes, the proof of Theorem 9.1 yields the proof of the results stated in Theorem 9.3. *Q.E.D.*

PROOF OF LEMMA 9.11: When $\beta_0 = 0$, $\widetilde{\pi}_n - \pi_n \to_p 0$ by Theorem 9.2(g) because sequences $\{\gamma_n\}$ in $\Gamma(\gamma_0, 0, b)$ with $\|b\| = \infty$ and $\beta_n/\|\beta_n\| \to \omega_0$ are in $\Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_0 = 0$. When $\beta_0 = 0$, $\widetilde{\psi}_n - \psi_n \to_p 0$ because $\|\widetilde{\psi}_n - \psi_n\| = \|\widetilde{\psi}_n(\widetilde{\pi}_n) - \psi_n\| \le \sup_{\pi \in \Pi} \|\widetilde{\psi}_n(\pi) - \psi_n\| = o_p(1)$ by Lemma 9.5(a).

When $\beta_0 \neq 0$, $\widetilde{\theta}_n \to_p \theta_0$ holds by an argument analogous to that given in the proof of Lemma 3.1(a) with $\widetilde{\theta}_n$, 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) in place of Assumption B3(ii), except that (9.3) needs to be altered. An alteration is needed because θ_0 does not necessarily satisfy the restrictions $r(\theta_0) = v_n$ $(= r(\theta_n))$, which invalidate the fourth inequality in (9.3). However, the fourth inequality holds with $Q_n(\theta_n; \gamma_0)$ in place of $Q_n(\psi_0, \pi; \gamma_0)$ in the second summand on the right-hand side of the fourth inequality, because the true value θ_n satisfies the restriction $r(\theta_n) = v_n$. With this change, the fifth inequality in (9.3) has the additional term $|Q(\theta_n; \gamma_0) - Q(\theta_0; \gamma_0)|$ on the r.h.s., which is $o(1)$ by Assumption RQ1(vi). This completes the proof. *Q.E.D.*

PROOF OF LEMMA 9.12: The proof is the same as the proof that $n^{1/2}B(\beta_n) \times (\widehat{\theta}_n - \theta_n) = O_p(1)$, which is given at the beginning of the proof of Theorem 3.2. In the proof, (9.104) is used in place of Lemma 3.4, and $\widetilde{\pi}_n - \pi_n = o_p(1)$ and $\widetilde{\psi}_n - \psi_n = o_p(1)$ by Lemma 9.11 are used in place of $\widehat{\pi}_n - \pi_n = o_p(1)$ and $\widehat{\psi}_n - \psi_n = o_p(1)$ by Lemma 3.3. The key inequality in (9.34) holds in the present case because the true value θ_n satisfies the restrictions. *Q.E.D.*

9.4.4. *QLR Statistic With Restrictions on $\pi + \beta$*

Here we provide more details concerning the claim in Comment (iv) following Theorem 4.2 that the QLR statistic has the same asymptotic distribution for restrictions of the form $r(\theta) = (r_1(\psi), \pi + \beta)$ as for restrictions of the form $r(\theta) = (r_1(\psi), \pi)$.

Roughly speaking, the reason the comment holds is as follows. First, suppose $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$. The restrictions do not effect the second component of the QLR statistic $Q_n(\widehat{\theta}_n)$ and we already have its asymptotic distribution after suitable normalization, so it suffices to focus on the first component $Q_n(\widetilde{\theta}_n)$. The limit set $\Pi_{r,0}$ is the same whether the restrictions are on $\pi + \beta$ or π because $\beta_n \to 0$. This leads to the same asymptotic distribution of $n(Q_n(\widetilde{\theta}_n) - Q_{0,n})$ for these two restrictions. Next, under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, weak identification is not an issue and so the QLR statistic has a $\chi^2_{d_r}$ asymptotic distribution whether $\pi + \beta$ or π is restricted (as in (4.15)).

Now we provide more details. As just stated, it suffices to focus on the normalized first component $n(Q_n(\widetilde{\theta}_n) - Q_{0,n})$. We consider a reparametrization

of the model/criterion function. The original model based on (β, ζ, π) can be reparametrized to depend on (β, ζ, π_1), where $\pi_1 = \pi + \beta$. The results of Theorem 9.1(a) can be applied to the reparametrized model with parameters (β, ζ, π_1). Denote the criterion function for the reparametrized model by $\overline{Q}_n(\beta, \zeta, \pi_1) = Q_n(\beta, \zeta, \pi_1 - \beta)$.

First, consider the asymptotic distribution $n(\overline{Q}_n(\widetilde{\theta}_n) - Q_{0,n})$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$ with the restrictions based on $r(\theta) = (r_1(\psi), \pi + \beta)$. Given these restrictions, for the results of Theorem 9.1(a), we do not need a quadratic expansion to hold for all π_1 in some set Π_1 that is analogous to Π in Assumption C1. Rather, we just need a version of Assumption C1 to hold for $\overline{Q}_n(\beta, \zeta, \pi_1)$ when $\pi_1 = \pi_{1,n} = \pi_n + \beta_n$, that is, for $\overline{Q}_n(\beta, \zeta, \pi_{1,n})$. This is obtained for the reparametrized criterion function when Assumptions C1–C4 hold for the original criterion function:

$$
\begin{aligned}
(9.112)\quad & \overline{Q}_n(\beta, \zeta, \pi_{1,n}) \\
&= Q_n(\beta, \zeta, \pi_{1,n} - \beta) \\
&= Q_n(0, \zeta, \pi_{1,n} - \beta) + D_\psi Q_n(\psi_{0,n}, \pi_{1,n} - \beta)'(\psi - \psi_{0,n}) \\
&\quad + \frac{1}{2}(\psi - \psi_{0,n})' D_{\psi\psi} Q_n(\psi_{0,n}, \pi_{1,n} - \beta)(\psi - \psi_{0,n}) \\
&\quad + R_n(\psi, \pi_{1,n} - \beta) \\
&= Q_n(0, \zeta, \pi_{1,n}) + D_\psi Q_n(\psi_{0,n}, \pi_{1,n})'(\psi - \psi_{0,n}) \\
&\quad + \frac{1}{2}(\psi - \psi_{0,n})' D_{\psi\psi} Q_n(\psi_{0,n}, \pi_{1,n})(\psi - \psi_{0,n}) \\
&\quad + R_n(\psi, \pi_{1,n} - \beta) + R_{2,n}(\psi),
\end{aligned}
$$

where $R_{2,n}(\psi)$ is defined implicitly by the third equality, the first equality holds by the definition of $\overline{Q}_n(\beta, \zeta, \pi_1)$, the second equality holds by Assumption C1 for $Q_n(\theta)$, and the third equality uses the fact that $Q_n(0, \zeta, \pi)$ does not depend on π. The additional remainder term $R_{2,n}(\psi)$ satisfies Assumption C1(ii) with $R_{2,n}(\psi)$ in place of $R_n(\psi, \pi)$, using Assumptions C2–C4 for $Q_n(\theta)$. This relies on the fact that the true values $\theta_n = (\beta_n, \zeta_n, \pi_n) \in \Theta^* \subset \text{int}(\Theta)$ by Assumption B1(i). In consequence, for some set Π^*, we have $\pi_n \in \Pi^* \subset \text{int}(\Pi)$ for all n and, hence, $\pi_{1,n} - \beta$ $(= \pi_n + \beta_n - \beta)$ is in Π for all β with $\|\beta\| \leq \delta_n$ for all n large, where $\delta_n \to 0$.

Similarly, under the given restrictions, for the results of Theorem 9.1(a) to hold for $\overline{Q}_n(\beta, \zeta, \pi_1)$, Assumptions B1–B3 and C2–C5 for $\overline{Q}_n(\beta, \zeta, \pi_1)$ do not need to hold for all $\pi_1 - \beta_n \in \Pi$. It suffices for them to hold with $\pi_1 \in \Pi^* + \beta_n \subset \Pi$, which they do by Assumptions B1–B3 and C2–C5 for $Q_n(\theta)$. Assumption A clearly holds for $\overline{Q}_n(\beta, \zeta, \pi_1)$. This completes the verification of the required assumptions for $\overline{Q}_n(\beta, \zeta, \pi_1)$. In turn, this completes the proof for sequences $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$.

Next, suppose $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$. We apply the results of Theorem 9.3(a) to the reparametrized model with criterion function $\overline{Q}_n(\beta, \zeta, \pi_1)$. In addition to Assumptions C1–C5, we suppose Assumptions D1–D3 and C8 hold for the original criterion function $Q_n(\theta)$. Then Assumption D1 holds for $\overline{Q}_n(\beta, \zeta, \pi_1)$ by the following calculation. For notational simplicity, suppose no parameter ζ appears. For $\pi_1 = \pi + \beta$ and $\pi_{1,n} = \pi_n + \beta_n$, we have

$$
\begin{aligned}
(9.113)\quad & \overline{Q}_n(\beta, \pi_1) \\
&= Q_n(\beta, \pi) \\
&= Q_n(\theta_n) + DQ_n(\theta_n)'(\theta - \theta_n) + \frac{1}{2}(\theta - \theta_n)D^2Q_n(\theta_n)(\theta - \theta_n) + R_n^*(\theta) \\
&= \overline{Q}_n(\beta_n, \pi_{1,n}) + \begin{pmatrix} \partial_\beta Q_n(\theta_n) \\ \partial_\pi Q_n(\theta_n) \end{pmatrix}' \begin{pmatrix} \beta - \beta_n \\ \pi_1 - \pi_{1,n} - (\beta - \beta_n) \end{pmatrix} \\
&\quad + \frac{1}{2}\begin{pmatrix} \beta - \beta_n \\ \pi_1 - \pi_{1,n} - (\beta - \beta_n) \end{pmatrix}' \begin{bmatrix} \partial_{\beta\beta} Q_n(\theta_n) & \partial_{\beta\pi} Q_n(\theta_n) \\ \partial_{\pi\beta} Q_n(\theta_n) & \partial_{\pi\pi} Q_n(\theta_n) \end{bmatrix} \\
&\quad \times \begin{pmatrix} \beta - \beta_n \\ \pi_1 - \pi_{1,n} - (\beta - \beta_n) \end{pmatrix} + R_n^*(\theta) \\
&= \overline{Q}_n(\beta_n, \pi_{1,n}) + \begin{pmatrix} \partial_\beta Q_n(\theta_n) - \partial_\pi Q_n(\theta_n) \\ \partial_\pi Q_n(\theta_n) \end{pmatrix}' \begin{pmatrix} \beta - \beta_n \\ \pi_1 - \pi_{1,n} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \beta - \beta_n \\ \pi_1 - \pi_{1,n} \end{pmatrix}' \\
&\quad \times \begin{bmatrix} \partial_{\beta\beta} Q_n(\theta_n) - 2\partial_{\beta\pi} Q_n(\theta_n) + \partial_{\pi\pi} Q_n(\theta_n) & \partial_{\beta\pi} Q_n(\theta_n) - \partial_{\pi\pi} Q_n(\theta_n) \\ \partial_{\pi\beta} Q_n(\theta_n) - \partial_{\pi\pi} Q_n(\theta_n) & \partial_{\pi\pi} Q_n(\theta_n) \end{bmatrix} \\
&\quad \times \begin{pmatrix} \beta - \beta_n \\ \pi_1 - \pi_{1,n} \end{pmatrix} + R_n^*(\theta),
\end{aligned}
$$

where the first equality holds by definition, the second equality holds by Assumption D1 for $Q_n(\theta)$, the quantities $\partial_\beta Q_n(\theta_n)$, $\partial_\pi Q_n(\theta_n)$, $\partial_{\beta\beta} Q_n(\theta_n), \ldots,$ on the r.h.s. of the third equality are subvectors and submatrices of $DQ_n(\theta_n)$ and $D^2Q_n(\theta_n)$ by definition, and the fourth equality holds by algebra. Equation (9.113) establishes Assumption D1 for $\overline{Q}_n(\beta, \pi_1)$ because the properties of $R_n^*(\theta)$ in Assumption D2(ii) for $Q_n(\theta)$ yield the appropriate properties for the remainder $R_n^*(\theta) = R^*(\beta, \pi_1 - \beta)$ for $\overline{Q}_n(\beta, \pi_1)$.

Assumptions D2 and D3 for $Q_n(\theta)$ imply Assumptions D2 and D3 for $\overline{Q}_n(\beta, \pi_1)$ with the limit quantities $J(\gamma_0)$ and $V(\gamma_0)$ changed to correspond to the changes in (9.113) from $DQ_n(\theta_n)$ and $D^2Q_n(\theta_n)$ to

$$
(9.114)\quad \begin{pmatrix} \partial_\beta Q_n(\theta_n) - \partial_\pi Q_n(\theta_n) \\ \partial_\pi Q_n(\theta_n) \end{pmatrix} \quad \text{and}
$$

$$
\begin{bmatrix} \partial_{\beta\beta} Q_n(\theta_n) - 2\partial_{\beta\pi} Q_n(\theta_n) + \partial_{\pi\pi} Q_n(\theta_n) & \partial_{\beta\pi} Q_n(\theta_n) - \partial_{\pi\pi} Q_n(\theta_n) \\ \partial_{\pi\beta} Q_n(\theta_n) - \partial_{\pi\pi} Q_n(\theta_n) & \partial_{\pi\pi} Q_n(\theta_n) \end{bmatrix},
$$

respectively. Assumption C7 for $\overline{Q}_n(\beta, \pi_1)$ is not needed to obtain the result in Theorem 9.1(a) for the restrictions given because there is a unique value of π_1 that satisfies the restrictions. Assumption C8 for $\overline{Q}_n(\beta, \pi_1)$ is implied by Assumption C8 for $Q_n(\beta, \pi)$. This completes the verification of the assumptions needed for $\overline{Q}_n(\beta, \pi_1)$ in Theorem 9.1(a). Combining this result with the asymptotic distribution of $n(Q_n(\widehat{\theta}_n) - Q_{0,n})$, which does not depend on the form of the restrictions, yields the result of Theorem 9.1(b), which is the same as the result in Theorem 4.2. This result, combined with (4.15) (using the assumption that Assumption RQ2 holds), yields a $\chi^2_{d_r}$ distribution for the QLR statistic under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ when $r(\theta) = (r_1(\psi), \pi + \beta)$, just as it does when $r(\theta) = (r_1(\psi), \pi)$.

This completes the proof of the assertion in Comment (iv) to Theorem 4.2.

9.5. *Proofs of Asymptotic Size Results*

PROOF OF THEOREM 4.4: We only prove the asymptotic size result of Theorem 4.4 for the symmetric two-sided CI, which is based on $|T_n|$. The proofs for the one-sided CI's and the QLR CS, which are based on T_n, $-T_n$, and QLR_n, respectively, are analogous. For the QLR CS, one uses Theorems 9.1 and 9.3 in place of Theorem 4.1 in the proof below.

By definition, $\text{CP}_n(\gamma_n) = P_{\gamma_n}(|T_n| \le z_{1-\alpha/2})$. By Theorem 4.1 and Assumption V3, $\text{CP}_n(\gamma_n) \to P(|T(h)| \le z_{1-\alpha/2})$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$ and $\text{CP}_n(\lambda_n) \to P(|Z| \le z_{1-\alpha/2}) = 1 - \alpha$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$. This implies Assumption ACP(i)–(iii). Assumption ACP(iv) holds by Assumption B2(ii). Given this, the desired result holds by Lemma 2.1. *Q.E.D.*

PROOF OF THEOREM 5.1: The proof of Theorem 5.1(a)(i) for the LF critical value is the same as that of Theorem 4.4 but with $c^{\text{LF}}_{\mathcal{T},1-\alpha}$ $(= \max\{\sup_{h\in H} c_{\mathcal{T},1-\alpha}(h), c_{\mathcal{T},1-\alpha}(\infty)\})$ for $\mathcal{T}_n = |T_n|$, T_n, $-T_n$, and QLR_n in place of $z_{1-\alpha/2}$, $z_{1-\alpha}$, $z_{1-\alpha}$, and $\chi^2_{d_r,1-\alpha}$, respectively, using Assumption LF(i) in place of Assumption V3. For the case of $\mathcal{T}_n = |T_n|$, this proof delivers

$$\text{AsySz} = \min\left\{\inf_{h\in H} P\big(|T(h)| \le c^{\text{LF}}_{|t|,1-\alpha}\big), P\big(|Z| \le c^{\text{LF}}_{|t|,1-\alpha}\big)\right\}, \tag{9.115}$$

where $Z \sim N(0,1)$. The r.h.s. of (9.115) is greater than or equal to $1-\alpha$ because (i) $P(|T(h)| \le c^{\text{LF}}_{|t|,1-\alpha}) \ge P(|T(h)| \le c_{|t|,1-\alpha}(h)) \ge 1-\alpha$ $\forall h \in H$, where the second inequality holds by the definition of the quantile $c_{|t|,1-\alpha}(h)$, and (ii) $P(|Z| \le c^{\text{LF}}_{|t|,1-\alpha}) \ge P(|Z| \le z_{1-\alpha/2}) = 1-\alpha$. The r.h.s. of (9.115) is less than or equal to $1-\alpha$ because if $c^{\text{LF}}_{|t|,1-\alpha} = z_{1-\alpha/2}$, then $P(|Z| \le c^{\text{LF}}_{|t|,1-\alpha}) = 1-\alpha$ and if $c^{\text{LF}}_{|t|,1-\alpha} > z_{1-\alpha/2}$, then $P(|T(h_{\max})| \le c^{\text{LF}}_{|t|,1-\alpha}) = P(|T(h_{\max})| \le c_{|t|,1-\alpha}(h_{\max})) = 1-\alpha$, where both equalities hold using Assumption LF. Hence, $\text{AsySz} = 1-\alpha$. The proofs for $\mathcal{T}_n = T_n$, $-T_n$, and QLR_n are analogous using Theorems 9.1

and 9.3 in place of Theorem 4.1 when considering QLR CS's. The assumptions are different for QLR CS's because of the latter change.

The proofs of Theorem 5.1(a)(ii) and (b)(ii) for the NI-LF critical value are the same as that just given for the LF critical value except that H, $c^{\mathrm{LF}}_{|t|,1-\alpha}$, $h_{\max}$, and Assumption LF are replaced by $H(v)$, $c^{\mathrm{LF}}_{|t|,1-\alpha}(v)$ $(= \max\{\sup_{h\in H(v)} c_{|t|,1-\alpha}(h), z_{1-\alpha/2}\})$, $h_{\max}(v)$, and Assumption NI-LF, respectively, for $v \in V_r$ and the r.h.s. of (9.115) has $\inf_{v\in V_r}$ added.

Theorem 5.1(a)(iii) is proved by verifying Assumption ACP and invoking Lemma 2.1. Consider the case where $\mathcal{T}_n = |T_n|$. First, we show $\widetilde{c}_{|t|,1-\alpha,n} \to_p c^{\mathrm{LF}}_{|t|,1-\alpha}$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$. By the construction of $\widetilde{c}_{|t|,1-\alpha,n}$, it suffices to show that $P_{\gamma_n}(A_n \le \kappa_n) \to 1$. This holds if $A_n = O_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$, because $\kappa_n \to \infty$ by Assumption K(i).

When β is a scalar, we have

$$
\begin{aligned}
(9.116)\quad A_n &= \big(n^{1/2}\widehat{\beta}_n' \widehat{\Sigma}^{-1}_{\beta\beta,n} n^{1/2}\widehat{\beta}_n/d_\beta\big)^{1/2} \\
&\to_d (\tau_\beta(\pi^*)' \Sigma^{-1}_{\beta\beta}(\pi^*;\gamma_0)\tau_\beta(\pi^*)/d_\beta)^{1/2},
\end{aligned}
$$

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 3.1(a) and Assumption V1. By Assumptions B1(iii) and V1(ii) and (iii), $\inf_{\pi\in\Pi} \Sigma_{\beta\beta}(\pi;\gamma_0) > 0$. Hence, $A_n = O_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$, as desired.

When β is a vector, (9.116) holds with $\Sigma_{\beta\beta}(\pi^*;\gamma_0)$ replaced by $\Sigma_{\beta\beta}(\pi^*, \omega^*(\pi^*);\gamma_0,\omega_0)$ by Theorem 3.1(a), Assumption V1, and the joint convergence $(n^{1/2}\widehat{\beta}_n, \widehat{\pi}_n, \widehat{\omega}_n) \to_d (\tau_\beta(\pi^*), \pi^*, \omega^*(\pi^*))$. By Assumptions B1(iii) and V1(ii) and (iii), $\inf_{\pi\in\Pi,\|\omega\|=1} \lambda_{\min}(\Sigma_{\beta\beta}(\pi,\omega;\gamma_0,\omega_0)) > 0$. Hence, $A_n = O_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$, as desired.

Using Theorem 4.1(a) and (b), $\widetilde{c}_{|t|,1-\alpha,n} \to_p c^{\mathrm{LF}}_{|t|,1-\alpha}$, and Assumption V3, we obtain $\mathrm{CP}_n(\gamma_n) = P_{\gamma_n}(|T_n| \le \widetilde{c}_{|t|,1-\alpha,n}) \to P(|T(h)| \le c^{\mathrm{LF}}_{|t|,1-\alpha})$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$. Hence, Assumption ACP(i) holds with $\mathrm{CP}(h) = P(|T(h)| \le c^{\mathrm{LF}}_{|t|,1-\alpha})$.

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

$$
(9.117)\quad P_{\gamma_n}\big(|T_n| \le z_{1-\alpha/2}\big) \le P_{\gamma_n}\big(|T_n| \le \widetilde{c}_{|t|,1-\alpha,n}\big) \le P_{\gamma_n}\big(|T_n| \le c^{\mathrm{LF}}_{|t|,1-\alpha}\big).
$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,

$$
\begin{aligned}
(9.118)\quad & P_{\gamma_n}\big(|T_n| \le z_{1-\alpha/2}\big) \to P\big(|Z| \le z_{1-\alpha/2}\big) = 1-\alpha, \\
& P_{\gamma_n}\big(|T_n| \le c^{\mathrm{LF}}_{|t|,1-\alpha}\big) \to P\big(|Z| \le c^{\mathrm{LF}}_{|t|,1-\alpha}\big) \ge 1-\alpha.
\end{aligned}
$$

By (9.117) and (9.118), Assumption ACP(ii) holds with $\mathrm{CP}_\infty = 1-\alpha$.

Next, we verify Assumption ACP(iii) by showing $\widetilde{c}_{|t|,1-\alpha,n} \to_p z_{1-\alpha/2}$ 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

$$
(9.119)\quad \kappa_n^{-1} A_n = \big(n^{1/2}\kappa_n^{-1}\big)(\widehat{\beta}_n' \widehat{\Sigma}^{-1}_{\beta\beta,n}\widehat{\beta}_n/d_\beta)^{1/2} \to_p \infty,
$$

where the divergence to infinity holds because $n^{1/2}\kappa_n^{-1} \to \infty$ by Assumption K(ii), $\widehat{\beta}_n \to_p \beta_0 \neq 0$ by Lemma 3.1(b), $\widehat{\Sigma}_{\beta\beta,n} \to_p \Sigma_{\beta\beta}(\gamma_0)$ by Assumption V2, where $\Sigma_{\beta\beta}(\gamma_0)$ denotes the upper left $d_\beta \times d_\beta$ submatrix 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 4.1(c), $\widetilde{c}_{|t|,1-\alpha,n} \to_p z_{1-\alpha/2}$, and the continuity of the d.f. of Z, we obtain $\mathrm{CP}_n(\gamma_n) = P_{\gamma_n}(|T_n| \leq \widetilde{c}_{|t|,1-\alpha,n}) \to 1-\alpha$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_0 \neq 0$. This completes the verification of Assumption ACP(iii). Assumption ACP(iv) holds by Assumption B2(ii).

Applying Lemma 2.1, we conclude that the nominal $1-\alpha$ type 1 robust two-sided t CI has $\mathrm{AsySz} = 1-\alpha$. This completes the proof of Theorem 5.1(a)(iii) for $\mathcal{T}_n = |T_n|$. The proofs for one-sided t CI's and QLR CS's are analogous. Note that the use of Theorem 3.1(a) above can be replaced by Lemma 9.2(a), which shows that $n^{1/2}\widehat{\beta}_n = O_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$. In consequence, the proof of Theorem 5.1(b)(iii) for QLR CS's requires Assumptions V1 and V2, but not C6. (The same is true for Theorem 5.1(b)(iv), but Theorem 5.1(b)(v) and (vi) require Assumptions V1, V2, and C6 because the asymptotic distribution of $n^{1/2}\widehat{\beta}_n$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ given in Theorem 3.1(a) is required.)

The proofs of Theorem 5.1(a)(iv) and 5.1(b)(iv) for the type 1 NI robust critical value are analogous to that just given for the type 1 robust critical value except that H, $c^{\mathrm{LF}}_{|t|,1-\alpha}$, and $\widetilde{c}_{|t|,1-\alpha,n}$ are replaced by $H(v)$, $c^{\mathrm{LF}}_{|t|,1-\alpha}(v)$, and $\widetilde{c}_{|t|,1-\alpha,n}(v)$, respectively, for $v \in V_r$.

The proof of Theorem 5.1(a)(v) for the type 2 robust critical value is proved by verifying Assumption ACP and invoking Lemma 2.1. Again, consider the case when $\mathcal{T}_n = |T_n|$. First, under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$, we have

$$(9.120) \quad \big(|T_n|, \widehat{c}_{|t|,1-\alpha,n}\big) \to_d \big(|T(h)|, \widehat{c}_{|t|,1-\alpha}(h)\big),$$

because (i) $T_n \to_d T(h)$ by Theorem 4.1, (ii) $A_n \to_d A(h)$ by (9.116), (iii) $\widehat{c}_{|t|,1-\alpha,n} \to_d \widehat{c}_{|t|,1-\alpha}(h)$ by the continuous mapping theorem using result (ii), (5.5), (8.3), 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 (9.120) and Assumption Rob2(i) imply that under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$,

$$(9.121) \quad P\big(|T_n| \leq \widehat{c}_{|t|,1-\alpha,n}\big) \to_d P\big(|T(h)| \leq \widehat{c}_{|t|,1-\alpha}(h)\big) \quad \forall h = (b, \gamma_0) \in H.$$

This verifies Assumption ACP(i) with $\mathrm{CP}(h) = P(|T(h)| \leq \widehat{c}_{|t|,1-\alpha}(h))$.

Second, under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, we have (i) $A_n \to_p \infty$ by Theorem 4.1(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) $s(A_n - \kappa) \to_p 0$ by results (i) and (ii) and the assumption that $s(x) \to 0$ as $x \to \infty$, and (iii) $\widehat{c}_{|t|,1-\alpha,n} \to_p c_{|t|,1-\alpha}(\infty) + \Delta_2 = z_{1-\alpha/2} + \Delta_2$ using result (ii)

and (5.5). Result (iii) and $|T_n| \to_d |Z|$ for $Z \sim N(0,1)$, which holds by Theorem 4.1(c), yield that under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,

$$(9.122) \quad P\big(|T_n| \le \widehat{c}_{|t|,1-\alpha,n}\big) \to_d P\big(|Z| \le z_{1-\alpha/2} + \Delta_2\big).$$

This verifies Assumption ACP(ii) and (iii) with $\mathrm{CP}_\infty = P(|Z| \le z_{1-\alpha/2} + \Delta_2)$.

Lemma 2.1 now gives

$$(9.123) \quad \mathrm{AsySz} = \min\Big\{\inf_{h\in H} P\big(|T(h)| \le \widehat{c}_{|t|,1-\alpha}(h)\big), P\big(|Z| \le z_{1-\alpha/2} + \Delta_2\big)\Big\}.$$

It remains to show that the right-hand side equals $1-\alpha$. We have

$$(9.124) \quad \mathrm{AsySz} = \min\Big\{\inf_{h\in H}(1-\mathrm{NRP}(\Delta_1, \Delta_2; h)), P\big(|Z| \le z_{1-\alpha/2}+\Delta_2\big)\Big\} \ge 1-\alpha,$$

where $\mathrm{NRP}(\Delta_1, \Delta_2; h)$ is defined in (5.7) with $\mathcal{T}(h) = |T(h)|$, the equality holds by (5.7) and (8.3) with $\mathcal{T}(h) = |T(h)|$ and (9.123), and the inequality holds by the definitions of Δ_1 and Δ_2 in (5.8), $P(|Z| \le z_{1-\alpha/2}) = 1-\alpha$, and $\Delta_2 \ge 0$.

If $\Delta_2 = 0$, then $P(|Z| \le z_{1-\alpha/2} + \Delta_2) = 1-\alpha$ and $\mathrm{AsySz} \le 1-\alpha$ by (9.124). Alternatively, if $\Delta_2 > 0$, we have

$$(9.125) \quad \mathrm{AsySz} \le 1 - \mathrm{NRP}(\Delta_1, \Delta_2; h^*) = 1-\alpha,$$

where the inequality holds using the equality in (9.124) and the equality holds by Assumption Rob2(ii). This completes the proof that $\mathrm{AsySz} = 1-\alpha$ in Theorem 5.1(a)(v) for the case $\mathcal{T}_n = |T_n|$. The proofs of Theorem 5.1(a)(v) and (b)(v) for the cases $\mathcal{T}_n = T_n - T_n$ and QLR_n are analogous.

The proofs of Theorem 5.1(a)(vi) and (b)(vi) are analogous to that of Theorem 5.1(a)(v) using Assumption NI-Rob2 in place of Assumption Rob2. *Q.E.D.*

9.6. *Proofs of Sufficient Conditions*

9.6.1. *Assumption B3*

PROOF OF LEMMA 8.1: Assumption B3*(i) and (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 standard 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 Assumption B3*(i) and (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 $\|\pi_1 - \pi_2\| < \delta_1$ imply 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

$$\left|Q(\psi^{**}(\pi_1), \pi_1; \gamma_0) - Q(\psi^*(\pi_2), \pi_2; \gamma_0)\right| < \varepsilon \tag{9.126}$$

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

$$Q(\psi^*(\pi_1), \pi_1; \gamma_0) \le Q(\psi^{**}(\pi_1), \pi_1; \gamma_0) < Q(\psi^*(\pi_2), \pi_2; \gamma_0) + \varepsilon \tag{9.127}$$

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 (9.126).

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. *Q.E.D.*

9.6.2. *Assumption C5*

PROOF OF LEMMA 8.2: We now verify Assumption C5. Without loss of generality, suppose $\beta \in R$. Let $\{\beta_k^* : k \ge 1\}$ 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 $E_{\gamma^*} m(W_i, \theta)$ w.r.t. β^* is

$$\begin{aligned}
&\lim_{k\to\infty} \frac{E_{\gamma_k^*} m(W_i, \theta) - E_{\gamma^*} m(W_i, \theta)}{\beta_k^* - \beta^*} \\
&\quad = \lim_{k\to\infty} \int_{\mathcal{W}} m(w, \theta) \frac{f_{W_i}(w; \gamma_k^*) - f_{W_i}(w; \gamma^*)}{\beta_k^* - \beta^*} \, d\mu(w) \\
&\quad = \int_{\mathcal{W}} m(w, \theta) \left(\lim_{k\to\infty} \frac{f_{W_i}(w; \gamma_k^*) - f_{W_i}(w; \gamma^*)}{\beta_k^* - \beta^*} \right) d\mu(w) \\
&\quad = \int_{\mathcal{W}} m(w, \theta) f_{\beta, W_i}(w; \gamma^*) \, d\mu(w),
\end{aligned} \tag{9.128}$$

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 $f_{W_i}(w;\gamma^*)$ w.r.t. β^*. The DCT holds in the second equality using

$$\frac{f_{W_i}(w;\gamma_k^*)-f_{W_i}(w;\gamma^*)}{\beta_k^*-\beta^*}=f_{\beta,W_i}(w;\widetilde{\gamma}_k(w)), \tag{9.129}$$

$$\int_{\mathcal{W}}\sup_{\theta\in\Theta}\|m(w,\theta)\|\cdot\sup_{\gamma\in N(\gamma^*,\varepsilon)}\left|f_{\beta,W_i}(w;\gamma)\right|d\mu(w)<\infty,$$

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

$$\begin{aligned}
&\sup_{\pi\in\Pi}|K_n(\psi_n,\pi;\widetilde{\gamma}_n)-K(\psi_0,\pi;\gamma_0)| \\
&\quad\le\int\sup_{\pi\in\Pi}\left|m(w,\psi_n,\pi)\left(n^{-1}\sum_{i=1}^n f_{\beta,W_i}(w;\widetilde{\gamma}_n)\right)\right. \\
&\qquad\left.-m(w,\psi_0,\pi)f_{\beta,W}(w;\gamma_0)\right|d\mu(w) \\
&\quad\le\int\sup_{\theta\in\Theta}|m(w,\theta)|\cdot\left|n^{-1}\sum_{i=1}^n f_{\beta,W_i}(w;\widetilde{\gamma}_n)-f_{\beta,W}(w;\gamma_0)\right|d\mu(w) \\
&\qquad+\int\sup_{\pi\in\Pi}|m(w,\psi_n,\pi)-m(w,\psi_0,\pi)|f_{\beta,W}(w;\gamma_0)\,d\mu(w),
\end{aligned} \tag{9.130}$$

where the first inequality is obvious and the second inequality holds by the triangle inequality. The fourth line of (9.130) converges to 0 by the DCT under Assumption C5*(ii), (iii), and (v) using $\widetilde{\gamma}_n\to\gamma_0$. The fifth line of (9.130) converges to 0 by Assumption C5*(iv) and (v). This yields Assumption C5(ii).

Assumption C5(iii) holds by the DCT using Assumption C5*(iv) and (v). *Q.E.D.*

9.6.3. *Assumption C6*

PROOF OF LEMMA 8.3: We block diagonalize $H(\pi;\gamma_0)$ using the $d_\psi\times d_\psi$ matrix $A(\pi)$ defined by

$$A(\pi)=\begin{bmatrix} I_{d_\beta} & -H_{12}(\pi)H_{22}^{-1} \\ 0_{d_\zeta\times d_\beta} & I_{d_\zeta}\end{bmatrix}. \tag{9.131}$$

Simple calculations yield

$$A(\pi)H(\pi;\gamma_0)A(\pi)' = \begin{bmatrix} H_{11}^*(\pi) & 0_{d_\beta \times d_\zeta} \\ 0_{d_\zeta \times d_\beta} & H_{22} \end{bmatrix}, \tag{9.132}$$

$$A(\pi)[G(\pi;\gamma_0) + K(\pi;\gamma_0)b] = \begin{bmatrix} G_1^*(\pi;\gamma_0) + K_1^*(\pi;\gamma_0)b \\ G_2 + K_2 b \end{bmatrix},$$

$$A(\pi)K(\pi;\gamma_0)\omega_0 = K_1^*(\pi;\gamma_0)\omega_0.$$

In consequence, we have

$$\begin{aligned} &\xi(\pi;\gamma_0,b) \\ &\quad = -\frac{1}{2}(G(\pi;\gamma_0) + K(\pi;\gamma_0)b)'A(\pi)'[A(\pi)H(\pi;\gamma_0)A(\pi)']^{-1} \\ &\qquad \times A(\pi)(G(\pi;\gamma_0) + K(\pi;\gamma_0)b) \\ &\quad = \xi_1(\pi;\gamma_0,b) + \xi_2(\gamma_0,b). \end{aligned} \tag{9.133}$$

Similarly, we have

$$\begin{aligned} &\eta(\pi;\gamma_0,\omega_0) \\ &\quad = -\frac{1}{2}\omega_0' K(\pi;\gamma_0)'A(\pi)'[A(\pi)H(\pi;\gamma_0)A(\pi)']^{-1}A(\pi)K(\pi;\gamma_0)\omega_0 \\ &\quad = \eta_1(\pi;\gamma_0,\omega_0) + \eta_2(\gamma_0,\omega_0), \end{aligned} \tag{9.134}$$

which completes the proof. *Q.E.D.*

Lemma 8.4 follows immediately from the following lemma, which is an extension of Lemma 2.6 of Kim and Pollard (1990).

LEMMA 9.13: *Let* $\{Z(t): t \in T\}$ *be a univariate Gaussian process with continuous sample paths, indexed by a* σ*-compact metric space* T*. If* $\mathrm{Var}(Z(s) - Z(t)) \neq 0$ *and* $\mathrm{Var}(Z(s) + Z(t)) \neq 0\ \forall s, t \in T$ *with* $s \neq t$*, then, with probability* 1*, no sample path of* $Z^2(\cdot)$ *can achieve its supremum at two distinct points of* T*.*

PROOF OF LEMMA 9.13: A sample path of Z^2 achieves its supremum only where Z achieves its supremum or infimum. By Lemma 2.6 of KP, if $\mathrm{Var}(Z(s) - Z(t)) \neq 0\ \forall s \neq t$, no sample path of Z achieves its supremum at two distinct points of T with probability 1. By the same argument, no sample path of Z achieves its infimum at two distinct points in T with probability 1.

It only remains to show that with probability 1, no sample path of Z has its supremum equal to minus its infimum at two distinct points. To show this, we use the condition

$$\mathrm{Var}(Z(s) + Z(t)) \neq 0 \quad \forall s \neq t. \tag{9.135}$$

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, $\mathrm{Cov}(Z(t_0), -Z(t_1)) = -H(t_0, t_1)$. By (9.135), $-H(t_0, t_1)$ 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 $P\{\sup_{t\in N_0} Z(t) = \sup_{t\in N_1}(-Z(t))\} = 0$. *Q.E.D.*

PROOF OF LEMMA 8.5: For any $\pi_1, \pi_2 \in \Pi$,

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

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 $\mathrm{Var}(G_1^*(\pi_1; \gamma_0) + G_1^*(\pi_2; \gamma_0)) \neq 0$ $\forall \pi_1, \pi_2 \in \Pi$ with $\pi_1 \neq \pi_2$. Hence, Assumption C6* holds. By Lemma 8.4, Assumption C6 holds as well. *Q.E.D.*

9.6.4. *Quadratic Expansions: Assumptions C1 and D1*

PROOF OF LEMMA 8.6: 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

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

where $\psi_{0,n}^{\dagger}(\pi)$ lies between ψ and $\psi_{0,n}$, and the $o_{p\pi}(\|\psi - \psi_{0,n}\|^2)$ 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 $Q_n(\theta)$ w.r.t. θ,

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

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). *Q.E.D.*

PROOF OF LEMMA 8.7: We first prove part (a). For any function $f(w,\theta)$, define the empirical process $\{\nu_n f(\theta):\theta\in\Theta\}$ by $\nu_n f(\theta)=n^{-1/2}\sum_{i=1}^n(f(W_i,\theta)-E_{\gamma_n}f(W_i,\theta))$. Note that

$$
\begin{aligned}
(9.139)\quad &Q_n(\theta)-Q_n(\psi_{0,n},\pi) \\
&\quad= n^{-1/2}(\nu_n\rho(\theta)-\nu_n\rho(\psi_{0,n},\pi))+Q_n^*(\theta)-Q_n^*(\psi_{0,n},\pi).
\end{aligned}
$$

The expansion in (8.10) implies that

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

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

$$
\begin{aligned}
(9.141)\quad &Q_n^*(\theta)-Q_n^*(\psi_{0,n},\pi) \\
&\quad=\frac{\partial}{\partial\psi}Q_n^*(\psi_{0,n},\pi)'(\psi-\psi_{0,n})
\end{aligned}
$$

$$+\frac{1}{2}(\psi-\psi_{0,n})'\left(\frac{\partial^2}{\partial\psi\,\partial\psi'}Q_n^*(\psi_{0,n},\pi)\right)(\psi-\psi_{0,n})$$
$$+o_\pi\big(\|\psi-\psi_{0,n}\|^2\big),$$

using Assumption Q2(v) (where $o_\pi(\cdot)$ denotes $o(\cdot)$ uniformly over $\pi\in\Pi$). From (9.139)–(9.141), we have

$$\begin{aligned}(9.142)\quad &Q_n(\theta)-Q_n(\psi_{0,n},\pi)\\ &=\left(n^{-1/2}\nu_n\Delta_\psi(\psi_{0,n},\pi)+\frac{\partial}{\partial\psi}Q_n^*(\psi_{0,n},\pi)\right)'(\psi-\psi_{0,n})\\ &\quad+\frac{1}{2}(\psi-\psi_{0,n})'\frac{\partial^2}{\partial\psi\,\partial\psi'}Q_n^*(\psi_{0,n},\pi)(\psi-\psi_{0,n})\\ &\quad+n^{-1/2}\nu_n r_\psi(\theta)+o_\pi\big(\|\psi-\psi_{0,n}\|^2\big).\end{aligned}$$

When $D_\psi Q_n(\theta)$ and $D_{\psi\psi}Q_n(\theta)$ take the form as in Lemma 8.7(a), the quadratic approximation in Assumption C1(i) holds with

$$(9.143)\quad R_n(\psi,\pi)=n^{-1/2}\nu_n r_\psi(\theta)+o_\pi\big(\|\psi-\psi_{0,n}\|^2\big).$$

To verify Assumption C1(ii), we have

$$\begin{aligned}(9.144)\quad &\sup_{\psi\in\Psi(\pi):\|\psi-\psi_{0,n}\|\le\delta_n}\frac{|a_n^2(\gamma_n)R_n(\psi,\pi)|}{(1+\|a_n(\gamma_n)(\psi-\psi_{0,n})\|)^2}\\ &\le\sup_{\psi\in\Psi(\pi):\|\psi-\psi_{0,n}\|\le\delta_n}\frac{|a_n^2(\gamma_n)n^{-1/2}\nu_n r_\psi(\theta)|}{(1+\|a_n(\gamma_n)(\psi-\psi_{0,n})\|)^2}+o_\pi(1)=o_{p\pi}(1),\end{aligned}$$

where the inequality follows from (9.143) 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

$$(9.145)\quad Q_n(\theta)-Q_n(\theta_n)=n^{-1/2}(\nu_n\rho(\theta)-\nu_n\rho(\theta_n))+Q_n^*(\theta)-Q_n^*(\theta_n).$$

The expansion in (8.9) gives

$$(9.146)\quad \nu_n\rho(\theta)-\nu_n\rho(\theta_n)=\nu_n\Delta(\theta_n)'(\theta-\theta_n)+\nu_n r(\theta).$$

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

$$\begin{aligned}(9.147)\quad &Q_n^*(\theta)-Q_n^*(\theta_n)\\ &=\frac{\partial}{\partial\theta}Q_n^*(\theta_n)'(\theta-\theta_n)+\frac{1}{2}(\theta-\theta_n)'\frac{\partial^2}{\partial\theta\,\partial\theta'}Q_n^*(\theta_n^\dagger)(\theta-\theta_n),\end{aligned}$$

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

$$
\begin{aligned}
(9.148)\quad & B^{-1}(\beta_n)\frac{\partial^2}{\partial\theta\,\partial\theta'}Q_n^*(\theta_n^\dagger)B^{-1}(\beta_n) \\
&= B^{-1}(\beta_n)\frac{\partial^2}{\partial\theta\,\partial\theta'}Q_n^*(\theta_n)B^{-1}(\beta_n) + o(1),
\end{aligned}
$$

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

Equations (9.145)–(9.148) yield

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

When $DQ_n(\theta)$ and $D^2Q_n(\theta)$ take the form in Lemma 8.7(b), the quadratic approximation in Assumption D1 holds with

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

To verify Assumption D1(ii), we have

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

where the inequality holds by (9.150) 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. *Q.E.D.*

PROOF OF LEMMA 8.8: Lemma 8.8(a) is proved using the proof of Lemma 8.6 with (9.137) and (9.138) changed to

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

respectively. By Assumption Q3(ii), Assumptions C1 and D1 follow from the same arguments as those in the proof of Lemma 8.6.

Lemma 8.8(b) is proved using the proof of Lemma 8.7 with (9.143) and (9.150) changed to

$$R_n(\psi,\pi) = n^{-1/2}\nu_n r_\psi(\theta) + o_\pi\big(\|\psi-\psi_{0,n}\|^2\big) + Q_n^{\mathrm{IC}}(\psi,\pi) - Q_n^{\mathrm{IC}}(\psi_{0,n},\pi) \quad \text{and} \tag{9.153}$$

$$R_n^*(\theta) = n^{-1/2}\nu_n r(\theta) + o\big(\|B(\beta_n)(\theta-\theta_n)\|^2\big) + Q_n^{\mathrm{IC}}(\theta) - Q_n^{\mathrm{IC}}(\theta_n),$$

respectively. By Assumption Q3(ii), Assumptions C1 and D1 follow from the same arguments as those in the proof of Lemma 8.7. *Q.E.D.*

10. SUPPLEMENTAL 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 6.

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 $\varepsilon_{t-1},\dots,\varepsilon_1$ in (1.1), we have

$$\varepsilon_t = \sum_{j=0}^{t-1}\pi_0^j(Y_{t-j}-\rho_0 Y_{t-j-1}) + \pi_0^t\varepsilon_0. \tag{10.1}$$

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

$$-\frac{n}{2}\log\zeta - \frac{1}{2\zeta}\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1}\pi^j[Y_{t-j}-(\pi+\beta)Y_{t-j-1}]+\pi^t\varepsilon_0\right)^2. \tag{10.2}$$

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

$$Q_n(\theta) = \frac{1}{2}\log\zeta + \frac{1}{2\zeta}n^{-1}\sum_{t=1}^{n}\left(Y_t-\beta\sum_{j=0}^{t-1}\pi^j Y_{t-j-1}\right)^2. \tag{10.3}$$

10.1. *ARMA Example: Initial Conditions Adjustment*

We use the initial conditions adjustment of the criterion function given in Lemma 8.8(a) of Section 8.7.3. 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 8.8(a) relies on Assumption Q3. We verify Assumption Q3 with

$$Q_n^\infty(\theta) = n^{-1}\sum_{t=1}^{n}\rho_t(\theta), \quad \text{where} \tag{10.4}$$

$$\rho_t(\theta) = \frac{1}{2}\log\zeta + \frac{1}{2\zeta}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)^2,$$

$$\begin{aligned} Q_n^{\mathrm{IC}}(\theta) &= Q_n(\theta) - Q_n^\infty(\theta) \\ &= -\frac{\beta^2}{2\zeta}n^{-1}\sum_{t=1}^{n}\left(\sum_{j=t}^{\infty}\pi^j Y_{t-j-1}\right)^2 \\ &\quad + \frac{\beta}{\zeta}n^{-1}\sum_{t=1}^{n}\left(Y_t - \beta\sum_{j=0}^{t-1}\pi^j Y_{t-j-1}\right)\sum_{j=t}^{\infty}\pi^j Y_{t-j-1}. \end{aligned}$$

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 (10.4), $W_t = (Y_t, Y_{t-1})'$ and $\rho_t(\theta)$ depends not only on W_t, but also on $W_{t-1}, \ldots, W_1$. This does not affect the results in Lemma 8.8(a).

LEMMA 10.1: *For the* ARMA(1, 1) *model*, $\{Q_n^{\mathrm{IC}}(\theta) : n \geq 1\}$ *satisfies the following statements*:

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

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

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

for all constants $\delta_n \to 0$.

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

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

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: (i) Lemma 10.1(a) implies that it suffices to establish Assumption B3 with $Q_n^\infty(\theta)$ in place of $Q_n(\theta)$.

(ii) Assumption Q3 holds by Lemma 10.1(b) and (c).

The proof of Lemma 10.1 is given in Section 10.4 below.

10.2. *ARMA Example: Derivation of Formulae for Key Quantities*

The quantities that appear in Assumptions B1–B3, C1–C8, and D1–D3, namely, $Q(\theta;\gamma_0)$, $D_\psi Q_n(\theta)$, $\Omega(\pi_1,\pi_2;\gamma_0)$, $D_{\psi\psi}Q_n(\theta)$, $H(\pi;\gamma_0)$, $K(\pi;\gamma_0)$, $\Omega_G(\pi_1,\pi_2;\gamma_0)$, $DQ_n(\theta)$, $D^2Q_n(\theta)$, $J(\gamma_0)$, and $V(\gamma_0)$, as well as $\pi^*(\gamma_0,b)$ and $\Sigma_{\pi\pi}(\pi)$, are specified in Section 3 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 10.3.4 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 $n^{-1}\sum_{t=1}^n \rho_t(\theta)$ w.r.t. ψ for $\rho_t(\theta)$ defined in (10.4). Analogously, $DQ_n(\theta)$ and $D^2Q_n(\theta)$ are the ordinary first and second partial derivatives of $n^{-1}\sum_{t=1}^n \rho_t(\theta)$ w.r.t. 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

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

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

$$
\begin{aligned}
(10.6)\quad & \rho_{\beta,t}(\psi_0,\pi) = -\zeta_0^{-1}\varepsilon_t \sum_{j=0}^{\infty} \pi^j \varepsilon_{t-j-1}, \\
& \rho_{\zeta,t}(\psi_0,\pi) = -(1/2)\zeta_0^{-2}(\varepsilon_t^2-\zeta_0)
\end{aligned}
$$

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 (6.5). The off-diagonal elements in (10.5) 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 10.3.3 to equal $E_{\gamma_0}\rho_{\psi\psi,t}(\psi_0, \pi)$. Using the definitions of $\rho_{\psi\psi,t}(\theta), \ldots, \rho_{\zeta\zeta,t}(\theta)$ in (6.8), when the true parameter is γ_0 with $\beta_0 = 0$, we have

$$\rho_{\beta\beta,t}(\psi_0, \pi) = \zeta_0^{-1}\left(\sum_{j=0}^{\infty} \pi^j \varepsilon_{t-j-1}\right)^2, \tag{10.7}$$

$$\rho_{\beta\zeta,t}(\psi_0, \pi) = \zeta_0^{-2}\varepsilon_t \sum_{j=0}^{\infty} \pi^j \varepsilon_{t-j-1},$$

$$\rho_{\zeta\zeta,t}(\psi_0, \pi) = -(1/2)\zeta_0^{-2} + \zeta_0^{-3}\varepsilon_t^2.$$

Using these expressions, we obtain

$$H(\pi; \gamma_0) = E_{\gamma_0}\rho_{\psi\psi,t}(\psi_0, \pi) = \begin{bmatrix} \zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty} \pi^j \varepsilon_{t-j-1}\right)^2 & 0 \\ 0 & (2\zeta_0^2)^{-1} \end{bmatrix} \tag{10.8}$$

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

Now, we calculate the covariance kernel $\Omega_G(\pi_1, \pi_2; \gamma_0)$ that appears in Assumption C6**. For $\beta_0 = 0$, we define

$$\rho^*_{\psi,t}(\psi_0, \pi_1, \pi_2) = (\rho_{\beta,t}(\psi_0, \pi_1), \rho_{\beta,t}(\psi_0, \pi_2), \rho_{\zeta,t}(\psi_0, \pi)')', \quad \text{where} \tag{10.9}$$

$$\rho_{\beta,t}(\psi_0, \pi) = -\zeta_0^{-1}\varepsilon_t \sum_{k=0}^{\infty} \pi^k Y_{t-k-1} = -\zeta_0^{-1}\varepsilon_t \sum_{k=0}^{\infty} \pi^k \varepsilon_{t-k-1},$$

$$\rho_{\zeta,t}(\psi_0, \pi) = -(1/2)\zeta_0^{-2}(\varepsilon_t^2 - \zeta_0).$$

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

$$\Omega_G(\pi_1, \pi_2; \gamma_0) \tag{10.10}$$

$$= \sum_{m=-\infty}^{\infty} \mathrm{Cov}_{\gamma_0}(\rho^*_{\psi,t}(\psi_0, \pi_1, \pi_2), \rho^*_{\psi,t+m}(\psi_0, \pi_1, \pi_2))$$

$$= \mathrm{Var}_{\gamma_0}(\rho^*_{\psi,t}(\psi_0, \pi_1, \pi_2))$$

$$
\begin{aligned}
&= \begin{bmatrix}
\zeta_0^{-1} E_{\gamma_0}\left(\sum_{j=0}^{\infty} \pi_1^j \varepsilon_{t-j-1}\right)^2 & \zeta_0^{-1} E_{\gamma_0}\left(\sum_{j=0}^{\infty} \pi_1^j \varepsilon_{t-j-1}\right)\left(\sum_{j=0}^{\infty} \pi_2^j \varepsilon_{t-j-1}\right) & 0 \\
\zeta_0^{-1} E_{\gamma_0}\left(\sum_{j=0}^{\infty} \pi_1^j \varepsilon_{t-j-1}\right)\left(\sum_{j=0}^{\infty} \pi_2^j \varepsilon_{t-j-1}\right) & \zeta_0^{-1} E_{\gamma_0}\left(\sum_{j=0}^{\infty} \pi_2^j \varepsilon_{t-j-1}\right)^2 & 0 \\
0 & 0 & (1/4)\zeta_0^{-4} E_{\gamma_0}(\varepsilon_t^2 - \zeta_0)^2
\end{bmatrix} \\
&= \begin{bmatrix}
(1-\pi_1^2)^{-1} & (1-\pi_1\pi_2)^{-1} & 0 \\
(1-\pi_1\pi_2)^{-1} & (1-\pi_2^2)^{-1} & 0 \\
0 & 0 & (1/4)\zeta_0^{-4} E_{\gamma_0}(\varepsilon_t^2 - \zeta_0)^2
\end{bmatrix}.
\end{aligned}
$$

The second and third equalities of (10.10) hold using (10.9) 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

$$
\begin{aligned}
D^2 Q_n(\theta) &= n^{-1}\sum_{t=1}^{n} \rho_{\theta\theta,t}(\theta) \\
&= n^{-1}\sum_{t=1}^{n} \begin{bmatrix}
\rho_{\beta\beta,t}(\theta) & \rho_{\beta\zeta,t}(\theta) & \rho_{\beta\pi,t}(\theta) \\
\rho_{\beta\zeta,t}(\theta) & \rho_{\zeta\zeta,t}(\theta) & \rho_{\zeta\pi,t}(\theta) \\
\rho_{\beta\pi,t}(\theta) & \rho_{\zeta\pi,t}(\theta) & \rho_{\pi\pi,t}(\theta)
\end{bmatrix},
\end{aligned} \tag{10.11}
$$

where

$$
\rho_{\beta\beta,t}(\theta) = \zeta^{-1}\left(\sum_{j=0}^{\infty} \pi^j Y_{t-j-1}\right)^2, \tag{10.12}
$$

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

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

and

$$\rho_{\zeta\zeta,t}(\theta) = -(1/2)\zeta^{-2} + \zeta^{-3}\left(Y_t - \beta\sum_{j=0}^{\infty}\pi^j Y_{t-j-1}\right)^2, \tag{10.13}$$

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

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

To determine $J(\gamma_0)$ via the expression $J(\gamma_0) = E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0)$ given in (10.51) below (in the verification of Assumption D2), we define $\rho^{\dagger}_{\theta\theta,t}(\theta)$ and $\chi_t(\theta)$ via

$$B^{-1}(\beta)\rho_{\theta\theta,t}(\theta)B^{-1}(\beta) = \rho^{\dagger}_{\theta\theta,t}(\theta) + \beta^{-1}\chi_t(\theta), \tag{10.14}$$

where $\rho_{\theta\theta,t}(\theta)$ is defined in (10.11)–(10.13) and $\rho^{\dagger}_{\theta\theta,t}(\theta)$ is defined by

$$\rho^{\dagger}_{\theta\theta,t}(\theta) = \begin{bmatrix} \rho_{\beta\beta,t}(\theta) & \rho_{\beta\zeta,t}(\theta) & \rho^{\dagger}_{\beta\pi,t}(\theta) \\ \rho_{\beta\zeta,t}(\theta) & \rho_{\zeta\zeta,t}(\theta) & \rho^{\dagger}_{\zeta\pi,t}(\theta) \\ \rho^{\dagger}_{\beta\pi,t}(\theta) & \rho^{\dagger}_{\zeta\pi,t}(\theta) & \rho^{\dagger}_{\pi\pi,t}(\theta) \end{bmatrix}, \tag{10.15}$$

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

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

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

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

$$\chi_t(\theta) = \begin{bmatrix} 0 & 0 & \chi_{\beta\pi,t}(\theta) \\ 0 & 0 & 0 \\ \chi_{\beta\pi,t}(\theta) & 0 & \chi_{\pi\pi,t}(\theta) \end{bmatrix}, \quad \text{where} \tag{10.16}$$

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

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

Now, using $J(\gamma_0) = E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0)$ and (10.12), (10.13), and (10.15), we have

$$\begin{aligned} J(\gamma_0) &= E_{\gamma_0}\rho^{\dagger}_{\theta\theta,t}(\theta_0) \\ &= \operatorname{Diag}\left\{\zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_0^j Y_{t-j-1}\right)^2, \frac{1}{2\zeta_0^2},\right. \\ &\quad \left.\zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty} j\pi_0^{j-1}Y_{t-j-1}\right)\sum_{k=0}^{\infty} k\pi_0^{k-1}Y_{t-k-1}\right\} \\ &\quad + \left(\zeta_0^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty}\pi_0^j Y_{t-j-1}\right)\sum_{k=0}^{\infty} k\pi_0^{k-1}Y_{t-k-1}\right) \\ &\quad \times \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{10.17}$$

As shown in Section 10.3.7 below, the matrix $n^{-1}\sum_{t=1}^{n}\beta^{-1}\chi_t(\theta)$ evaluated at $\theta = \theta_n$ ($\to \theta_0$) does not contribute to $J(\gamma_0)$ because its probability limit is zero.

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

$$\rho^{\dagger}_{\theta,t}(\theta) = B^{-1}(\beta)\rho_{\theta,t}(\theta) = (\rho_{\beta,t}(\theta), \rho_{\zeta,t}(\theta), \beta^{-1}\rho_{\pi,t}(\theta))', \tag{10.18}$$

$$V^{\dagger}(\theta_1, \theta_2; \gamma_0) = \sum_{m=-\infty}^{\infty}\operatorname{Cov}_{\gamma_0}(\rho^{\dagger}_{\theta,t}(\theta_1), \rho^{\dagger}_{\theta,t+m}(\theta_2)).$$

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

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

where the first equality holds because the convergence in distribution result in Assumption D3(i) is obtained by a CLT (see (10.56) 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 (10.20) below. The last two equalities in (10.19) hold because

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

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

The expression for $\pi^*(\gamma_0, b)$ given in (6.19) holds using the expression for $\xi(\pi; \gamma_0, b)$ for this example given in (6.10) plus simplifications based on (6.7)–(6.9). In particular, it uses the block diagonality of $H(\pi; \gamma_0)$ in (6.8) and the fact that the second element of $G(\pi; \gamma_0)$ in (6.7) does not depend on π. The expression for $\Sigma_{\pi\pi}(\pi)$ in (6.19) uses the expression for $\tau_\beta(\pi; \gamma_0, b)$ given just above (6.16) and the equality $\overline{\Sigma}_{\pi\pi}(\pi; \gamma_0, b) = \Sigma_{\pi\pi}(\pi)_{22}$, which holds using the

expressions for $J(\theta; \gamma_0)$ and $V(\theta; \gamma_0)$ in (10.57) and (10.58) and some calculations.

10.3. *ARMA Example: Verification of Assumptions*

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

10.3.1. *ARMA Example: Verification of Assumptions A and B1–B3*

Assumption A holds immediately given the definition of $\rho_t(\theta)$ in (10.4).

Assumption B1(i) holds by the definitions of Θ and Θ^* in (6.1). 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 (6.1).

Assumption B2(i) holds by the definition of Γ in (6.2). Assumption B2(ii) 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 B2(iii) holds by the definitions of Γ and Θ^*, and the condition $\rho_L^* < \pi_L^* < \pi_U^* < \rho_U^*$.

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 (1.1), 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 (10.33) below). For brevity, the details are omitted.

Assumption B3(ii) and (iii) are verified using Assumption B3* and Lemma 8.1 in Supplemental 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 (6.1). 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 (1.1), 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

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

using $\zeta_0 = E_{\gamma_0}\varepsilon_t^2\ \forall t = 0, 1, \ldots$. The l.h.s. is zero for $\psi = \psi_0$; the r.h.s. 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), that is, $Q(\theta; \gamma_0)$ is uniquely minimized by $\theta_0\ \forall\gamma_0 \in \Gamma$ with $\beta_0 \neq 0$. Using (10.4), we have

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

The first summand on the r.h.s. is uniquely minimized by $\zeta = \zeta_0$ by the argument following (10.21).

We now show that the second summand on the r.h.s. of (10.22) equals zero when $(\beta, \pi) = (\beta_0, \pi_0)$ and is positive for $(\beta, \pi) \neq (\beta_0, \pi_0)$. We have

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

where the first equality uses (1.1) and the second equality uses the independence of ε_{t-1}, and $(Y_{t-2}, \varepsilon_{t-2}, \ldots)$ and $E\varepsilon_{t-1} = 0$. The r.h.s. of (10.23) 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

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

The r.h.s. of (10.24) 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 summand on the r.h.s. of (10.22) is zero if and only if (iff) $(\beta, \pi) = (\beta_0, \pi_0)$. Hence, Assumption B3*(iii) holds. This completes the verification of Assumption B3*.

10.3.2. *ARMA Example: Verification of Assumptions C1 and D1*

We verify the quadratic expansions that appear in Assumptions C1 and D1 using Lemma 8.6, 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 $Y_t, Y_{t-1}, \ldots$, rather than just W_t, does not effect the result of Lemma 8.6.) 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 $\sup_{\pi\in\Pi}\sup_{\psi\in\Psi(\pi):\|\psi-\psi_{0,n}\|\le\delta_n} |E_{\gamma_n}\rho_{\psi\psi,t}(\psi, \pi) - E_{\gamma_n}\rho_{\psi\psi,t}(\psi_{0,n}, \pi)| \to 0$ 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 (10.11)–(10.13). 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 $n \ge 1$ and hence the limit is zero. For the $(1, 2)$ element of $\rho_{\psi\psi,t}(\theta)$, we have

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

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

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

where $\rho_{\theta\theta,t}^{\dagger}(\theta)$ and $\chi_t(\theta)$ are defined in (10.14). In (10.26), the second equality holds because $|\beta| \le |\beta - \beta_n| + |\beta_n| \le (1 + \delta_n)|\beta_n|$ and $\delta_n = o(1)$. By (10.26) 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 $n^{-1}\sum_{t=1}^{n}\rho^{\dagger}_{\theta\theta,t}(\theta)$ 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 $n^{-1}\sum_{t=1}^{n}\rho^{\dagger}_{\theta\theta,t}(\theta)$ follows by the same argument as used above to verify Assumption B3(i) with $\rho^{\dagger}_{\theta\theta,t}(\theta)$ 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

$$v_n^*(\pi)=n^{-1/2}\sum_{t=1}^{n}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^j k\pi^{k-1}\big(Y_{t-j-1}Y_{t-k-1}-E_{\gamma_n}Y_{t-j-1}Y_{t-k-1}\big) \tag{10.27}$$

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 (10.16). For any $\varepsilon>0$, we have

$$\begin{aligned}
&\varepsilon^2 P_{\gamma_n}\Big(\sup_{|\pi_1-\pi_2|<\delta}|v_n^*(\pi_1)-v_n^*(\pi_2)|>\varepsilon\Big) \\
&\quad\le E_{\gamma_n}\sup_{|\pi_1-\pi_2|<\delta}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}k\frac{(\pi_1^{j+k-1}-\pi_2^{j+k-1})}{a_{jk}^{1/2}}\right. \\
&\qquad\left.\times a_{jk}^{1/2}n^{-1/2}\sum_{t=1}^{n}(Y_{t-j-1}Y_{t-k-1}-E_{\gamma_n}Y_{t-j-1}Y_{t-k-1})\right)^2 \\
&\quad\le \sup_{|\pi_1-\pi_2|<\delta}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}k^2\frac{(\pi_1^{j+k-1}-\pi_2^{j+k-1})^2}{a_{jk}} \\
&\qquad\times\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}a_{jk}E_{\gamma_n}\left(n^{-1/2}\sum_{t=1}^{n}(Y_{t-j-1}Y_{t-k-1}-E_{\gamma_n}Y_{t-j-1}Y_{t-k-1})\right)^2 \\
&\quad\le\varepsilon^3
\end{aligned} \tag{10.28}$$

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) $\lim_{\delta\to0}\sup_{|\pi_1-\pi_2|<\delta}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}k^2((\pi_1/\pi_{\#})^{j+k-1}-(\pi_2/\pi_{\#})^{j+k-1})^2=0$, 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 $(\pi_2/\pi_{\#})^{j+k-1}$ $\forall j,k\ge0$, (ii) $\mathrm{Var}_{\gamma_n}(n^{-1/2}\sum_{t=1}^{n}Y_{t-j-1}Y_{t-k-1})\le C$ $\forall n\ge1$ 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

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

To show that the first term on the r.h.s. of (10.29) is $o(1)$, we write

$$
\begin{aligned}
(10.30)\quad & E_{\gamma_n} \chi_{\beta\pi,t}(\theta) \\
&= -\zeta^{-1} E_{\gamma_n} \left(\beta_n \sum_{j=0}^{\infty} \pi_n^j Y_{t-j-1} - \beta \sum_{j=0}^{\infty} \pi^j Y_{t-j-1} \right) \sum_{k=0}^{\infty} k \pi^{k-1} Y_{t-k-1}, \\
& E_{\gamma_n} \chi_{\beta\pi,t}(\psi_n, \pi) \\
&= -\zeta_n^{-1} E_{\gamma_n} \left(\beta_n \sum_{j=0}^{\infty} (\pi_n^j - \pi^j) Y_{t-j-1} \right) \sum_{k=0}^{\infty} k \pi^{k-1} Y_{t-k-1},
\end{aligned}
$$

using the definition of $\chi_{\beta\pi,t}(\theta)$ in (10.16).

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

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

for some constant $C < \infty$, where the inequality uses the definition of $\Theta_n(\delta_n)$ and $|E_{\gamma_n} Y_{t-j-1} Y_{t-k-1}| \le E_{\gamma_n} Y_t^2 \le C_1 \ \forall n \ge 1$ for some constant $C_1 < \infty$. Combining (10.30), (10.31), and $\sup_{n \ge 1} |\zeta_n E_{\gamma_n} \chi_{\beta\pi,t}(\theta_n)| < \infty$ (which holds by standard calculations) establishes that the $(3,1)$ element (i.e., the $\beta\pi$ element) of the first term on the r.h.s. of (10.29) is $o(1)$:

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

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

The proof for the $(3,3)$ element (i.e., the $\pi\pi$ element) of the first term on the r.h.s. of (10.29), which is the only other nonzero 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 r.h.s. of (10.29) is $o(1)$.

Let $c_j = |E_{\gamma_0} Y_1 Y_{1+j}|$. The second summand on the r.h.s. of (10.29) is $O(\delta_n) = o(1)$ by the following calculations. For $\theta \in \Theta_n(\delta_n)$, we have

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

where the first equality holds by (10.30), the second inequality holds because $|\pi^j - \pi_n^j| \le |j \pi_{n*}^{j-1} (\pi - \pi_n)| \le j \pi_U^{j-1} |\pi - \pi_n|$ for some π_{n*} between π and π_n by a mean-value expansion and $\sup_{j \ge 1} c_j < \infty$, 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 $|\beta_n^{-1} E_{\gamma_n} \chi_{\pi\pi,t}(\psi_n, \pi)| \le |\pi - \pi_n| C^* = O(\delta_n) = o(1)$ for a constant $C^* < \infty$ by the same argument as in (10.33) with $k(k-1)\pi^{k-2}$ in place of $k\pi^{k-1}$. This concludes the proof that the second summand on the r.h.s. of (10.29) is $o(1)$, which completes the verification of Assumption Q1(iv). In turn, this completes the verification of Assumptions C1 and D1.

10.3.3. *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 (6.6). The covariance matrix of the stochastic process $\{G(\pi; \gamma_0) : \pi \in \Pi\}$ that appears in Assumption C3 is defined and derived in (10.5). 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 (6.8), 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 ver-

ification of Assumption Q1(iii); see (10.25). The simplified expression for $H(\pi;\gamma_0) = E_{\gamma_0}\rho_{\psi\psi,t}(\psi_0,\pi)$ is derived in (10.8).

Assumption C4(ii) holds because $H(\pi;\gamma_0) = \text{Diag}\{(1-\pi^2)^{-1}, (2\zeta_0^2)^{-1}\}$ by (10.8), $\inf_{\pi\in\Pi}(1-\pi^2)^{-1} \geq 1$, and $\zeta^* \geq \zeta_L^* > 0$ by the definition of Θ^*.

10.3.4. *ARMA Example: Verification of Assumption C5*

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

$$K_n(\theta;\gamma^*) = n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\beta^*}E_{\gamma^*}\rho_{\psi,t}(\theta) = \begin{pmatrix} \dfrac{\partial}{\partial\beta^*}E_{\gamma^*}\rho_{\beta,t}(\theta) \\ \dfrac{\partial}{\partial\beta^*}E_{\gamma^*}\rho_{\zeta,t}(\theta) \end{pmatrix}. \tag{10.34}$$

The terms on the r.h.s. of (10.34) are calculated as

$$\begin{aligned} &E_{\gamma^*}\rho_{\beta,t}(\theta) \\ &\quad = -\zeta^{-1}E_{\gamma^*}\left(\varepsilon_t + \beta^*\sum_{j=0}^{\infty}\pi^{*j}Y_{t-j-1} - \beta\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)\sum_{k=0}^{\infty}\pi^{k}Y_{t-k-1} \\ &\quad = -\zeta^{-1}\beta^*\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{*j}\pi^{k}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1} \\ &\qquad + \zeta^{-1}\beta\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{j}\pi^{k}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1} \end{aligned} \tag{10.35}$$

and

$$\begin{aligned} \frac{\partial}{\partial\beta^*}E_{\gamma^*}\rho_{\beta,t}(\theta) &= -\zeta^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{*j}\pi^{k}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1} \\ &\quad - \zeta^{-1}\beta^*\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{*j}\pi^{k}\frac{\partial}{\partial\beta^*}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1} \\ &\quad + \zeta^{-1}\beta\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{j}\pi^{k}\frac{\partial}{\partial\beta^*}E_{\gamma^*}Y_{t-j-1}Y_{t-k-1}. \end{aligned} \tag{10.36}$$

In addition, we have

$$\begin{aligned} &E_{\gamma^*}\rho_{\zeta,t}(\theta) \\ &\quad = -(1/2)\zeta^{-2}\left(E_{\gamma^*}\left(\varepsilon_t + \beta^*\sum_{j=0}^{\infty}\pi^{*j}Y_{t-j-1} - \beta\sum_{j=0}^{\infty}\pi^{j}Y_{t-j-1}\right)^2 - \zeta\right) \end{aligned} \tag{10.37}$$

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

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

$$\left. \times E_{\gamma^*} Y_{t-j-1} Y_{t-k-1}\right).$$

This gives

$$\text{(10.38)} \quad \frac{\partial}{\partial \beta^*} E_{\gamma^*} \rho_{\zeta,t}(\theta)$$

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

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

$$\times \frac{\partial}{\partial \beta^*} E_{\gamma^*} Y_{t-j-1} Y_{t-k-1}.$$

From (10.36), if $\widetilde{\gamma}_n \to \gamma_0$ with $\beta_0 = 0$ (for nonstochastic $\widetilde{\gamma}_n$) and $\psi_n \to \psi_0 = (0, \zeta_0)$, as in Assumption C5, then

$$\text{(10.39)} \quad \frac{\partial}{\partial \widetilde{\beta}_n} E_{\widetilde{\gamma}_n} \rho_{\beta,t}(\psi_0, \pi)$$

$$\to -\zeta_0^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \pi_0^j \pi^k E_{\gamma_0} Y_{t-j-1} Y_{t-k-1}$$

$$= -\zeta_0^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \pi_0^j \pi^k E_{\gamma_0} \varepsilon_{t-j-1} \varepsilon_{t-k-1}$$

$$= -\sum_{j=0}^{\infty} \pi_0^j \pi^j = -\frac{1}{1 - \pi_0 \pi}.$$

The convergence is uniform in $\pi \in \Pi$ because (i) $|\pi| \le \max\{|\pi_L|, |\pi_U|\} < 1$ $\forall \pi \in \Pi$ 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 $n \ge 1$. This holds because when the true pa-

rameter is $\widetilde{\gamma}_n$, we can write

$$(10.40)\quad Y_t = (\widetilde{\pi}_n + \widetilde{\beta}_n)Y_{t-1} + u_t = \sum_{j=0}^{\infty}(\widetilde{\pi}_n + \widetilde{\beta}_n)^j u_{t-j-1}, \quad \text{where}$$

$$u_t = \varepsilon_t - \widetilde{\pi}_n \varepsilon_{t-1} \quad \text{and}$$

$$\frac{\partial}{\partial \widetilde{\beta}_n} E_{\widetilde{\gamma}_n} Y_s Y_t$$

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

From (10.38), 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

$$(10.41)\quad \frac{\partial}{\partial \widetilde{\beta}_n} E_{\widetilde{\gamma}_n} \rho_{\zeta,t}(\psi_n, \pi) \to 0$$

due to the multiplicative terms β^*, β, β^{*2}, $\beta^*\beta$, and β^2 that appear in (10.38) and that converge to 0 when $\beta^* = \widetilde{\beta}_n \to 0$ and $\beta = \beta_n \to 0$.

Combining (10.34), (10.39), and (10.41) verifies Assumption C5(i) and (ii) with $K(\pi; \gamma_0) = (-(1 - \pi_0\pi)^{-1}, 0)$. Assumption C5(iii) holds because $1 - \pi_0\pi \neq 0\ \forall \pi \in \Pi$.

10.3.5. *ARMA Example: Verification of Assumption C6*

Now, we verify Assumption C6 using Assumption C6**, which is shown in Lemma 8.5 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 r.h.s. matrix in (10.10) 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 (6.1) and (6.2). This completes the verification of Assumption C6**. Hence, Assumption C6 holds.

10.3.6. *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 (10.35), we have

$$(10.42)\quad \frac{\partial}{\partial \beta} E_{\gamma^*} \rho_{\beta,t}(\theta) = \zeta^{-1} \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \pi^j \pi^k E_{\gamma^*} Y_{t-j-1} Y_{t-k-1},$$

which leads to

$$\begin{aligned}(10.43)\quad \frac{\partial}{\partial\beta}E_{\gamma_n}\rho_{\beta,t}(\psi,\pi_n)\Big|_{\psi=\psi_n} &= \zeta_n^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi_n^j\pi_n^k E_{\gamma_n}Y_{t-j-1}Y_{t-k-1}\\ &\to \zeta_0^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi_0^j\pi_0^k E_{\gamma_0}Y_{t-j-1}Y_{t-k-1}\\ &= \zeta_0^{-1}\sum_{j=0}^{\infty}\pi_0^{2j}E_{\gamma_0}\varepsilon_{t-j-1}^2\\ &= \frac{1}{1-\pi_0^2},\end{aligned}$$

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 (10.35), we also have

$$\begin{aligned}(10.44)\quad \frac{\partial}{\partial\zeta}E_{\gamma^*}\rho_{\beta,t}(\theta) &= \zeta^{-2}\beta^*\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^{*j}\pi^k E_{\gamma^*}Y_{t-j-1}Y_{t-k-1}\\ &\quad - \zeta^{-2}\beta\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi^j\pi^k E_{\gamma^*}Y_{t-j-1}Y_{t-k-1},\end{aligned}$$

which yields

$$(10.45)\quad \frac{\partial}{\partial\zeta}E_{\gamma_n}\rho_{\beta,t}(\psi,\pi_n)\Big|_{\psi=\psi_n} = 0 \quad \forall n \geq 1.$$

From (10.37), we have

$$\begin{aligned}(10.46)\quad &\frac{\partial}{\partial\beta}E_{\gamma^*}\rho_{\zeta,t}(\theta)\\ &\quad = \zeta^{-2}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(\beta^*\pi^{*j}\pi^k - \beta\pi^{j+k})E_{\gamma^*}Y_{t-j-1}Y_{t-k-1}\right),\end{aligned}$$

which yields

$$(10.47)\quad \frac{\partial}{\partial\beta}E_{\gamma_n}\rho_{\zeta,t}(\psi,\pi_n)\Big|_{\psi=\psi_n} = 0 \quad \forall n \geq 1.$$

From (10.37), we also have

$$\begin{aligned}(10.48)\quad &\frac{\partial}{\partial\zeta}E_{\gamma^*}\rho_{\zeta,t}(\theta)\\ &= \zeta^{-3}\Bigg(\zeta^* - \zeta + \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\big(\beta^{*2}\pi^{*(j+k)} - 2\beta^*\beta\pi^{*j}\pi^k + \beta^2\pi^{j+k}\big)\\ &\qquad \times E_{\gamma^*}Y_{t-j-1}Y_{t-k-1}\Bigg) + (1/2)\zeta^{-2},\end{aligned}$$

which yields

$$(10.49)\quad \frac{\partial}{\partial\zeta}E_{\gamma_n}\rho_{\zeta,t}(\psi,\pi_n)\bigg|_{\psi=\psi_n} = (1/2)\zeta_n^{-2} \to (1/2)\zeta_0^{-2}.$$

Combining (10.43), (10.45), (10.47), and (10.49) gives

$$\begin{aligned}(10.50)\quad \frac{\partial}{\partial\psi'}E_{\gamma_n}D_\psi Q_n(\psi,\pi_n)\bigg|_{\psi=\psi_n} &= \frac{\partial}{\partial\psi'}E_{\gamma_n}\rho_{\psi,t}(\psi,\pi_n)\bigg|_{\psi=\psi_n}\\ &\to \begin{bmatrix}(1-\pi_0^2)^{-1} & 0\\ 0 & (1/2)\zeta_0^{-2}\end{bmatrix} = H(\pi_0;\gamma_0),\end{aligned}$$

where the first equality holds by (6.5). This completes the verification of Assumption C8.

10.3.7. *ARMA Example: Verification of Assumption D2*

Next, we verify Assumption D2. By (10.26), we have

$$\begin{aligned}(10.51)\quad J_n &= B^{-1}(\beta_n)n^{-1}\sum_{t=1}^{n}\rho_{\theta\theta,t}(\theta_n)B^{-1}(\beta_n)\\ &= \left(n^{-1}\sum_{t=1}^{n}\rho^\dagger_{\theta\theta,t}(\theta_n)\right)(1+o(1))\\ &\quad + \left(n^{-1/2}\sum_{t=1}^{n}\big(\chi_t(\theta_n) - E_{\gamma_n}\chi_t(\theta_n)\big)\right)\big(n^{1/2}\beta_n\big)^{-1}(1+o(1))\\ &\quad + \big(E_{\gamma_n}\chi_t(\theta_n)/\beta_n\big)(1+o(1))\\ &= \left(n^{-1}\sum_{t=1}^{n}\rho^\dagger_{\theta\theta,t}(\theta_n)\right)(1+o(1)) + o(1)\\ &= E_{\gamma_n}\rho^\dagger_{\theta\theta,t}(\theta_n) + o_p(1) \to_p E_{\gamma_0}\rho^\dagger_{\theta\theta,t}(\theta_0) = J(\gamma_0),\end{aligned}$$

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 (10.30) 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 $\mathrm{Var}_{\gamma_n}(n^{-1/2}\sum_{t=1}^n \chi_{\beta\pi,t}(\theta_n))^2 = O(1)$ 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 $t = 1, \ldots, n$ and likewise for $n^{-1/2} \times \sum_{t=1}^n \chi_{\pi\pi,t}(\theta_n)$; 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 (10.12)–(10.15) and $\gamma_n \to \gamma_0$.

The form of the matrix $J(\gamma_0)$ given in (6.13) is derived in (10.11)–(10.17) 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 (10.17), 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 submatrix 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

$$\text{(10.52)}\quad E_{\gamma_0}A_tA_t', \quad \text{where} \quad A_t = \begin{pmatrix} A_{1t} \\ A_{2t} \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^\infty \pi_0^j Y_{t-j-1} \\ \sum_{j=1}^\infty j\pi_0^{j-1}Y_{t-j-1} \end{pmatrix}.$$

Now, by (1.1), $Y_t = \varepsilon_t + (\pi_0 + \beta_0)Y_{t-1} - \pi_0\varepsilon_{t-1}$. Hence,

$$\text{(10.53)}\quad A_{1t} = Y_{t-1} + \sum_{j=1}^\infty \pi_0^j Y_{t-j-1} = \varepsilon_{t-1} + \xi_{t-2}, \quad \text{where}$$

$$\xi_{t-2} = (\pi_0 + \beta_0)Y_{t-2} - \pi_0\varepsilon_{t-2} + \sum_{j=1}^\infty \pi_0^j Y_{t-j-1}$$

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

$$\text{(10.54)}\quad \lambda' E_{\gamma_0}A_tA_t'\lambda = E_{\gamma_0}\left(\lambda_1\varepsilon_{t-1} + \lambda_1\xi_{t-2} + \lambda_2\sum_{j=1}^\infty j\pi_0^{j-1}Y_{t-j-1}\right)^2$$

$$= \lambda_1^2 E_{\gamma_0}\varepsilon_{t-1}^2 + E_{\gamma_0}\left(\lambda_1\xi_{t-2} + \lambda_2\sum_{j=1}^\infty j\pi_0^{j-1}Y_{t-j-1}\right)^2.$$

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

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

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 (p.d.). This completes the verification that $J(\gamma_0)$ is positive definite.

10.3.8. *ARMA Example: Verification of Assumption D3*

Assumption D3(i) is verified as follows. By the definitions in (6.5) and (6.12), and $B(\beta) = \text{Diag}\{1, 1, \beta\}$, we have

$$
\begin{aligned}
(10.56)\quad & n^{1/2} B^{-1}(\beta_n) DQ_n(\theta_n) \\
&= n^{-1/2} \sum_{t=1}^{n} B^{-1}(\beta_n) \rho_{\theta,t}(\theta_n) \\
&= -n^{-1/2} \sum_{t=1}^{n} \begin{pmatrix} \zeta_n^{-1} \varepsilon_t \sum_{k=0}^{\infty} \pi_n^k Y_{t-k-1} \\ (1/2)\zeta_n^{-2}(\varepsilon_t^2 - \zeta_n) \\ \zeta_n^{-1} \varepsilon_t \sum_{k=0}^{\infty} k\pi_n^{k-1} Y_{t-k-1} \end{pmatrix} \\
&\to_d N(0, V(\gamma_0)),
\end{aligned}
$$

where the convergence in distribution holds by a triangular array martingale difference CLT for rowwise stationary random variables (e.g., see Hall and Hyde (1980, Theorem 3.1)) and $V(\gamma_0) = \lim_{n\to\infty} \text{Var}_{\gamma_n}(n^{-1/2} \sum_{t=1}^{n} B^{-1}(\beta_n) \times \rho_{\theta,t}(\theta_n))$. The verification of the conditions of Hall and Hyde's martin-

gale difference CLT is essentially the same as given in the proof of Theorem 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 (6.2), to verify a Lyapounov-type condition. The formula for $V(\gamma_0)$ given in (6.15) is derived in (10.18)–(10.20).

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 $(1/4)\zeta_0^{-4}E_{\gamma_0}(\varepsilon_t^2 - \zeta_0)^2$ in place of $(2\zeta_0^2)^{-1}$; see (10.17) and (10.19). Because $(1/4)\zeta_0^{-4}E_{\gamma_0}(\varepsilon_t^2 - \zeta_0)^2 > 0$ by the definition of the parameter spaces Θ^* and Φ^*, the same argument as used above to show that $J(\gamma_0)$ is p.d. also shows that $V(\gamma_0)$ is p.d. Hence, Assumption D3(ii) holds.

10.3.9. *ARMA Example: Verification of Assumptions V1 and V2*

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

$$
\begin{aligned}
(10.57)\quad & J(\theta; \gamma_0) \\
&= \operatorname{Diag}\left\{ \zeta^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty} \pi^j Y_{t-j-1}\right)^2, (2\zeta^2)^{-1}, \right. \\
&\quad \left. \zeta^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty} j\pi^{j-1} Y_{t-j-1}\right)^2 \right\} \\
&\quad + \left(\zeta^{-1}E_{\gamma_0}\left(\sum_{j=0}^{\infty} \pi^j Y_{t-j-1}\right)\sum_{k=0}^{\infty} k\pi^{k-1} Y_{t-k-1}\right) \times \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}
\end{aligned}
$$

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

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

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 $E_\phi|\xi_t|^{8+\delta_2} \le K$ in the definition of Φ in (6.2).

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

$$A(\pi) = \zeta_0^{-1} E_{\gamma_0} \begin{pmatrix} \sum_{j=0}^{\infty} \pi^j Y_{t-j-1} \\ \sum_{j=0}^{\infty} j\pi^{j-1} Y_{t-j-1} \end{pmatrix} \begin{pmatrix} \sum_{j=0}^{\infty} \pi^j Y_{t-j-1} \\ \sum_{j=0}^{\infty} j\pi^{j-1} Y_{t-j-1} \end{pmatrix}' \tag{10.59}$$

$$= \begin{pmatrix} \sum_{j=0}^{\infty} \pi^{2j} & \sum_{j=0}^{\infty} j\pi^{2j-1} \\ \sum_{j=0}^{\infty} j\pi^{2j-1} & \sum_{j=0}^{\infty} j^2\pi^{2(j-1)} \end{pmatrix},$$

where the first equality holds by (10.57) 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

$$\left(\sum_{j=0}^{\infty} j\pi^{2j-1}\right)^2 < \left(\sum_{j=0}^{\infty} \pi^{2j}\right)\left(\sum_{j=0}^{\infty} j^2\pi^{2(j-1)}\right) \quad \forall \pi \in \Pi \tag{10.60}$$

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 (10.57) and (10.58), we have $\Sigma_{22}(\pi;\gamma_0) = (2\zeta_0^2)(4\zeta_0^2)^{-1} E_{\gamma_0}(Y_t^2 - \zeta_0)^2 (2\zeta_0^2) = \zeta_0^2 E_{\gamma_0}(\varepsilon_t^2 - \zeta_0)^2$, which lies in $(0,\infty)$ because $\zeta_0 = \mathrm{Var}(\varepsilon_t) > 0$ and $E_{\gamma_0}\varepsilon_t^4 < \infty$. This completes the verification of Assumption V1(iii).

Assumption V1(i) and (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 3.3, imply that Assumption V2 holds.

10.3.10. *ARMA Example: Verification of Assumptions RQ and RQ3*

Assumptions RQ2(ii) and RQ3 hold with $s(\gamma_0) = \widehat{s}_n = 1$ in the ARMA$(1,1)$ example for restrictions $r(\theta)$ that only involve the parameters (β, π) because (i) $V(\gamma_0)$ and $J(\gamma_0)$ are block diagonal between the parameters (β, π) and ζ, where ζ is the innovation variance, and (ii) the blocks of $V(\gamma_0)$ and $J(\gamma_0)$ that correspond to (β, π) are equal whether or not the innovations are normally distributed. (In contrast, the blocks corresponding to ζ are equal under normality, but not for more general error distributions.)

10.4. *Proof of the ARMA Initial Conditions Lemma*

PROOF OF LEMMA 10.1: To prove part (a), we write

$$
\begin{aligned}
2\zeta_L Q_n^{\mathrm{IC}}(\theta) &= 2\zeta_L |Q_n^{\infty}(\theta) - Q_n(\theta)| \\
&\le \left| n^{-1} \sum_{t=1}^{n} \left[(A_t - B_t)^2 - A_t^2 \right] \right| \\
&= \left| n^{-1} \sum_{t=1}^{n} [-2A_t B_t + B_t^2] \right| \\
&\le 2 \left(n^{-1} \sum_{t=1}^{n} A_t^2 \right)^{1/2} \left(n^{-1} \sum_{t=1}^{n} B_t^2 \right)^{1/2} + n^{-1} \sum_{t=1}^{n} B_t^2,
\end{aligned} \tag{10.61}
$$

where

$$
A_t = A_t(\theta) = Y_t - \beta \sum_{j=0}^{t-1} \pi^j Y_{t-j-1} \quad \text{and} \tag{10.62}
$$

$$
B_t = B_t(\theta) = \beta \sum_{j=t}^{\infty} \pi^j Y_{t-j-1}.
$$

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

$$
\sup_{\theta \in \Theta} n^{-1} \sum_{t=1}^{n} A_t^2(\theta) = O_p(1) \quad \text{and} \quad \sup_{\theta \in \Theta} n^{-1} \sum_{t=1}^{n} B_t^2(\theta) = o_p(1). \tag{10.63}
$$

To show (10.63), we have

$$
\begin{aligned}
n^{-1} \sum_{t=1}^{n} B_t^2(\theta) &= \beta^2 n^{-1} \sum_{t=1}^{n} \left(\sum_{j=t}^{\infty} \pi^j Y_{t-j-1} \right)^2 \\
&= \beta^2 n^{-1} \sum_{t=1}^{n} \left(\sum_{k=0}^{\infty} \pi^{t+k} Y_{-k-1} \right)^2 \\
&\le n^{-1} \beta_U^2 \sum_{t=1}^{\infty} \pi_+^{2t} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_+^{j+k} |Y_{-j-1} Y_{-k-1}|,
\end{aligned} \tag{10.64}
$$

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 (10.64), we obtain

$$E_{\gamma_n} \sup_{\theta\in\Theta} n^{-1} \sum_{t=1}^{n} B_t^2(\theta) \le n^{-1}\beta_U^2 \sum_{t=1}^{\infty} \pi_+^{2t} \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \pi_+^{j+k} E_{\gamma_n} Y_1^2 \to 0, \tag{10.65}$$

where the inequality uses $E_{\gamma_n}|Y_{-j-1}Y_{-k-1}| \le \sup_{n\ge1} E_{\gamma_n} Y_1^2 \le C < \infty$ by the Cauchy–Schwarz inequality and stationarity.

Next, we have

$$\begin{aligned}
&E_{\gamma_n} \sup_{\theta\in\Theta} n^{-1} \sum_{t=1}^{n} A_t^2(\theta) \\
&\quad\le \sup_{t\ge1} E_{\gamma_n} \sup_{\theta\in\Theta} A_t^2(\theta) \\
&\quad\le 2\sup_{t\ge1} E_{\gamma_n} Y_t^2 + 2\sup_{t\ge1} E_{\gamma_n} \sup_{\theta\in\Theta} \left(\beta \sum_{j=0}^{t-1} \pi^j Y_{t-j-1}\right)^2 \\
&\quad\le 2\sup_{n,t\ge1} E_{\gamma_n} Y_t^2 \\
&\qquad + 2\beta_U^2 \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \pi_+^{j+k} \sup_{n,t\ge1,j,k\ge0} E_{\gamma_n}|Y_{t-j-1}Y_{t-k-1}| < \infty.
\end{aligned} \tag{10.66}$$

This completes the proof of part (a).

Next, we establish part (b). By (10.61) and (10.62),

$$A_t(\psi_{0,n}, \pi) = Y_t, \quad B_t(\psi_{0,n}, \pi) = 0, \quad \text{and} \quad Q_n^{\mathrm{IC}}(\psi_{0,n}, \pi) = 0. \tag{10.67}$$

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

$$\sup_{\pi\in\Pi} \sup_{\psi\in\Psi(\pi):\|\psi-\psi_{0,n}\|\le\delta_n} \frac{|a_n^2(\gamma_n) Q_n^{\mathrm{IC}}(\psi, \pi)|}{(1 + \|a_n(\gamma_n)(\psi - \psi_{0,n})\|)^2} = o_p(1) \tag{10.68}$$

for all constants $\delta_n \to 0$. The l.h.s. of (10.68) is less than or equal to

$$\sup_{\theta\in\Theta:|\beta|\le\delta_n} |n Q_n^{\mathrm{IC}}(\theta)| = o_p(1), \tag{10.69}$$

where the equality holds by (10.61) and (10.64)–(10.66) because (10.64) and (10.65) hold with β_U replaced by δ_n and $\delta_n \to 0$.

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

$$\sup_{\theta\in\Theta_n(\delta_n)} |Q_n^{\mathrm{IC}}(\theta) - Q_n^{\mathrm{IC}}(\theta_n)| = o_p(n^{-1}) \tag{10.70}$$

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

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

First, suppose $\zeta = \zeta_n$. Then, using (10.61), we have

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

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

To bound the first two terms on the r.h.s. of (10.71), we have

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

where the last inequality holds by mean-value expansions of π^j around π_n^j for $j \ge 1$ and $\pi_+ = \max\{|\pi_L|, |\pi_U|\}$, and

$$
(10.73)\quad B_t(\theta) = \beta\sum_{j=t}^{\infty}\pi^j Y_{t-j-1} = \beta\sum_{k=0}^{\infty}\pi^{t+k}Y_{-k-1},
$$

$$
\begin{aligned}
&|B_t(\theta) - B_t(\theta_n)| \\
&\quad \le \left| (\beta - \beta_n) \sum_{k=0}^{\infty} \pi^{t+k} Y_{-k-1} + \beta_n \sum_{k=0}^{\infty} (\pi^{t+k} - \pi_n^{t+k}) Y_{-k-1} \right| \\
&\quad \le \delta_n \beta_U \sum_{k=0}^{\infty} \pi_+^{t+k} |Y_{-k-1}| + |\pi - \pi_n| \beta_U \sum_{k=0}^{\infty} (t+k) \pi_+^{t+k-1} |Y_{-k-1}|, \\
&\sup_{\theta \in \Theta_n(\delta_n)} |B_t(\theta) - B_t(\theta_n)| \le \delta_n \beta_U \pi_+^t \sum_{k=0}^{\infty} [\pi_+^k + (t+k)\pi_+^{k-1}] |Y_{-k-1}|,
\end{aligned}
$$

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 $k \ge 0$.

Using (10.72) and (10.73), we have the following bound on the expectation of the supremum over $\theta \in \Theta_n(\delta_n)$ of the first term on the r.h.s. of (10.71):

$$
\begin{aligned}
(10.74) \quad & 2E_{\gamma_n} \sup_{\theta \in \Theta_n(\delta_n)} n^{-1} \sum_{t=1}^{n} |A_t(\theta)| \cdot |B_t(\theta) - B_t(\theta_n)| \\
&\le 2n^{-1} \delta_n \sum_{t=1}^{\infty} \pi_+^t \sum_{k=0}^{\infty} [\pi_+^k + \beta_U (t+k) \pi_+^{k-1}] E_{\gamma_n} |Y_t Y_{-k-1}| \\
&\quad + 2n^{-1} \delta_n \beta_U \sum_{t=1}^{\infty} \pi_+^t \sum_{j=0}^{\infty} \pi_+^{j-1} \sum_{k=0}^{\infty} [\pi_+^k + \beta_U (t+k) \pi_+^{k-1}] \\
&\quad \times E_{\gamma_n} |Y_{t-j-1} Y_{-k-1}| = o(n^{-1}),
\end{aligned}
$$

using $E_{\gamma_n} |Y_{t-j-1} Y_{-k-1}| \le \sup_{n \ge 1} E_{\gamma_n} Y_1^2 \le C < \infty$ and $\pi_+ \in (0,1)$. By Markov's inequality, (10.74) implies that the l.h.s. quantity with E_{γ_n} deleted is $o_p(n^{-1})$, as desired.

Similarly, using (10.72) and (10.73), we have the following bound on the expectation of the supremum over $\theta \in \Theta_n(\delta_n)$ of the second term on the r.h.s. of (10.71):

$$
\begin{aligned}
(10.75) \quad & E_{\gamma_n} \sup_{\theta \in \Theta_n(\delta_n)} n^{-1} \sum_{t=1}^{n} |A_t(\theta) - A_t(\theta_n)| \cdot |B_t(\theta_n)| \\
&\le n^{-1} \delta_n \beta_U^2 \sum_{t=1}^{\infty} \pi_+^t \sum_{j=1}^{\infty} [\pi_+^j + j \pi_+^{j-1}] \sum_{k=0}^{\infty} \pi_+^k \\
&\quad \times \sup_{n,t \ge 1, j,k \ge 0} E_{\gamma_n} |Y_{t-j-1} Y_{-k-1}| = o(n^{-1}).
\end{aligned}
$$

Hence, the l.h.s. of (10.75) with E_{γ_n} deleted is $o_p(n^{-1})$.

Next, we consider the third term on the r.h.s. of (10.71):

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

The supremum over $\theta\in\Theta_n(\delta_n)$ of the absolute value of the first term on the r.h.s. of (10.76) 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 (10.64) and (10.65). The expectation of the supremum over $\theta\in\Theta_n(\delta_n)$ of the absolute value of the second term on the r.h.s. of (10.76) is bounded by

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

The equality in (10.77) holds because

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

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 (10.77) implies that the supremum over $\theta\in\Theta_n(\delta_n)$ of the absolute value of the second term on the r.h.s. of (10.76) is $o_p(n^{-1})$. Hence, we conclude that the supremum over $\theta\in\Theta_n(\delta_n)$ of the absolute value of the l.h.s. of (10.76), which is the third summand in (10.71), is $o_p(n^{-1})$.

This completes the verification of (10.70) for the case where $\zeta=\zeta_n$.

Last, we consider the case where $\zeta\ne\zeta_n$. We have

$$
\begin{aligned}
(10.79)\quad |Q_n^{\mathrm{IC}}(\theta)-Q_n^{\mathrm{IC}}(\theta_n)| &= |Q_n^{\mathrm{IC}}(\theta)-Q_n^{\mathrm{IC}}(\beta_n,\zeta,\pi_n)| \\
&\quad + |Q_n^{\mathrm{IC}}(\beta_n,\zeta,\pi_n)-Q_n^{\mathrm{IC}}(\beta_n,\zeta_n,\pi_n)|.
\end{aligned}
$$

The proof of part (c) for the case where $\zeta = \zeta_n$ gives $\sup_{\theta\in\Theta_n(\delta_n)} |Q_n^{\mathrm{IC}}(\theta) - Q_n^{\mathrm{IC}}(\beta_n, \zeta, \pi_n)| = o_p(n^{-1})$. It remains to show

$$\sup_{\theta\in\Theta_n(\delta_n)} |Q_n^{\mathrm{IC}}(\beta_n, \zeta, \pi_n) - Q_n^{\mathrm{IC}}(\beta_n, \zeta_n, \pi_n)| = o_p(n^{-1}). \tag{10.80}$$

We have

$$\begin{aligned} Q_n^{\mathrm{IC}}(\beta_n, \zeta, \pi_n) &= Q_n(\beta_n, \zeta, \pi_n) - Q_n^{\infty}(\beta_n, \zeta, \pi_n) \\ &= \frac{1}{2\zeta} n^{-1} \sum_{t=1}^{n} \left(Y_t - \beta_n \sum_{j=0}^{t-1} \pi_n^j Y_{t-j-1} \right)^2 - \frac{1}{2\zeta} n^{-1} \sum_{t=1}^{n} \varepsilon_t^2 \\ &= \frac{1}{2\zeta} n^{-1} \sum_{t=1}^{n} \left(\varepsilon_t + \beta_n \sum_{j=t}^{\infty} \pi_n^j Y_{t-j-1} \right)^2 - \frac{1}{2\zeta} n^{-1} \sum_{t=1}^{n} \varepsilon_t^2 \\ &= \frac{1}{\zeta} n^{-1} \sum_{t=1}^{n} \varepsilon_t \beta_n \sum_{k=0}^{\infty} \pi_n^{t+k} Y_{-k-1} \\ &\quad + \frac{1}{2\zeta} n^{-1} \sum_{t=1}^{n} \left(\beta_n \sum_{k=0}^{\infty} \pi_n^{t+k} Y_{-k-1} \right)^2. \end{aligned} \tag{10.81}$$

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

$$\begin{aligned} &|Q_n^{\mathrm{IC}}(\beta_n, \zeta, \pi_n) - Q_n^{\mathrm{IC}}(\beta_n, \zeta_n, \pi_n)| \\ &\quad \le \frac{|\zeta - \zeta_n|}{\zeta\zeta_n} \left| n^{-1} \sum_{t=1}^{n} \varepsilon_t \beta_n \sum_{k=0}^{\infty} \pi_n^{t+k} Y_{-k-1} \right| \\ &\qquad + \frac{|\zeta - \zeta_n|}{2\zeta\zeta_n} n^{-1} \sum_{t=1}^{n} \left(\beta_n \sum_{k=0}^{\infty} \pi_n^{t+k} Y_{-k-1} \right)^2. \end{aligned} \tag{10.82}$$

We have

$$\begin{aligned} &E_{\gamma_n} \sup_{\theta\in\Theta_n(\delta_n)} \left| n^{-1} \sum_{t=1}^{n} \varepsilon_t \beta_n \sum_{k=0}^{\infty} \pi_n^{t+k} Y_{-k-1} \right| \\ &\quad \le n^{-1} \beta_U \sum_{t=1}^{\infty} \sum_{k=0}^{\infty} \pi_+^{t+k} \sup_{n\ge 1, k\ge 0} E_{\gamma_n} |\varepsilon_t Y_{-k-1}| = O(n^{-1}), \end{aligned} \tag{10.83}$$

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

$$E_{\gamma_n} n^{-1} \sum_{t=1}^{n} \left(\beta_n \sum_{k=0}^{\infty} \pi_n^{t+k} Y_{-k-1} \right)^2 \tag{10.84}$$
$$\leq n^{-1} \beta_U^2 \sum_{t=1}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_+^{t+j} \pi_+^{t+k} \sup_{n \geq 1, j, k \geq 0} E_{\gamma_n} |Y_{-j-1} Y_{-k-1}| = O(n^{-1}).$$

Equations (10.83) and (10.84) and Markov's inequality, coupled with (10.82) and $\sup_{\theta \in \Theta_n(\delta_n)} |\zeta - \zeta_n| \leq \delta_n = o(1)$, establish (10.80), which completes the proof of part (c). *Q.E.D.*

11. SUPPLEMENTAL APPENDIX D: ARMA(1, 1) NUMERICAL RESULTS

This appendix provides (i) a table containing the constants $c^{\mathrm{LF}}_{\mathcal{T},1-\alpha}(v)$, $\Delta_1(v)$, and $\Delta_2(v)$ that are used to compute the type 2 NI robust critical values that are used to construct CI's for the MA and AR CI's, (ii) details concerning the ARMA(1, 1) simulation computations, and (iii) additional numerical results.

11.1. *Table of Constants for Type 2 Robust CI's With NI Critical Values*

Table S-I provides the $c^{\mathrm{LF}}_{\mathcal{T},1-\alpha}(v)$, $\Delta_1(v)$, and $\Delta_2(v)$ values necessary to compute the type 2 NI robust critical values for the $|t|$ and QLR test statistics for computing CI's for the MA and AR parameters. These CI's employ the unrestricted ICS A_n. (The same values apply to both the MA and AR parameters.) In this case, v denotes the null hypothesis value of π (or ρ), which we denote by π_{H_0} (or ρ_{H_0}) in the table. For π_{H_0} (or ρ_{H_0}) values between those given in Table S-I, linear interpolation can be used.

11.2. *Simulation Details*

To achieve an approximately stationary startup, 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.

TABLE S-I

NI-LF CRITICAL VALUES AND VALUES OF $\Delta_1(\pi_{H_0})$ AND $\Delta_2(\pi_{H_0})$ FOR SIZE CORRECTION IN THE ARMA(1, 1) MODEL

$\lvert t\rvert$	π_{H_0}/ρ_{H_0}	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
	$c^{\mathrm{LF}}_{\lvert t\rvert,0.95}(\pi_{H_0})$	6.43	6.43	6.43	6.43	6.57	6.81	7.09	7.39	7.69	8.01	8.31
	$\Delta_1(\pi_{H_0})$	1.22	1.21	1.19	1.12	0.90	0.64	0.32	0.22	0.20	0.19	0.20
	$\Delta_2(\pi_{H_0})$	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.06	0.05	0.06	0.06
	π_{H_0}/ρ_{H_0}	0.55	0.60	0.625	0.65	0.675	0.70	0.725	0.75	0.775	0.80	0.825
	$c^{\mathrm{LF}}_{\lvert t\rvert,0.95}(\pi_{H_0})$	8.62	8.94	9.09	9.24	9.40	9.55	9.70	9.86	10.01	10.17	10.25
	$\Delta_1(\pi_{H_0})$	0.21	0.22	0.22	0.23	0.24	0.25	0.25	0.26	0.26	0.27	0.26
	$\Delta_2(\pi_{H_0})$	0.05	0.03	0.02	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.01
QLR	π_{H_0}/ρ_{H_0}	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
	$c^{\mathrm{LF}}_{\mathrm{QLR},0.95}(\pi_{H_0})$	4.30	4.31	4.32	4.32	4.33	4.32	4.31	4.30	4.29	4.28	4.25
	$\Delta_1(\pi_{H_0})$	0.60	0.62	0.71	0.73	0.76	0.81	0.82	0.77	0.68	0.64	0.55
	$\Delta_2(\pi_{H_0})$	0.08	0.08	0.08	0.09	0.10	0.10	0.08	0.09	0.09	0.09	0.09
	π_{H_0}/ρ_{H_0}	0.55	0.60	0.625	0.65	0.675	0.70	0.725	0.75	0.775	0.80	0.825
	$c^{\mathrm{LF}}_{\mathrm{QLR},0.95}(\pi_{H_0})$	4.21	4.13	4.08	4.07	4.09	4.12	4.16	4.22	4.29	4.36	4.37
	$\Delta_1(\pi_{H_0})$	0.57	0.55	0.54	0.45	0.29	0.18	0.07	0.09	0.11	0.12	0.12
	$\Delta_2(\pi_{H_0})$	0.06	0.04	0.04	0.03	0.04	0.04	0.04	0.02	0.01	0.00	0.00

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 minima 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 4, 6, and 7 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 nonnegative.

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, that is, $\rho_n \in [-0.85, 0.85]$. 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, that is, $\pi_n \in [-0.8, 0.8]$.

In Figure 1 of AC1 and Figures S-1 and S-2 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-11–S-13 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 2 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

$$-(1 - \pi^2)\left(\sum_{j=0}^{\infty} \pi^j Z_j - (1 - \pi_0 \pi)^{-1} b\right) + b. \tag{11.1}$$

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 1.

When determining κ for use with the robust CI's, we compute FCP's using $n = 500$.

11.3. *Additional Simulation Results*

In this section, we provide additional numerical results to those given in AC1. Figures S-1–S-9 provide results analogous to those in AC1, but for $\pi = 0.0$ and 0.7, rather than $\pi = 0.4$. Figure S-10 gives asymptotic 0.95 quantile

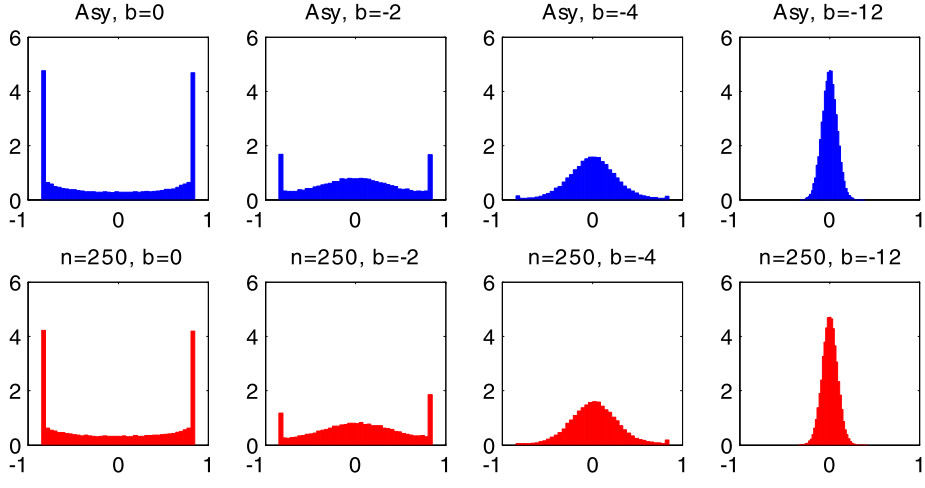

FIGURE S-1.—Asymptotic and finite-sample ($n = 250$) densities of the estimator of the MA parameter π in the ARMA(1, 1) model when $\pi_0 = 0$.

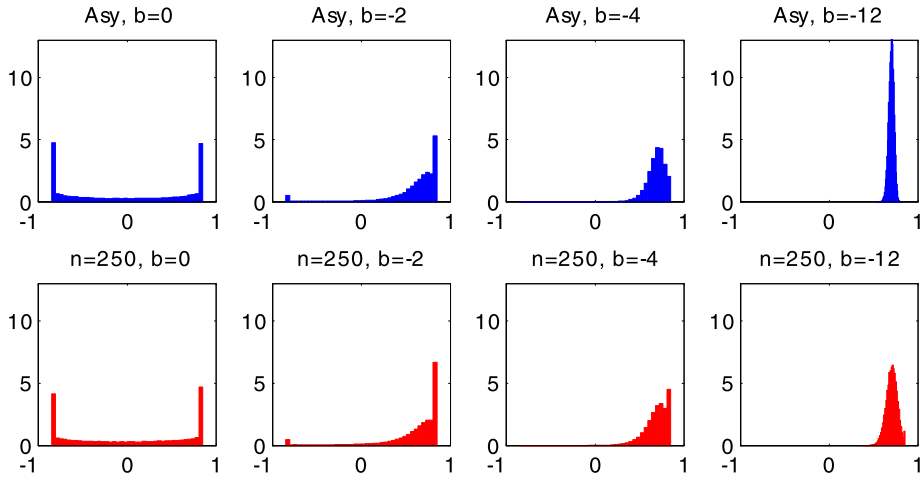

FIGURE S-2.—Asymptotic and finite-sample ($n = 250$) densities of the estimator of the MA parameter π in the ARMA$(1, 1)$ model when $\pi_0 = 0.7$.

graphs for the $|t|$ and QLR statistics for tests concerning β. Figures S-11–S-25 provide figures for the AR parameter ρ that are analogous to the figures given for the MA parameter π.

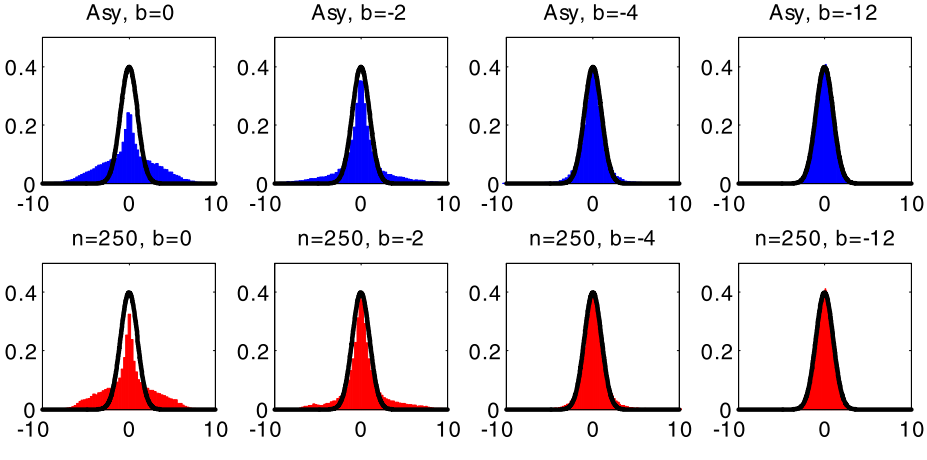

FIGURE S-3.—Asymptotic and finite-sample ($n = 250$) densities of the t statistic for the MA parameter π in the ARMA$(1, 1)$ model when $\pi_0 = 0$ and the standard normal density (black line).

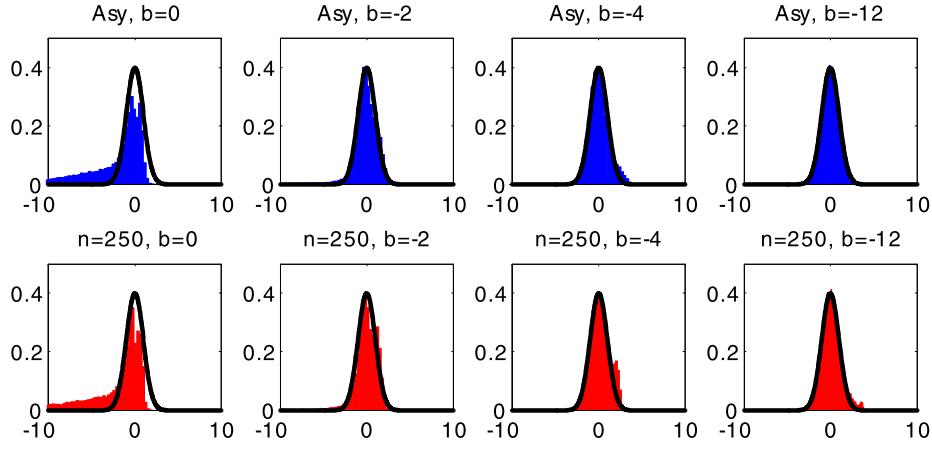

FIGURE S-4.—Asymptotic and finite-sample ($n = 250$) densities of the t statistic for the MA parameter π in the ARMA(1, 1) model when $\pi_0 = 0.7$ and the standard normal density (black line).

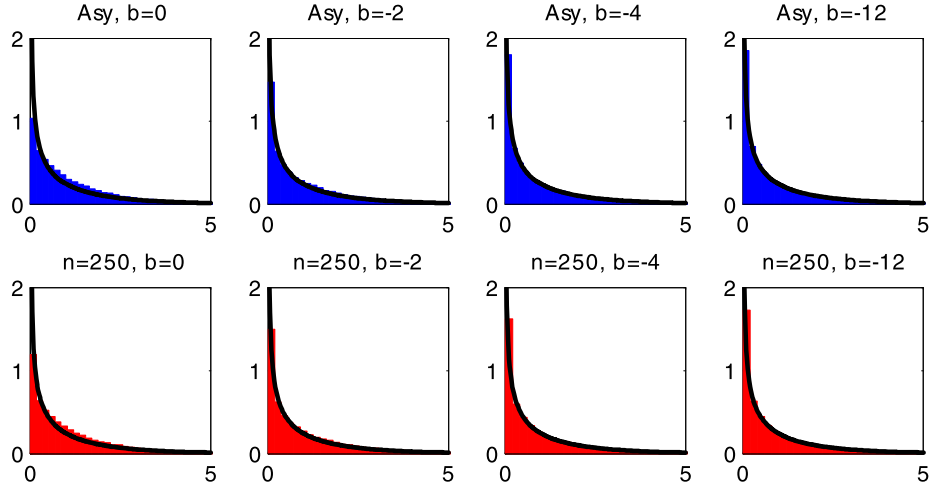

FIGURE S-5.—Asymptotic and finite-sample ($n = 250$) densities of the QLR statistic for the MA parameter π in the ARMA(1, 1) model when $\pi_0 = 0$ and the χ^2_1 density (black line).

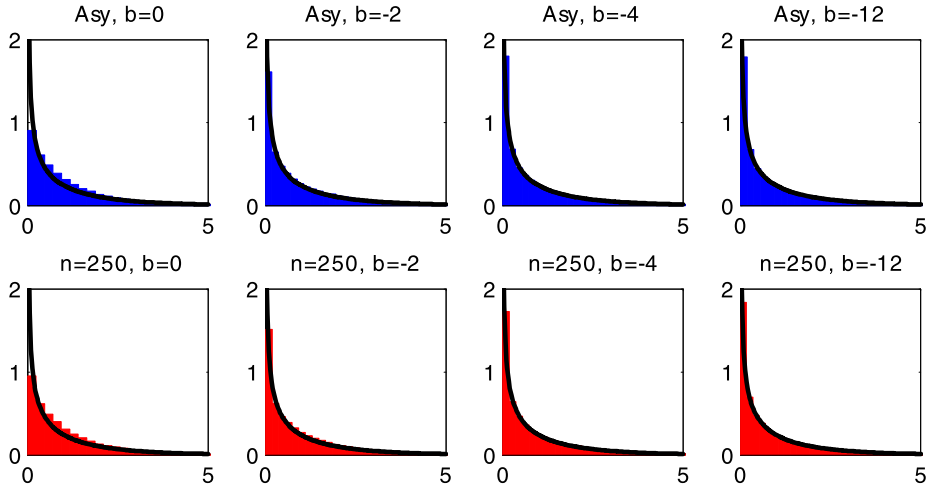

FIGURE S-6.—Asymptotic and finite-sample ($n = 250$) densities of the QLR statistic for the MA parameter π in the ARMA(1, 1) model when $\pi_0 = 0.7$ and the χ^2_1 density (black line).

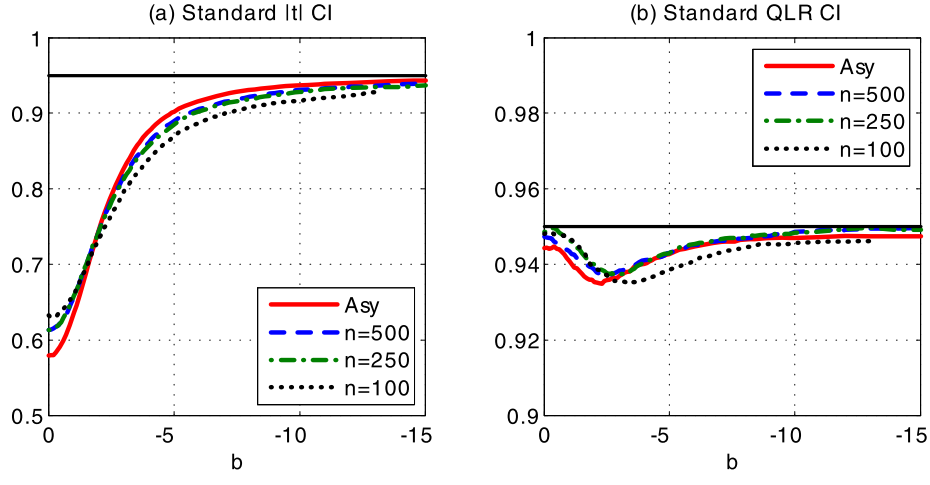

FIGURE S-7.—Coverage probabilities of standard $|t|$ and QLR CI's for the MA parameter π in the ARMA(1, 1) model when $\pi_0 = 0.4$.

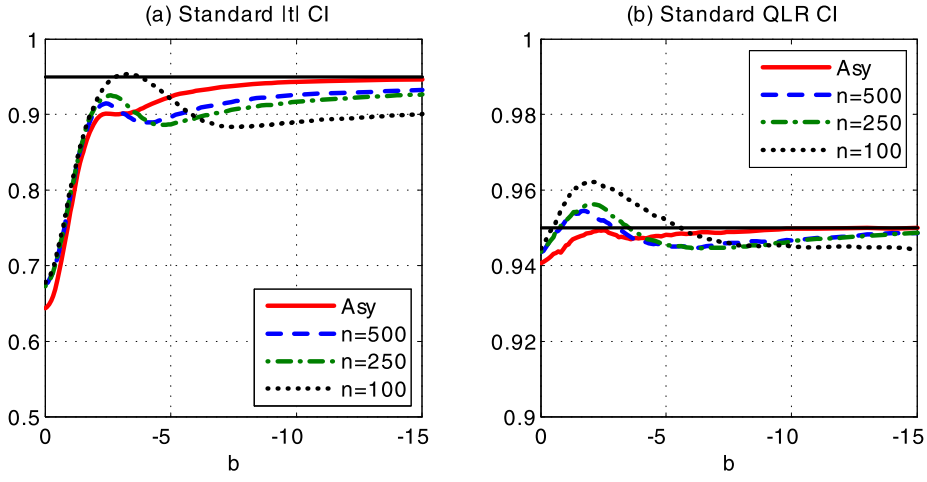

FIGURE S-8.—Coverage probabilities of standard $|t|$ and QLR CI's for the MA parameter π in the ARMA(1, 1) model when $\pi_0 = 0.7$.

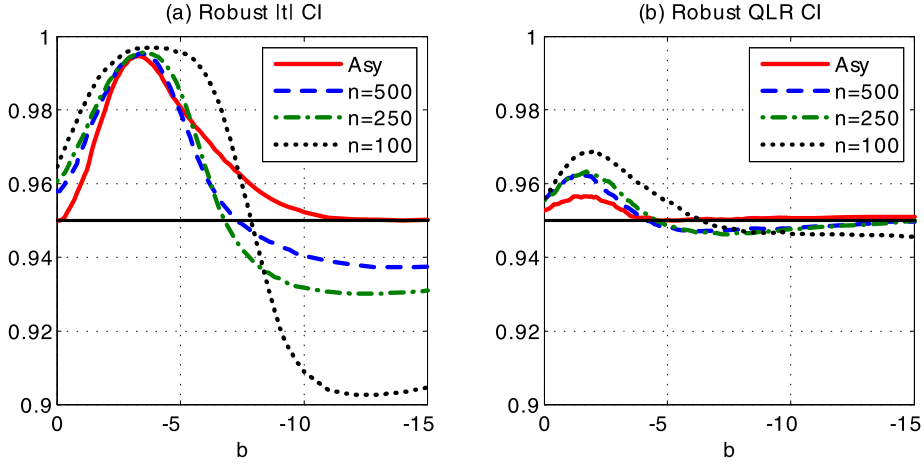

FIGURE S-9.—Coverage probabilities of robust $|t|$ and QLR CI's for the MA parameter π in the ARMA(1, 1) model when $\pi_0 = 0.7$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

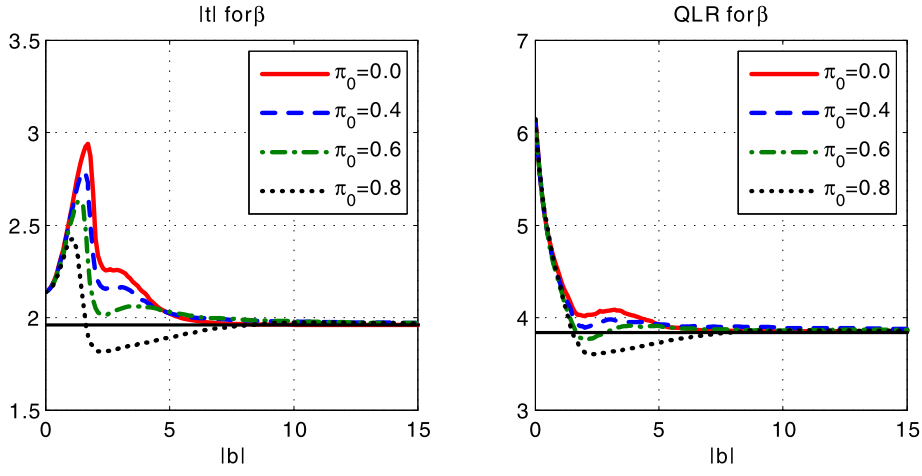

FIGURE S-10.—Asymptotic 0.95 quantiles of the $|t|$ and QLR statistics for tests concerning β in the ARMA(1, 1) model.

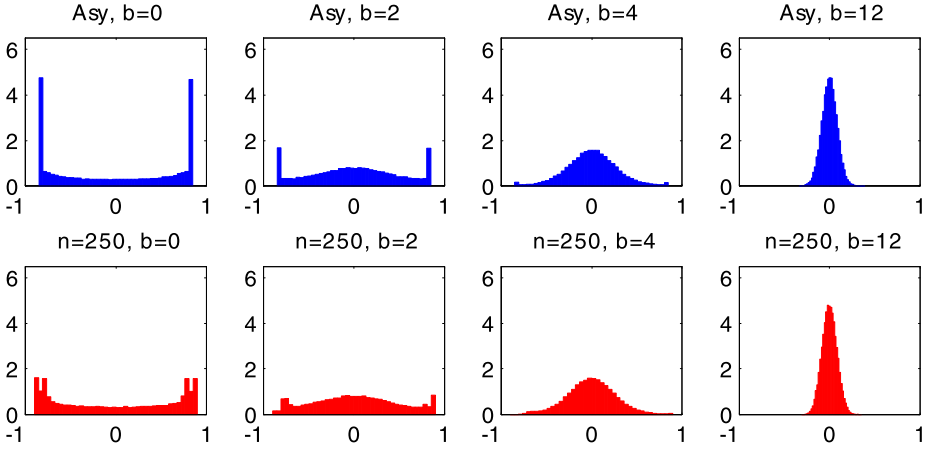

FIGURE S-11.—Asymptotic and finite-sample ($n = 250$) densities of the estimator of the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0$.

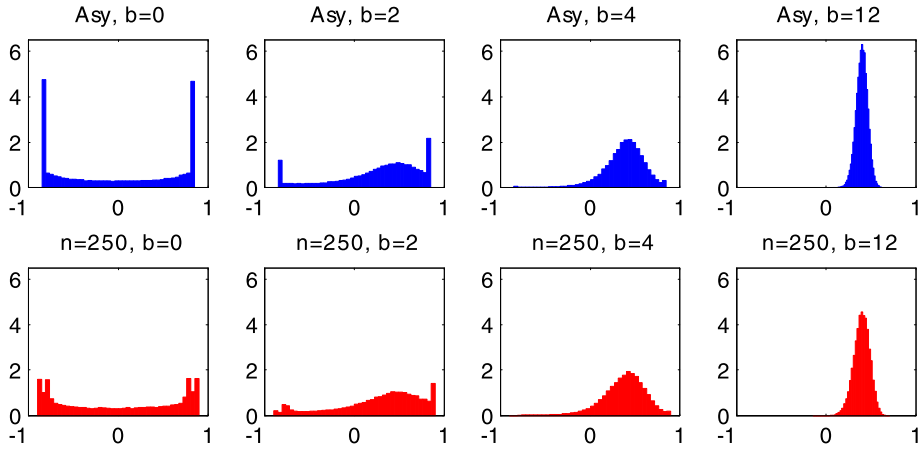

FIGURE S-12.—Asymptotic and finite-sample ($n = 250$) densities of the estimator of the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0.4$.

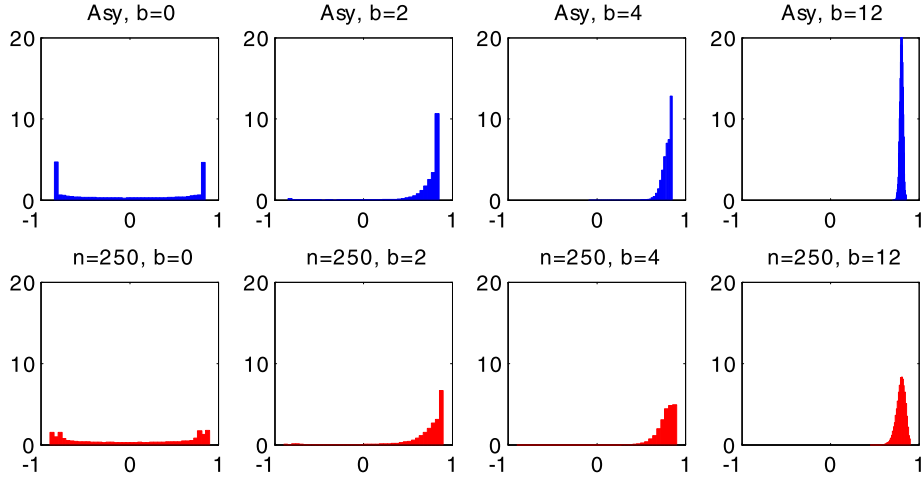

FIGURE S-13.—Asymptotic and finite-sample ($n = 250$) densities of the estimator of the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0.8$.

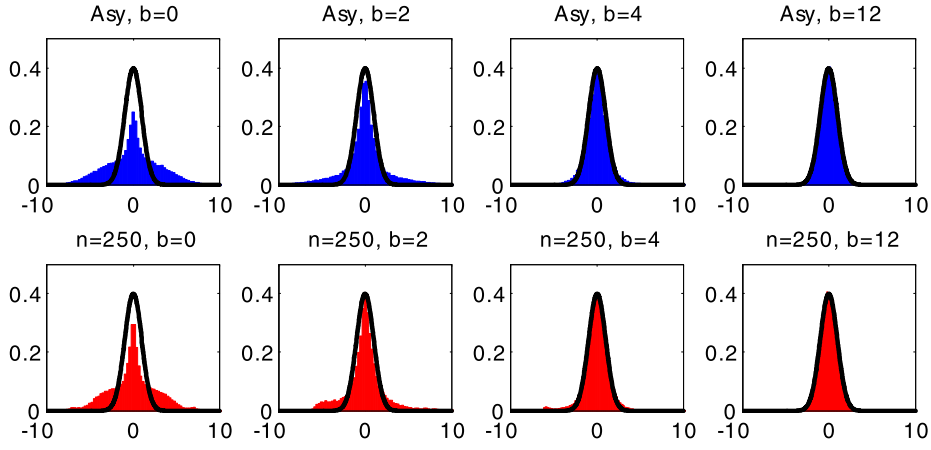

FIGURE S-14.—Asymptotic and finite-sample ($n = 250$) densities of the t statistic for the AR parameter ρ in the ARMA$(1, 1)$ model when $\rho_0 = 0$ and the standard normal density (black line).

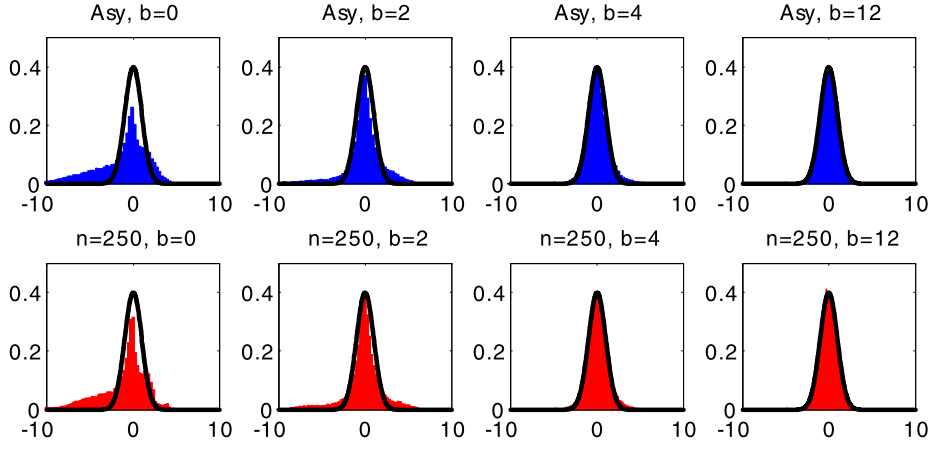

FIGURE S-15.—Asymptotic and finite-sample ($n = 250$) densities of the t statistic for the AR parameter ρ in the ARMA$(1, 1)$ model when $\rho_0 = 0.4$ and the standard normal density (black line).

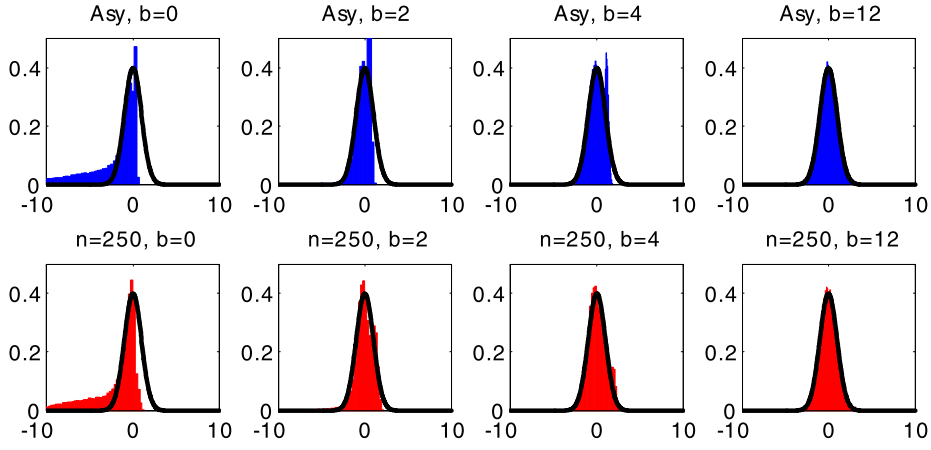

FIGURE S-16.—Asymptotic and finite-sample ($n = 250$) densities of the t statistic for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0.8$ and the standard normal density (black line).

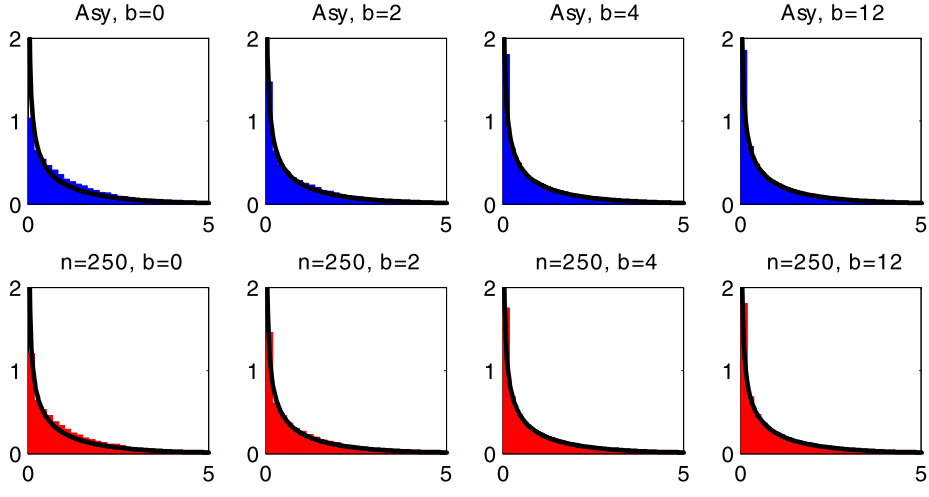

FIGURE S-17.—Asymptotic and finite-sample ($n = 250$) densities of the QLR statistic for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0$ and the χ^2_1 density (black line).

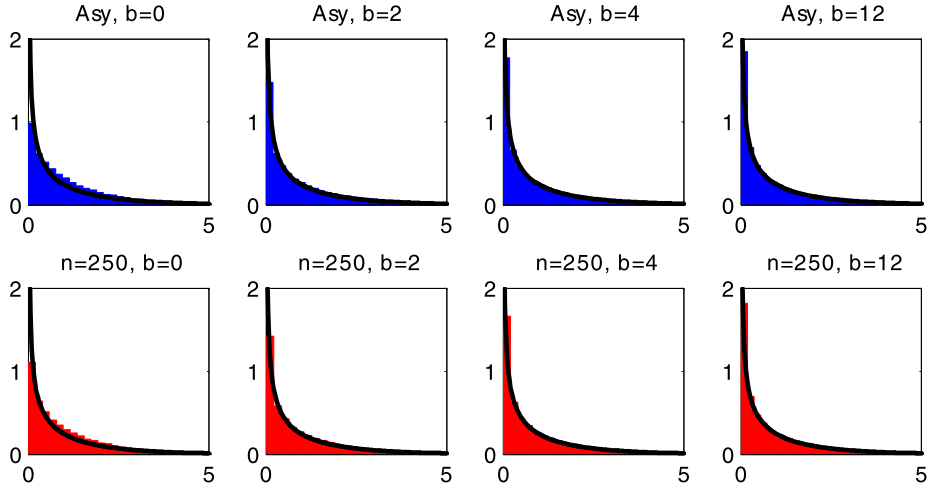

FIGURE S-18.—Asymptotic and finite-sample ($n = 250$) densities of the QLR statistic for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0.4$ and the χ^2_1 density (black line).

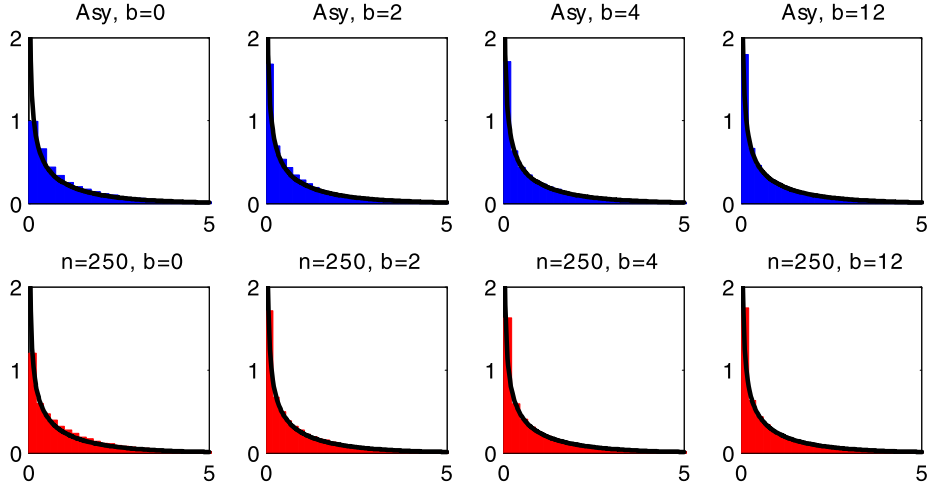

FIGURE S-19.—Asymptotic and finite-sample ($n = 250$) densities of the QLR statistic for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0.8$ and the χ^2_1 density (black line).

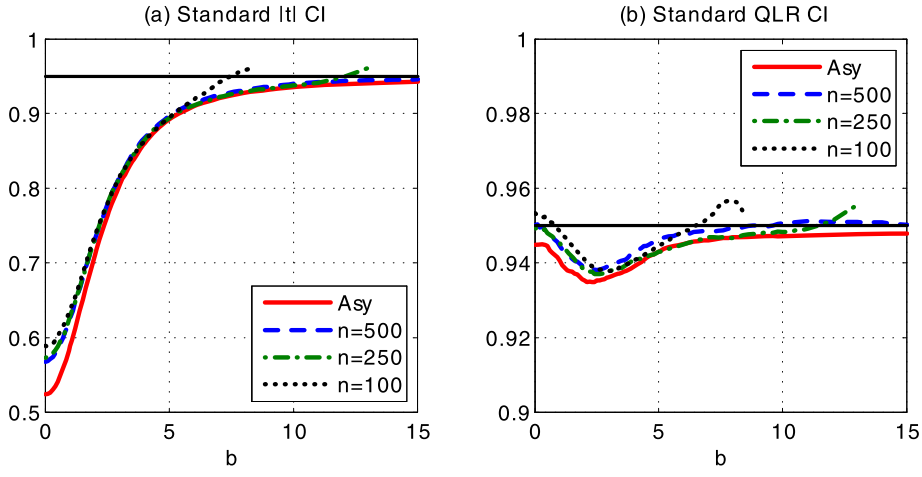

FIGURE S-20.—Coverage probabilities of standard $|t|$ and QLR CI's for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0$.

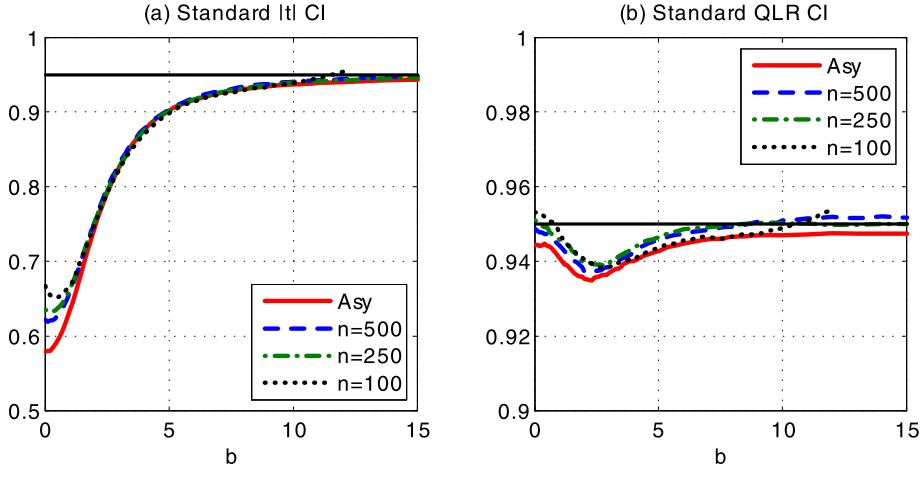

FIGURE S-21.—Coverage probabilities of standard $|t|$ and QLR CI's for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0.4$.

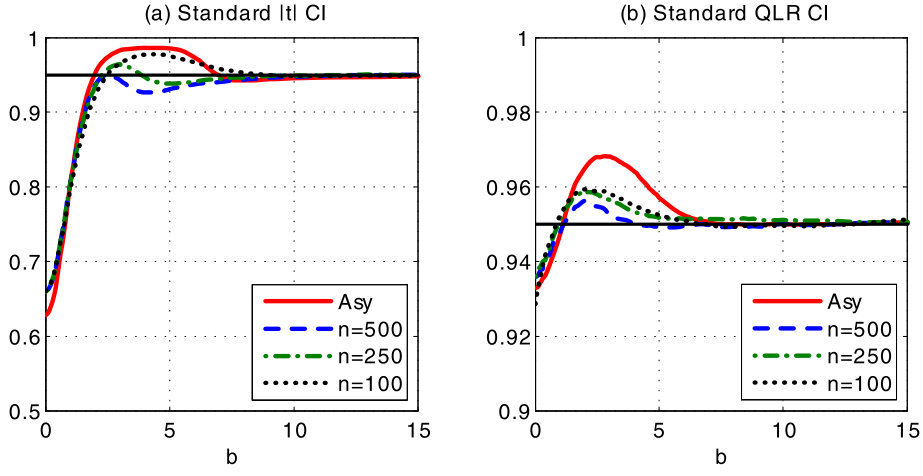

FIGURE S-22.—Coverage probabilities of standard $|t|$ and QLR CI's for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0.8$.

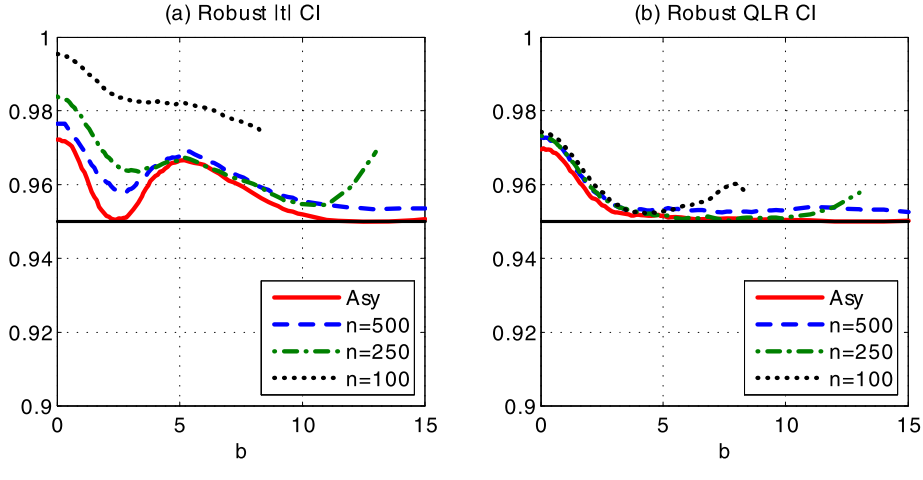

FIGURE S-23.—Coverage probabilities of robust $|t|$ and QLR CI's for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

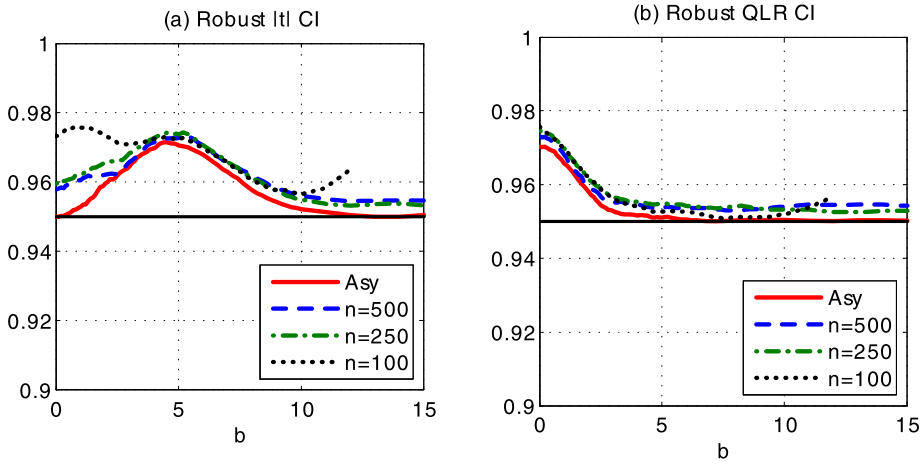

FIGURE S-24.—Coverage probabilities of robust $|t|$ and QLR CI's for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0.4$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

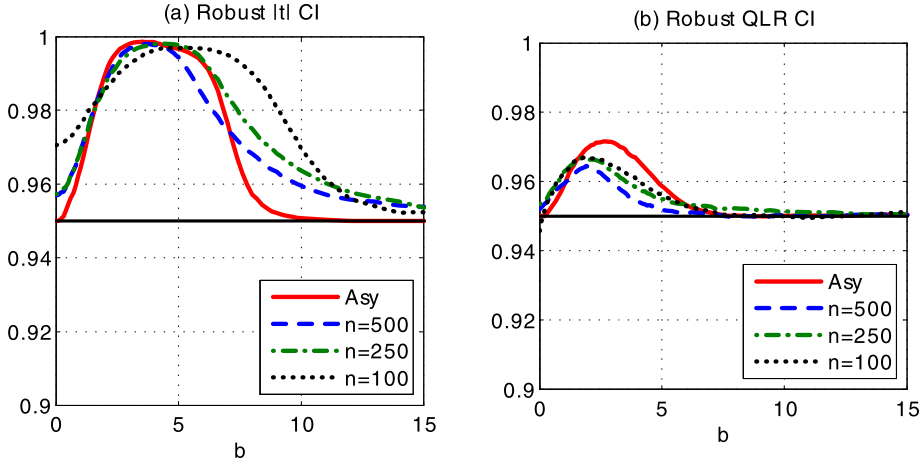

FIGURE S-25.—Coverage probabilities of robust $|t|$ and QLR CI's for the AR parameter ρ in the ARMA(1, 1) model when $\rho_0 = 0.8$, $\kappa = 1.5$, and $s(x) = \exp(-x/2)$.

TABLE S-II

ASYMPTOTIC COVERAGE PROBABILITIES (MINIMUM OVER b) OF NOMINAL 95% STANDARD CI'S FOR π AND ρ IN THE ARMA(1, 1) MODEL

π_0/ρ_0	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	Asy Size
$\|t\|$	0.523	0.527	0.534	0.552	0.578	0.612	0.642	0.643	0.627	0.523
QLR	0.935	0.933	0.933	0.934	0.935	0.936	0.940	0.941	0.933	0.933

Tables S-II to S-X provide (i) asymptotic and finite-sample coverage probabilities for $|t|$ and QLR CI's for π and ρ and (ii) FCP results for NI-LF and type 2 robust CI's for π and ρ.

Table S-II provides the minimum over b asymptotic CP's for π for a range of true π_0 values. It shows that the asymptotic size of the $|t|$ CI for π is 0.523.[75] Table S-II also shows that the undercoverage of the standard QLR CI for π is much less severe than for the $|t|$ CI. It shows that the asymptotic size of the nominal 95% standard QLR CI for π is 0.933. The results of Table S-II also apply to CI's for ρ.

Table S-III provides a summary of the finite-sample ($n = 250$) CP's of the CI's for both π and ρ based on critical values that are standard (normal or χ^2_1), NI-LF, and type 2 robust (using NI critical values and ICS statistic A_n). The standard $|t|$ CI's undercover considerably. The standard QLR CI's only undercover by a small amount. The NI-LF $|t|$ CI's overcover by a small amount.

TABLE S-III

FINITE-SAMPLE COVERAGE PROBABILITIES (MINIMUM OVER b) OF NOMINAL 95% CI'S FOR π AND ρ IN THE ARMA(1, 1) MODEL, $n = 250$

		$\|t\|$			QLR		
		Std	LF	Rob	Std	LF	Rob
MA	$\pi_0 = 0.0$	0.569	0.965	0.952	0.937	0.951	0.951
	$\pi_0 = 0.4$	0.613	0.961	0.943	0.937	0.953	0.951
	$\pi_0 = 0.7$	0.673	0.962	0.930	0.944	0.953	0.946
AR	$\rho_0 = 0.0$	0.573	0.967	0.955	0.937	0.952	0.950
	$\rho_0 = 0.4$	0.632	0.966	0.953	0.939	0.954	0.953
	$\rho_0 = 0.8$	0.660	0.965	0.952	0.936	0.954	0.950

[75]This is based on a grid of π_0 values with grid size 0.05 for $|\pi_0| \leq 0.60$ and grid size 0.025 for $0.625 \leq |\pi_0| \leq 0.825$.

TABLE S-IV

FINITE-SAMPLE COVERAGE PROBABILITIES (MINIMUM OVER b) OF NOMINAL 95% CI'S FOR π AND ρ IN THE ARMA(1, 1) MODEL, $n = 100, 500$

		$\lvert t\rvert$			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

The type 2 robust $|t|$ CI's are close to 0.95 except for some undercoverage for π when $\pi_0 = 0.4$ and 0.7. The NI-LF and type 2 robust QLR CI's are quite close to 0.95.

Table S-IV provides analogous results to Table S-III, but for $n = 100$ and 500. The results for the standard CI's are very similar to those in Table S-III. The discrepancies between the CP's and 0.95 for the NI-LF and type 2 robust $|t|$ CI's are magnified for $n = 100$ and lessened for $n = 500$. The CP's for the NI-LF and type 2 robust QLR CI's are quite close to 0.95 for $n = 100$ and 500.

Table S-V provides finite-sample FCP results for the NI-LF and type 2 robust CI's for the MA parameter π for $n = 500$.[76] Table S-V shows that the $|t|$ statistic combined with the NI-LF critical value yields a CI whose FCP's are very high—close to 1.0 for most values of b and π_0. This illustrates the poor performance of NI-LF critical values when a substantial amount of size correction is required. The NI-LF critical value performs much better in terms of FCP's when combined with the QLR statistic (because much less size correc-

[76]The true values considered are $\pi_0 = 0.0$, 0.4, and 0.7 and $b = -2, -5, -10$, and $-\infty$. The null values π_{H_0} are provided in the table. They are selected so that the robust QLR CI has FCP close to 0.50 for those cases where that is possible. When $b = 0$ or $|b|$ is small, all CI's have FCP greater than 0.50 for all values of π_{H_0} in the parameter space.

TABLE S-V

FINITE-SAMPLE FALSE COVERAGE PROBABILITIES OF 95% LEAST FAVORABLE AND ROBUST $|t|$ AND QLR CI'S FOR THE MA PARAMETER π 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	$-\infty$	
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	Avg
$\|t\|$													
LF	0.97	1.00	1.00	1.00	0.93	0.96	1.00	1.00	0.76	0.99	1.00	1.00	0.97
Rob	0.95	0.78	0.56	0.90	0.91	0.64	0.49	0.49	0.68	0.57	0.44	0.44	0.65
QLR													
LF	0.68	0.51	0.55	0.52	0.88	0.52	0.55	0.55	0.59	0.53	0.54	0.53	0.58
Rob	0.67	0.50	0.51	0.49	0.89	0.50	0.51	0.51	0.62	0.51	0.51	0.51	0.56

tion is needed). The type 2 robust critical values work quite well in terms of FCP's with both the $|t|$ and QLR statistics. Overall, the type 2 robust QLR CI performs best, followed closely by the NI-LF QLR CI, followed by the type 2 robust $|t|$ CI.

Analogous results to those in Table S-V, but for the AR parameter ρ, are provided in Table S-VI. Most of the results are quite similar.

Tables S-VII–S-X provide finite-sample false coverage probabilities of robust $|t|$ and QLR CI's for π and ρ for a range of values of κ in the ARMA(1, 1) model with $n = 500$.

TABLE S-VI

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	∞	
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	Avg
$\|t\|$													
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-VII

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	$-\infty$	
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	Avg
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-VIII

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	$-\infty$	
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	Avg
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-IX

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	∞	
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	Avg
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

TABLE S-X

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	∞	
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	Avg
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

12. SUPPLEMENTAL 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.

12.1. *Nonlinear Regression Model*

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

$$Y_i = \beta^* \cdot h(X_i, \pi^*) + Z_i'\zeta^* + U_i \quad \text{for} \quad i = 1, \ldots, n, \tag{12.1}$$

where $h(X_i, \pi) \in R$ is known up to the finite-dimensional parameter $\pi \in R^{d_\pi}$. When the true value β^* is 0, (12.1) becomes a linear model and π^* is not identified.

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 w.r.t. π $\forall \pi \in \Pi$, $\forall x \in \mathcal{X}$, although the general theory of AC1 allows for continuous nonsmooth 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 $h(x, \pi)$ w.r.t. π.

The LS sample criterion function is

$$Q_n(\theta) = n^{-1} \sum_{i=1}^{n} U_i^2(\theta)/2, \quad \text{where} \quad U_i(\theta) = Y_i - \beta h(X_i, \pi) - Z_i'\zeta. \tag{12.2}$$

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.

12.2. *Parameter Space*

In this example, the random variables $\{(X_i, Z_i, U_i) : i = 1, \ldots, n\}$ are i.i.d. with true distribution $\phi^* \in \Phi^*$, where Φ^* is a compact metric space with some metric that induces weak convergence. (The results can be extended to allow for stationary and ergodic observations under suitable weak dependence conditions, such as strong mixing conditions; see Andrews and Cheng (2011a).) 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

$$\Theta^* = \mathcal{B}^* \times \mathcal{Z}^* \times \Pi^*, \quad \text{where} \quad \mathcal{B}^* = [-b_1^*, b_2^*] \subset R \tag{12.3}$$

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

$$
\begin{aligned}
(12.4)\qquad \Phi^*(\theta^*) = \Big\{ & \phi \in \Phi^* : E_\phi(U_i|X_i, Z_i) = 0 \text{ a.s.}, \\
& E_\phi(U_i^2|X_i, Z_i) = \sigma^2(X_i, Z_i) > 0 \text{ a.s.}, E_\phi\Big(\sup_{\pi\in\Pi} \|h(X_i, \pi)\|^{4+\varepsilon} \\
& + \sup_{\pi\in\Pi} \|h_\pi(X_i, \pi)\|^{4+\varepsilon} + \sup_{\pi\in\Pi} \|h_{\pi\pi}(X_i, \pi)\|^{2+\varepsilon}\Big) \le C, \\
& \|h_{\pi\pi}(X_i, \pi_1) - h_{\pi\pi}(X_i, \pi_2)\| \le M(X_i)\|\pi_1 - \pi_2\| \\
& \forall \pi_1, \pi_2 \in \Pi \text{ for some function } M(X_i), E_\phi M(X_i)^{2+\varepsilon} \le C, \\
& E_\phi|U_i|^{4+\varepsilon} \le C, E_\phi\|Z_i\|^{4+\varepsilon} \le C, \\
& P_\phi\big(a'(h(X_i, \pi_1), h(X_i, \pi_2), Z_i) = 0\big) < 1 \\
& \forall \pi_1, \pi_2 \in \Pi \text{ with } \pi_1 \ne \pi_2, \ \forall a \in R^{d_\zeta+2} \text{ with } a \ne 0, \\
& \lambda_{\min}\big(E_\phi(h(X_i, \pi), Z_i')'(h(X_i, \pi), Z_i')\big) \ge \varepsilon \ \forall \pi \in \Pi, \text{ and} \\
& \lambda_{\min}(E_\phi d_i(\pi) d_i(\pi)') \ge \varepsilon \ \forall \pi \in \Pi \Big\}
\end{aligned}
$$

for some constants $C < \infty$ and $\varepsilon > 0$, and, by definition, $d_i(\pi) = (h(X_i, \pi), Z_i, h_\pi(X_i, \pi))'$. 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 $\beta \ne 0$.

Given the definitions above, the true parameter space Γ is of the form in (2.3). Thus, Assumption B2(i) holds immediately. Assumption B2(ii) follows from the form of $\mathcal{B}^*$ given in (12.3) and the fact that Θ^* is a product space and $\Phi^*(\theta^*)$ does not depend on β^*. Assumption B2(iii) follows from the form of $\mathcal{B}^*$. 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

$$
(12.5)\qquad \Theta = \mathcal{B} \times \mathcal{Z} \times \Pi, \quad \text{where} \quad \mathcal{B} = [-b_1, b_2] \subset R
$$

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

12.3. *Criterion Function Limit Assumption*

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

$$
\begin{aligned}
(12.6)\qquad Q(\theta; \gamma_0) = {} & E_{\phi_0} U_i^2/2 + E_{\phi_0}(\beta_0 h(X_i, \pi_0) \\
& + Z_i'\zeta_0 - \beta h(X_i, \pi) - Z_i'\zeta)^2/2,
\end{aligned}
$$

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 weak LLN given the moment and smoothness conditions in $\Phi^*(\theta^*)$ in (12.3).

LEMMA 12.1: *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: (i) The centering term in Lemma 12.1 is $E_{\gamma_0} s(W_i, \theta)$, rather than $E_{\gamma_n} s(W_i, \theta)$.

(ii) The proof of Lemma 12.1 is given in Andrews and Cheng (2011a).

Next, we verify Assumption B3* given in Supplemental Appendix A, which is a set of sufficient conditions for Assumption B3(ii) and (iii). Assumption B3*(i) holds with $Q(\theta; \gamma_0)$ defined in (12.6) by the continuity of $h(x, \pi)$ in π, the moment conditions in (12.4), and the DCT. Assumption B3*(iv) and (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 $\beta_0 = 0$,

$$Q(\psi, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0) = E_{\phi_0}(\beta h(X_i, \pi) + Z_i'(\zeta_0 - \zeta))^2/2 > 0 \tag{12.7}$$

$\forall \psi \neq \psi_0, \forall \pi \in \Pi$. The inequality in (12.7) holds unless

$$P_{\phi_0}(\beta h(X_i, \pi) + Z_i'(\zeta_0 - \zeta) = 0) = 1 \tag{12.8}$$

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 (12.4). Hence, (12.8) 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 $\beta_0 \neq 0$,

$$\begin{aligned} &Q(\theta; \gamma_0) - Q(\theta_0; \gamma_0) \\ &\quad = E_{\phi_0}(\beta h(X_i, \pi) - \beta_0 h(X_i, \pi_0) + Z_i'(\zeta_0 - \zeta))^2/2 > 0 \end{aligned} \tag{12.9}$$

$\forall \theta \neq \theta_0$. The inequality in (12.9) holds unless

$$P_{\phi_0}(\beta_0 h(X_i, \pi_0) - \beta h(X_i, \pi) + Z_i'(\zeta_0 - \zeta) = 0) = 1 \tag{12.10}$$

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 (12.4), the condition $\beta_0 \neq 0$ implies that (12.10) cannot hold for any θ such that $\pi \neq \pi_0$. When $\pi = \pi_0$, (12.10) becomes

$$P_{\phi_0}((\beta_0 - \beta)h(X_i, \pi_0) + Z_i'(\zeta_0 - \zeta) = 0) = 1. \tag{12.11}$$

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 (12.4), equation (12.11) cannot hold for $(\beta, \zeta) \neq (\beta_0, \zeta_0)$. This completes the verification of Assumption B3*.

12.4. *Close to* $\beta = 0$ *Assumptions*

12.4.1. *Assumptions C1 and D1*

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

$$Q_n(\theta) = n^{-1}\sum_{i=1}^{n}\rho(W_i, \theta), \quad \text{where} \quad \rho(W_i, \theta) = U_i^2(\theta)/2 \quad \text{and} \quad W_i = (Y_i, X_i, Z_i')'. \tag{12.12}$$

In consequence, we verify Assumptions C1 and D1 by verifying Assumption Q1 of Supplemental Appendix A. The latter is sufficient for the Assumptions C1 and D1 by Lemma 8.6 of Supplemental Appendix A (given Assumptions B1 and B2).

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

$$\begin{aligned}&\rho_\psi(W_i, \theta) = -U_i(\theta)d_{\psi,i}(\pi) \quad \text{and}\\ &\rho_{\psi\psi}(W_i, \theta) = d_{\psi,i}(\pi)d_{\psi,i}(\pi)', \quad \text{where}\\ &d_{\psi,i}(\pi) = (h(X_i, \pi), Z_i')'.\end{aligned} \tag{12.13}$$

Thus, by Lemma 8.6, we verify that Assumption C1 holds with

$$\begin{aligned}&D_\psi Q_n(\theta) = -n^{-1}\sum_{i=1}^{n}U_i(\theta)d_{\psi,i}(\pi) \quad \text{and}\\ &D_{\psi\psi}Q_n(\theta) = n^{-1}\sum_{i=1}^{n}d_{\psi,i}(\pi)d_{\psi,i}(\pi)'.\end{aligned} \tag{12.14}$$

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

$$\begin{aligned}&\rho_\theta(W_i, \theta) = -U_i(\theta)B(\beta)d_i(\pi),\\ &\rho_{\theta\theta}(W_i, \theta) = -U_i(\theta)D_i(\theta) + B(\beta)d_i(\pi)d_i(\pi)'B(\beta), \quad \text{where}\\ &d_i(\pi) = (h(X_i, \pi), Z_i', h_\pi(X_i, \pi)')',\\ &D_i(\theta) = \begin{bmatrix} 0 & 0_{1\times d_\zeta} & h_\pi(X_i, \pi)' \\ 0_{d_\zeta\times 1} & 0_{d_\zeta\times d_\zeta} & 0_{d_\zeta\times d_\pi} \\ h_\pi(X_i, \pi) & 0_{d_\pi\times d_\zeta} & h_{\pi\pi}(X_i, \pi)\beta \end{bmatrix},\end{aligned} \tag{12.15}$$

and $B(\beta)$ depends on β, not $\|\beta\|$, because β is a scalar. Hence, by Lemma 8.6, we verify that Assumption D1 holds with

$$DQ_n(\theta) = -n^{-1}\sum_{i=1}^{n} U_i(\theta)B(\beta)d_i(\pi), \tag{12.16}$$

$$D^2Q_n(\theta) = n^{-1}\sum_{i=1}^{n}(B(\beta)d_i(\pi)d_i(\pi)'B(\beta) - U_i(\theta)D_i(\theta))$$

by Lemma 8.6 in Supplemental Appendix A.[77]

Now, verify Assumption Q1. Assumption Q1(i) and (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 (12.13), verification of Assumption Q1(iv) is equivalent to showing the stochastic equicontinuity (SE) of $n^{-1}\sum_{i=1}^{n} U_i(\theta) \times h_\pi(X_i, \pi)/\beta_n$, $n^{-1}\sum_{i=1}^{n} U_i(\theta)h_{\pi\pi}(X_i, \pi) \times \beta/\beta_n^2$, and $n^{-1}\sum_{i=1}^{n} B(\beta/\beta_n) \times d_i(\pi)d_i(\pi)'B(\beta/\beta_n)$ over $\theta \in \Theta_n(\delta_n)$. We now show the SE of these three terms under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$.

The first term is

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n} U_i(\theta)h_\pi(X_i, \pi)/\beta_n \\ &= \left(n^{-1/2}\sum_{i=1}^{n} U_i h_\pi(X_i, \pi)\right)/(n^{1/2}\beta_n) \\ &\quad + \left(n^{-1}\sum_{i=1}^{n} h(X_i, \pi_n)h_\pi(X_i, \pi)\right) \\ &\quad - \left(n^{-1}\sum_{i=1}^{n} h(X_i, \pi)h_\pi(X_i, \pi)\right)\beta/\beta_n \\ &\quad + n^{-1}\sum_{i=1}^{n} Z_i'(\zeta_n - \zeta)h_\pi(X_i, \pi)/\beta_n. \end{aligned} \tag{12.17}$$

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 $\|\psi - \psi_n\| \le \delta_n|\beta_n|$ and $\delta_n \to 0$. Hence, under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, the

[77]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 $d_i(\beta, \pi) = (h(X_i, \pi), Z_i', \mathrm{sgn}(\beta)h_\pi(X_i, \pi)')'$. The appearance of $\mathrm{sgn}(\beta)$ 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.

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) $n^{-1/2}\sum_{i=1}^n U_i \times h_\pi(X_i,\pi)$ for $\pi\in\Pi$, (ii) $n^{-1}\sum_{i=1}^n h(X_i,\pi)h_\pi(X_i,\overline{\pi})$ for $(\pi,\overline{\pi})\in\Pi\times\Pi$, and (iii) $n^{-1}\sum_{i=1}^n Z_i h_\pi(X_i,\pi)'$ for $\pi\in\Pi$. 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 (12.4), and the compactness of Π. The SE of (ii) and (iii) follows from Lemma 12.1.

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 $n^{-1/2}\sum_{i=1}^n U_i h_{\pi\pi}(X_i,\pi)$ for $\pi\in\Pi$ (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 (12.4). 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 12.1.

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 12.1 using $|\beta/\beta_n| = 1 + o(1)$. This completes the verification of Assumption Q1.

12.4.2. *Assumption C2*

Assumption C2(i) holds in this example with

$$m(W_i,\theta) = -U_i(\theta)d_{\psi,i}(\pi). \tag{12.18}$$

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 $E_{\gamma^*}m(W_i,\psi^*,\pi) = -E_{\gamma^*}(U_i + \beta^*h(X_i,\pi^*) - \beta^*h(X_i,\pi))(h(X_i,\pi),Z_i')' = 0\ \forall\pi\in\Pi$ when $\beta^* = 0$.

12.4.3. *Assumption C3*

To verify Assumption C3, we have

$$U_i(\psi_{0,n},\pi) = Y_i - Z_i'\zeta_n = U_i + \beta_n h(X_i,\pi_n), \tag{12.19}$$

$$G_n(\pi) = -n^{-1/2}\sum_{i=1}^n \big(U_i d_{\psi,i}(\pi) + \beta_n[h(X_i,\pi_n)d_{\psi,i}(\pi) - E_{\phi_n}h(X_i,\pi_n)d_{\psi,i}(\pi)]\big).$$

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 rowwise 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.

12.4.4. *Assumption C4*

Assumption C4(i) holds in this example with

$$H(\pi;\gamma_0) = E_{\phi_0} d_{\psi,i}(\pi) d_{\psi,i}(\pi)' \tag{12.20}$$

by applying a uniform LLN for drifting true distributions, specifically, Lemma 12.1, 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 (12.4).

12.4.5. *Assumption C5*

To verify Assumption C5(i), we have

$$\begin{aligned} K_n(\theta;\gamma^*) &= \frac{\partial}{\partial\beta^*} E_{\phi^*} m(W_i,\theta) \\ &= -\frac{\partial}{\partial\beta^*} E_{\phi^*}(Y_i - \beta h(X_i,\pi) - Z_i'\zeta) d_{\psi,i}(\pi) \\ &= -\frac{\partial}{\partial\beta^*} E_{\phi^*}(U_i + \beta^* h(X_i,\pi^*) \\ &\quad - \beta h(X_i,\pi) - Z_i'(\zeta - \zeta^*)) d_{\psi,i}(\pi) \\ &= -E_{\phi^*} h(X_i,\pi^*) d_{\psi,i}(\pi). \end{aligned} \tag{12.21}$$

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

$$K(\pi;\gamma_0) = K(\psi_0,\pi;\gamma_0) = -E_{\phi_0} h(X_i,\pi_0) d_{\psi,i}(\pi). \tag{12.22}$$

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 Andrews and Cheng (2011a), 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 $E_{\phi_n} h(X_i,\pi_1) \times d_{\psi,i}(\pi_2) \to E_{\phi_0} h(X_i,\pi_1) d_{\psi,i}(\pi_2)$ 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 $\phi \in \Phi^*$, that is, $\sup_{\phi\in\Phi^*} E_\phi \|h(X_i,\pi_1) d_{\psi,i}(\pi_2)\|^{1+\delta} \le C < \infty$ (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, that is, $E_{\phi^*}\sup_{(\pi_1,\pi_2)\in\Pi\times\Pi}(\|h_\pi(X_i,\pi_1)' d_{\psi,i}(\pi_2)\| + \|h(X_i,\pi_1)(\partial d_{\psi,i}(\pi_2)/\partial\pi')\|) \le C$ for some $C < \infty$ for all $\phi^* \in \Phi^*(\theta^*)$.

12.4.6. *Assumption C6*

Next, we verify Assumption C6**, which implies Assumption C6 by Lemma 8.5 in Supplemental Appendix A. Assumption C6**(i) holds because β is a scalar. By the discussion following (12.19), $a'(G_1(\pi_1), G_1(\pi_2), G_2)$ 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 (12.4), $P_{\phi_0}(d_a(\pi_1, \pi_2) = 0) < 1$ $\forall a \in R^{d_\zeta + 2}$ 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.

12.4.7. *Assumption C7*

We verify Assumption C7 as follows. Given the form of $H(\pi; \gamma_0)$ and $K(\pi; \gamma_0)$ in (12.20) and (12.22), respectively, we have

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

where the inequality holds by the matrix Cauchy–Schwarz inequality in Tripathi (1999). The inequality 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 inequality 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 (12.4). This completes the verification of Assumption C7.

12.4.8. *Assumption C8*

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

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

by the form of $D_\psi Q_n(\theta_n)$ given in (12.14). 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 (12.4).

12.5. *Distant From* $\beta = 0$ *Assumptions*

12.5.1. *Assumption D2*

To verify Assumption D2 with $D^2 Q_n(\theta)$ given in (12.16), we have

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

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $n^{-1} \sum_{i=1}^{n} d_i(\pi_n) d_i(\pi_n)' \to_p E_{\phi_0} d_i(\pi_0) d_i(\pi_0)'$ because $n^{-1} \sum_{i=1}^{n} d_i(\pi) d_i(\pi)' \to_p E_{\phi_0} d_i(\pi) d_i(\pi)'$ uniformly over $\pi \in \Pi$ by Lemma 12.1 (stated earlier in this appendix) and the continuity of $E_{\phi_0} d_i(\pi) \times d_i(\pi)'$ in π. The second line of (12.25) is $o_p(1)$ because $n^{1/2} |\beta_n| \to \infty$, $n^{-1/2} \sum_{i=1}^{n} U_i h_\pi(X_i, \pi_n)' = O_p(1)$, 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 rowwise i.i.d. random variables under the moment conditions in (12.4). Hence, Assumption D2 holds with the matrix

$$
(12.26)\quad J(\gamma_0) = E_{\phi_0} d_i(\pi_0) d_i(\pi_0)',
$$

which is nonsingular by the conditions in (12.4).

12.5.2. *Assumption D3*

To verify Assumption D3 in this example, we have

$$
\begin{aligned}
(12.27)\quad n^{1/2} B^{-1}(\beta_n) D Q_n(\theta_n) &= -n^{-1/2} \sum_{i=1}^{n} U_i d_i(\pi_n) \\
&\to_d N(0_{d_\theta}, V(\gamma_0)), \quad \text{where} \\
V(\gamma_0) &= E_{\phi_0} U_i^2 d_i(\pi_0) d_i(\pi_0)'.
\end{aligned}
$$

The convergence in distribution holds by the CLT for a triangular array of rowwise i.i.d. random variables. Assumption D3(ii) holds because $E_{\phi_0} d_i(\pi_0) d_i(\pi_0)'$ is nonsingular and $E_{\phi_0}(U_i^2 | X_i, Z_i) > 0$ a.s. by (12.4).

12.6. *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 3 of AC1 are

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

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

12.7. *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

$$
\begin{aligned}
(12.29)\quad \widehat{J}_n(\theta) &= n^{-1} \sum_{i=1}^n d_i(\pi) d_i(\pi)', \\
\widehat{V}_n(\theta) &= n^{-1} \sum_{i=1}^n U_i^2(\theta) d_i(\pi) d_i(\pi)' \\
&= n^{-1} \sum_{i=1}^n U_i^2 d_i(\pi) d_i(\pi)' \\
&\quad + 2n^{-1} \sum_{i=1}^n U_i \big[\beta_n h(X_i, \pi_n) - \beta h(X_i, \pi) \\
&\quad + (\zeta_n - \zeta)' Z_i \big] d_i(\pi) d_i(\pi)' \\
&\quad + n^{-1} \sum_{i=1}^n \big[\beta_n h(X_i, \pi_n) - \beta h(X_i, \pi) \\
&\quad + (\zeta_n - \zeta)' Z_i \big]^2 d_i(\pi) d_i(\pi)'.
\end{aligned}
$$

These variance matrix estimators are used to construct t and Wald statistics, and also to construct the identification-category-selection statistic A_n in (5.3) of AC1.

Assumption V1(i) (scalar β) holds with

$$J(\theta;\gamma_0) = E_{\phi_0} d_i(\pi) d_i(\pi)', \tag{12.30}$$
$$V(\theta;\gamma_0) = E_{\phi_0} U_i^2 d_i(\pi) d_i(\pi)' + E_{\phi_0}\big[\beta_0 h(X_i,\pi_0) - \beta h(X_i,\pi) + (\zeta_0 - \zeta)' Z_i\big]^2 d_i(\pi) d_i(\pi)'$$

by Lemma 12.1, using the conditions in (12.4). Assumption V1(ii) holds by the continuity of $h(x,\pi)$ and $h_\pi(x,\pi)$ in π and the moment conditions in (12.4).

The quantity $\Sigma(\pi;\gamma_0)$ in (4.4) takes the form

$$\Sigma(\pi;\gamma_0) = (E_{\phi_0} d_i(\pi) d_i(\pi)')^{-1} E_{\phi_0} U_i^2 d_i(\pi) d_i(\pi)' (E_{\phi_0} d_i(\pi) d_i(\pi)')^{-1}. \tag{12.31}$$

Given this, Assumption V1(iii) holds by the nonsingularity conditions in (12.4).

Assumption V1(i) and (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 3.3 of AC1, imply that Assumption V2 holds.

12.8. *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 (12.1). 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 (12.1) and, for simplicity, suppose no $Z_i'\zeta$ summand appears:

$$Y_i = \beta \cdot h(X_i,\pi) + U_i. \tag{12.32}$$

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 (12.32), 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 moment functions

$$E_{\gamma_n}(Y_i - \beta h(X_i,\pi)) \begin{pmatrix} h(X_i,\pi) \\ h_\pi(X_i,\pi) \end{pmatrix}, \tag{12.33}$$

which equal 0_2 when $(\beta, \pi) = (\beta_n, \pi_0)$. 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 (12.33), where their t, T, and θ correspond to our i, n, and (β, π), respectively.

SW's population moments equal

$$\begin{aligned}(12.34)\quad \widetilde{m}_T(\alpha, \beta) &= E_{\gamma_n}(Y_i - \beta h(X_i, \pi)) \begin{pmatrix} h(X_i, \pi) \\ h_\pi(X_i, \pi) \end{pmatrix} \\ &= E_{\phi_0}(\beta_n h(X_i, \pi_0) - \beta h(X_i, \pi)) \begin{pmatrix} h(X_i, \pi) \\ h_\pi(X_i, \pi) \end{pmatrix}.\end{aligned}$$

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

$$\begin{aligned}(12.35)\quad \widetilde{m}_{1T}(\alpha, \beta) &= \widetilde{m}_T(\alpha, \beta) - \widetilde{m}_T(\alpha_0, \beta) \\ &= E_{\phi_0}(\beta_n h(X_i, \pi_0) - \beta h(X_i, \pi)) \begin{pmatrix} h(X_i, \pi) \\ h_\pi(X_i, \pi) \end{pmatrix} \\ &\quad - E_{\phi_0}(\beta_n h(X_i, \pi_0) - \beta h(X_i, \pi_0)) \begin{pmatrix} h(X_i, \pi_0) \\ h_\pi(X_i, \pi_0) \end{pmatrix} \\ &= A_{1n}(\pi) + A_2(\pi, \beta),\end{aligned}$$

where

$$\begin{aligned}(12.36)\quad & A_{1n}(\pi) \\ &\quad = n^{-1/2} C \cdot E_{\phi_0} h(X_i, \pi_0) \begin{pmatrix} h(X_i, \pi) - h(X_i, \pi_0) \\ h_\pi(X_i, \pi) - h_\pi(X_i, \pi_0) \end{pmatrix}, \\ & A_2(\pi, \beta) \\ &\quad = \beta E_{\phi_0} \left[h(X_i, \pi_0) \begin{pmatrix} h(X_i, \pi_0) \\ h_\pi(X_i, \pi_0) \end{pmatrix} - h(X_i, \pi) \begin{pmatrix} h(X_i, \pi) \\ h_\pi(X_i, \pi) \end{pmatrix} \right].\end{aligned}$$

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, is 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 (12.35) and (12.36), 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.

13. SUPPLEMENTAL APPENDIX F: LIML EXAMPLE

In this example, we consider a linear IV regression model estimated by the ML estimator, which is the limited information ML (LIML) estimator. We consider robust QLR-based tests concerning the coefficient π (in our notation) on the endogenous variable in the structural equation. The objective of this section is to compare the robust tests introduced in AC1 with the conditional likelihood ratio (CLR) test of Moreira (2003), the LM test of Kleibergen (2002) and Moreira (2009), and the well known Anderson–Rubin (AR in this section only) test. The CLR test is known to have approximate asymptotic optimality properties in the classes of invariant similar tests and invariant tests; see Andrews, Moreira, and Stock (2006, 2008) and Chernozhukov, Hansen, and Jansson (2009). Hence, this is a good benchmark test and model to assess the performance of the robust tests of AC1.

The asymptotic distributions of the LIML estimator and the QLR statistic, which are obtained here, also are given in Staiger and Stock (1997), Moreira (2003), and Andrews, Moreira, and Stock (2006). Hence, the point of this section is not to derive new asymptotic results, but rather to link the general results of AC1 to existing results in the literature and, more importantly, to assess the power properties of the robust tests introduced in AC1. A numerical study is conducted to compare the asymptotic power of the type 2 robust QLR test with that of the CLR, LM, and AR tests.

In short, we find that the type 2 robust test based on the NI-ICS statistic has power that is essentially equal to that of the CLR test. Hence, this robust test has approximately asymptotically optimal power in the same sense as the CLR test. The type 2 robust test based on the unrestricted ICS statistic has lower power than the CLR test in some areas of the parameter space and equal power in others.

13.1. *Key Quantities*

The structural model is

$$y_{1,i} = y_{2,i}\pi + u_i^*, \quad y_{2,i} = Z_i'\beta + v_i^*, \tag{13.1}$$

where $(u_i^*, v_i^*)' \sim N(0, \Upsilon^*)$ for a p.d. 2×2 matrix Υ^*, (u_i^*, v_i^*) and Z_i are independent, $\{(Z_i', u_i^*, v_i^*)' : i = 1, \ldots, n\}$ are i.i.d., $y_{1,i}, y_{2,i}, u_i^*, v_i^* \in R$, $Z_i \in R^k$,

$\pi \in R$, and $\beta \in R^k$.[78,79] The reduced-form equations are

$$y_{1,i} = \pi \cdot Z_i'\beta + u_i, \quad y_{2,i} = Z_i'\beta + v_i, \tag{13.2}$$

where $u_i = u_i^* + v_i^*\pi$, $v_i = v_i^*$, and $(u_i, v_i)' \sim N(0, \Upsilon)$. Note that the reparametrization between (π, Υ^*) and (π, Υ) is one-to-one and Υ is p.d.

Define $\zeta = \text{vech}(\Upsilon^{-1}) = S \cdot \text{vec}(\Upsilon^{-1}) \in R^3$, where $S \in R^{3\times 4}$ is a selector matrix.

The log-likelihood function for $\theta = (\beta, \zeta, \pi)$ multiplied by $-n^{-1}$ and ignoring a constant is

$$Q_n(\theta) = \frac{1}{2}\log|\Upsilon| + \frac{1}{2}n^{-1}\sum_{i=1}^{n}\varepsilon_i(\beta, \pi)'\Upsilon^{-1}\varepsilon_i(\beta, \pi), \quad \text{where} \tag{13.3}$$

$$\varepsilon_i(\beta, \pi) = (y_{1,i} - \pi \cdot Z_i'\beta, y_{2,i} - Z_i'\beta)' \in R^2.$$

Assumption A holds because $Q_n(\theta)$ does not depend on π when $\beta = 0$. Define $\varepsilon_i = (u_i, v_i)' = \varepsilon_i(\beta_0, \pi_0)$.

Below we verify Assumptions B1–B3, C1–C5, C7, C8, D1–D3, and RQ1–RQ3, and provide key quantities in these assumptions. We do not give all of the details of the verification, which are similar to those in the nonlinear regression example in Supplemental Appendix E.

The optimization and true parameter spaces Θ and Θ^* are $\Theta = \bigtimes_{j=1}^k [-b_{L,j}, b_{H,j}] \times \mathcal{Z} \times \Pi$ and $\Theta^* = \bigtimes_{j=1}^k [-b_{L,j}^*, b_{H,j}^*] \times \mathcal{Z}^* \times \Pi^*$, where $b_{L,j}, b_{H,j}, b_{L,j}^*, b_{H,j}^* \in R$, $0 \le b_{L,j}^* < b_{L,j}$, $0 \le b_{H,j}^* < b_{H,j}$, $b_{L,j}^*$ and $b_{H,j}^*$ are not both 0 for $j = 1, \ldots, k$, $\mathcal{Z}^* \subset \text{int}(\mathcal{Z}) \subset \{\zeta \in R^3 : \zeta = \text{vech}(A)$ for some 2×2 symmetric p.d. matrix $A\}$, $\Pi^* \subset \text{int}(\Pi) \subset R$, and $\mathcal{Z}^*$, $\mathcal{Z}$, Π^*, and Π are compact. Let ϕ denote the distribution of Z_i $\forall i \ge 1$. The true parameter space for $\gamma = (\theta, \phi)$ is

$$\Gamma = \{\gamma = (\theta, \phi) : \theta \in \Theta^*, \phi \in \Phi^*\}, \tag{13.4}$$

where Φ^* is some compact subset of Φ w.r.t. the metric d_Φ and $\Phi = \{\phi : E_\phi Z_i Z_i' = I_k\}$, where d_Φ is some metric on the space of distributions on

[78]We use the notation of AC1 in which the parameters (β, π) are reversed from the usual notation in the literature. The reason is that, in AC1, the parameter β is the parameter that determines the strength of identification of the parameter π.

[79]For simplicity, we consider a model without exogenous variables X_i in either equation. As is well known, such variables can be projected out and the results given here apply with Z_i being viewed as the projection residual; for example, see Section 2 of Andrews, Moreira, and Stock (2006) and consider a population projection in place of a sample projection. Provided X_i includes an intercept, this yields that Z_i has mean zero. Also for simplicity, we assume the errors are normally distributed. The results can be extended to nonnormal finite variance errors, provided (u_i^*, v_i^*) is symmetrically distributed or the instruments have mean zero. By the discussion above, the latter is not restrictive.

R^k that induces weak convergence.[80] With these definitions, Assumptions B1 and B2 hold.

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

$$
\begin{aligned}
(13.5)\qquad Q(\theta; \gamma_0) &= \frac{1}{2}(\log|Y| + E_{\gamma_0}\varepsilon_i(\beta, \pi)'Y^{-1}\varepsilon_i(\beta, \pi)) \\
&= \frac{1}{2}(\log|Y| + \operatorname{trace}(Y^{-1}Y_0) + \Delta(\beta, \pi; \gamma_0)), \quad \text{where} \\
\Delta(\beta, \pi; \gamma_0) &= E_{\gamma_0}\delta_i(\beta, \pi; \gamma_0)'Y^{-1}\delta_i(\beta, \pi; \gamma_0) \geq 0, \\
\delta_i(\beta, \pi; \gamma_0) &= \begin{pmatrix} \pi_0 Z_i'\beta_0 \\ Z_i'\beta_0 \end{pmatrix} - \begin{pmatrix} \pi Z_i'\beta \\ Z_i'\beta \end{pmatrix}.
\end{aligned}
$$

Because Y is p.d. and $Z_i'\beta = 0$ a.s. if and only if $\beta = 0$, we have (i) when $\beta_0 = 0$, $\forall \pi \in \Pi$, $\delta_i(\beta, \pi; \gamma_0) = 0$ if and only if $\beta = 0$ and (ii) when $\beta_0 \neq 0$, $\delta_i(\beta, \pi; \gamma_0) = 0$ if and only if $(\beta, \pi) = (\beta_0, \pi_0)$. For any $\theta \in \Theta$,

$$
(13.6)\qquad \frac{\partial}{\partial Y^{-1}}Q(\theta; \gamma_0) = \frac{1}{2}(-Y + Y_0) \quad \text{and} \quad \frac{\partial^2}{\partial^2 Y^{-1}}Q(\theta; \gamma_0) = I_2 \otimes I_2.
$$

Hence, $Q(\theta; \gamma_0)$ is minimized at $\zeta = \operatorname{vech}(Y_0^{-1})$ for any β and π. In consequence, Assumption B3 is verified using Assumption B3* and Lemma 8.1 in Supplemental Appendix A.

Denote the first derivative of $\varepsilon_i(\beta, \pi)$ w.r.t. β as

$$
(13.7)\qquad q_{\beta,i}(\pi) = -(\pi Z_i, Z_i)' = -(\pi, 1)' \otimes Z_i' \in R^{2\times k}.
$$

Note that $E_{\gamma_0}q_{\beta,i}(\pi_1)'Y_0^{-1}q_{\beta,i}(\pi_2) = a(\pi_1)'Y_0^{-1}a(\pi_2)I_k$, where $a(\pi) = (\pi, 1)' \in R^k$.

Assumption C1 is verified with

$$
\begin{aligned}
(13.8)\qquad D_\beta Q_n(\theta) &= n^{-1}\sum_{i=1}^n q_{\beta,i}(\pi)'Y^{-1}\varepsilon_i(\beta, \pi) \in R^k, \\
D_{\beta\beta}Q_n(\theta) &= n^{-1}\sum_{i=1}^n q_{\beta,i}(\pi)'Y^{-1}q_{\beta,i}(\pi) \in R^{k\times k}, \\
D_\zeta Q_n(\theta) &= \frac{1}{2}\operatorname{vech}\left(-Y + n^{-1}\sum_{i=1}^n \varepsilon_i(\beta, \pi)\varepsilon_i(\beta, \pi)'\right) \in R^3, \\
D_{\zeta\zeta}Q_n(\theta) &= \frac{1}{2}S \cdot (Y \otimes Y) \cdot S' \in R^{3\times 3},
\end{aligned}
$$

[80]There is no loss of generality in assuming $E_\phi Z_i Z_i' = I_k$ because β and Z_i in the original model can be reparametrized as $\beta^* = (E_\phi Z_i Z_i')^{1/2}\beta$ and $Z_i^* = (E_\phi Z_i Z_i')^{-1/2}Z_i$.

$$D_{\beta\zeta}Q_n(\theta) = n^{-1}\sum_{i=1}^{n} \varepsilon_i(\beta, \pi)' \otimes q_{\beta,i}(\pi)' \cdot S' \in R^{k\times 3}.$$

Assumption C1 is verified using the sufficient condition Assumption Q1 and Lemma 8.6 in Supplemental Appendix A. Assumption Q1 holds by a uniform LLN.

Assumption C2 holds with

$$m(W_i, \theta) = \begin{pmatrix} q_{\beta,i}(\pi)'\Upsilon^{-1}\varepsilon_i(\beta, \pi) \\ \frac{1}{2}\operatorname{vech}(-\Upsilon + \varepsilon_i(\beta, \pi)\varepsilon_i(\beta, \pi)') \end{pmatrix} \in R^{k+3} \tag{13.9}$$

because $\forall \pi \in \Pi$, $\varepsilon_i(0, \pi) = \varepsilon_i$ when $\beta_0 = 0$ and $\varepsilon_i \sim N(0, \Upsilon)$.

Assumption C3 holds with

$$G_n(\pi) = n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} q_{\beta,i}(\pi)'\Upsilon_n^{-1}\varepsilon_i - E_{\gamma_n} q_{\beta,i}(\pi)'\Upsilon_n^{-1}\varepsilon_i \\ \frac{1}{2}\operatorname{vech}(\varepsilon_i\varepsilon_i' - E_{\gamma_n}\varepsilon_i\varepsilon_i') \end{pmatrix}. \tag{13.10}$$

The weak convergence of the empirical process $\{G_n(\pi) : \pi \in \Pi\}$ is straightforward because $q_{\beta,i}(\pi)' = -(\pi, 1) \otimes Z_i$. The limit process $\{G(\pi; \gamma_0) : \pi \in \Pi\}$ in Assumption C3 is the mean zero Gaussian process with covariance kernel

$$\Omega(\pi_1, \pi_2; \gamma_0) = \begin{pmatrix} a(\pi_1)'\Upsilon_0^{-1}a(\pi_2)I_k & 0_{k\times 3} \\ 0_{3\times k} & \Omega_{\zeta\zeta}(\gamma_0) \end{pmatrix}, \quad \text{where} \tag{13.11}$$

$$\Omega_{\zeta\zeta}(\gamma_0) = \frac{1}{4}S \cdot \operatorname{Var}_{\gamma_0}(\varepsilon_i \otimes \varepsilon_i) \cdot S' = \frac{1}{4}S(I_4 + K_4)(\Upsilon_0 \otimes \Upsilon_0)S',$$

$I_4 \in R^{4\times 4}$ is the identity matrix, and $K_4 \in R^{4\times 4}$ is the communication matrix that transforms $\operatorname{vec}(A)$ to $\operatorname{vec}(A')$ for any $A \in R^{4\times 4}$. The equalities for $\Omega_{\zeta\zeta}(\gamma_0)$ hold by Theorem 4.3(iv) of Magnus and Neudecker (1979). In (13.11), the off-diagonal elements are zeros because the bivariate normal distribution is symmetric around 0.[81]

Assumption C4 holds with

$$H(\pi; \gamma_0) = \begin{pmatrix} a(\pi)'\Upsilon_0^{-1}a(\pi)I_k & 0_{k\times 3} \\ 0_{3\times k} & \frac{1}{2}S \cdot (\Upsilon_0 \otimes \Upsilon_0) \cdot S' \end{pmatrix} \tag{13.12}$$

by a uniform LLN, where the off-diagonal elements are zeros because $\varepsilon_i(0, \pi) = \varepsilon_i$ when $\beta_0 = 0$.

[81]Alternatively, the off-diagonal elements are zeros if $EZ_i = 0$ and ε_i has a nonsymmetric distribution.

To verify Assumption C5, note that

$$E_{\gamma_0} m(W_i, \theta) = \begin{pmatrix} E_{\gamma_0} q_{\beta,i}(\pi)' \Upsilon^{-1} \varepsilon_i(\beta, \pi) \\ \frac{1}{2} \operatorname{vech}(-\Upsilon + E_{\gamma_0} \varepsilon_i(\beta, \pi) \varepsilon_i(\beta, \pi)') \end{pmatrix} \in R^{k+3}, \tag{13.13}$$

$$K_n(\theta; \gamma_0) = \begin{pmatrix} -E_{\gamma_0} q_{\beta,i}(\pi)' \Upsilon^{-1} q_{\beta,i}(\pi_0) \\ \frac{1}{2} S \cdot E_{\gamma_0}(\varepsilon_i(\beta, \pi) \otimes I_2) q_{\beta,i}(\pi_0) \end{pmatrix} \in R^{(k+3)\times k},$$

where the second equality uses $(\partial/\partial A)(AA') = A \otimes I_2$ for $A \in R^2$. Assumption C5 holds with

$$K(\pi; \gamma_0) = \begin{pmatrix} -a(\pi)' \Upsilon_0^{-1} a(\pi_0) I_k \\ 0_{3\times k} \end{pmatrix} \in R^{(k+3)\times k}, \tag{13.14}$$

where the second element is zero because $\varepsilon_i(0, \pi) = \varepsilon_i$ when $\beta_0 = 0$.

Assumption C6 is not needed in deriving the asymptotic null distributions of the QLR statistic for π and the null-imposed ICS statistic.[82] Assumption C7 holds by the matrix Cauchy–Schwarz inequality because $K(\pi; \gamma_0)' H^{-1}(\pi; \gamma_0) K(\pi; \gamma_0) = a(\pi)' \Upsilon_0^{-1} a(\pi_0) [a(\pi)' \Upsilon_0^{-1} a(\pi)]^{-1} a(\pi)' \Upsilon_0^{-1} a(\pi_0) I_k$. Assumption C8 follows from the switch of E and ∂ and a uniform LLN.

Define

$$\begin{aligned} q_{\pi,i}(\beta) &= \frac{\partial}{\partial \pi} \varepsilon_i(\beta, \pi) = -(Z_i' \beta, 0^{k\times 1})' \in R^2, \\ q_{\beta\pi,i} &= -(Z_i, 0^{k\times 1})' \in R^{2\times k}, \\ q_i(\omega) &= q_{\pi,i}(\beta)/\|\beta\| = -(Z_i' \omega, 0)' \in R^2. \end{aligned} \tag{13.15}$$

Assumption D1 holds with the partial derivatives in (13.8) plus

$$\begin{aligned} D_\pi Q_n(\theta) &= n^{-1} \sum_{i=1}^n q_{\pi,i}(\pi)' \Upsilon^{-1} \varepsilon_i(\beta, \pi) \in R, \\ D_{\pi\pi} Q_n(\theta) &= n^{-1} \sum_{i=1}^n q_{\pi,i}(\pi)' \Upsilon^{-1} q_{\pi,i}(\pi) \in R, \\ D_{\beta\pi} Q_n(\theta) &= n^{-1} \sum_{i=1}^n (q_{\beta\pi,i}' \Upsilon^{-1} \varepsilon_i(\beta, \pi) + q_{\beta,i}(\pi)' \Upsilon^{-1} q_{\pi,i}(\beta)) \in R^k, \\ D_{\zeta\pi} Q_n(\theta) &= n^{-1} \sum_{i=1}^n S \cdot (\varepsilon_i(\beta, \pi) \otimes I_2) q_{\pi,i}(\pi) \in R^3. \end{aligned} \tag{13.16}$$

[82] If the ICS statistic involves an unrestricted estimator, we assume Assumption C6 holds.

Assumption D1 is verified using the sufficient condition Assumption Q1 and Lemma 8.6 in Supplemental Appendix A.

Assumption D2 holds with

$$J(\gamma_0) = \begin{pmatrix} E_{\gamma_0} q_{\beta,i}(\pi_0)' Y_0^{-1} q_{\beta,i}(\pi_0) & 0_{k\times 3} & E_{\gamma_0} q_{\beta,i}(\pi_0)' Y_0^{-1} q_i(\omega_0) \\ 0_{3\times k} & \frac{1}{2} S \cdot (Y_0 \otimes Y_0) \cdot S' & 0_3 \\ E_{\gamma_0} q_i(\omega_0)' Y_0^{-1} q_{\beta,i}(\pi_0) & 0_3' & E_{\gamma_0} q_i(\omega_0)' Y_0^{-1} q_i(\omega_0) \end{pmatrix}, \tag{13.17}$$

where the zero elements follow from $\varepsilon_i(\beta_0, \pi_0) = \varepsilon_i$. Assumption D3 holds with $V(\gamma_0)$ equal to $J(\gamma_0)$ except that $\frac{1}{2} S \cdot (Y_0 \otimes Y_0) \cdot S'$ is replaced by $\frac{1}{4} S(I_4 + K_4)(Y_0 \otimes Y_0) S'$. Because $H(\pi; \gamma_0)$ and $J(\gamma_0)$ are block diagonal, the first- and second-order derivatives of $Q_n(\theta)$ w.r.t. ζ do not effect the asymptotic distributions of the estimators and the QLR statistic for π.

We consider the QLR test and CI's involving π. In consequence, Assumption RQ2(ii) holds for the QLR statistic with $\widehat{s}_n = 1$ and the standard critical value is $\chi^2_{1,1-\alpha}$. Assumptions RQ1 and RQ3 hold automatically.

13.2. *Asymptotic Distributions of the Statistics*

Let $\mathrm{QLR}_n(\pi_{H_0})$ denote the QLR statistic for the null hypothesis $H_0 : \pi = \pi_{H_0}$, where π_{H_0} may be different from the limit π_0 of the true value of π.

Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $b \in R^k$, the asymptotic distribution of $\mathrm{QLR}_n(\pi_{H_0})$ is the distribution of

$$\begin{aligned} &\mathrm{QLR}(h, \pi_{H_0}) = 2\Big(\xi(\pi_{H_0}; \gamma_0, b) - \inf_{\pi \in \Pi} \xi(\pi; \gamma_0, b)\Big), \quad \text{where} \\ &\xi(\pi; \gamma_0, b) \\ &\quad = -\frac{(G_\beta(\pi; \gamma_0) - a(\pi)' Y_0^{-1} a(\pi_0) b)' (G_\beta(\pi; \gamma_0) - a(\pi)' Y_0^{-1} a(\pi_0) b)}{2 a(\pi)' Y_0^{-1} a(\pi)}, \\ &G_\beta(\pi; \gamma_0) = \big(a(\pi)' Y_0^{-1/2} \eta\big)' \in R^k, \\ &\eta = (\eta_1, \ldots, \eta_k) \in R^{2\times k}, \quad \eta_j \sim N(0, I_2) \quad \text{are i.i.d.} \\ &\quad \text{for} \quad j = 1, \ldots, k. \end{aligned} \tag{13.18}$$

By construction, $\{G_\beta(\pi; \gamma_0) : \pi \in \Pi\}$ is a Gaussian process with covariance kernel $a(\pi_1)' Y_0^{-1} a(\pi_2) I_k$. Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $\mathrm{QLR}_n(\pi_{H_0}) \sim \chi^2_1$ when $\pi_{H_0} = \pi_0$.

The null-imposed ICS statistic is[83]

$$A_n(\pi_{H_0}) = (n\widehat{\beta}_n(\pi_{H_0})'\widehat{\Sigma}^{-1}_{\beta\beta,n}(\pi_{H_0})\widehat{\beta}_n(\pi_{H_0})/k)^{1/2}, \quad \text{where} \tag{13.19}$$

$$\widehat{\Sigma}_{\beta\beta,n}(\pi_{H_0}) = \left(a(\pi_{H_0})'\widehat{Y}_n^{-1}(\pi_{H_0})a(\pi_{H_0})n^{-1}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1},$$

$$\widehat{Y}_n(\pi_{H_0}) = n^{-1}\sum_{i=1}^{n} \varepsilon_i(\widehat{\beta}_n(\pi_{H_0}), \pi_{H_0})\varepsilon_i(\widehat{\beta}_n(\pi_{H_0}), \pi_{H_0})'.$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $b \in R^k$, $\widehat{Y}_n(\pi_{H_0}) \to_p Y_0$ by a uniform LLN, $\widehat{\beta}_n(\pi_{H_0}) \to_p 0$, and $\varepsilon_i(0, \pi)$ does not depend on π. Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $\widehat{Y}_n(\pi_{H_0}) \to_p Y_0$ when $\pi_{H_0} = \pi_0$ by a uniform LLN and $\widehat{\beta}_n(\pi_0) \to_p \beta_0$. This replaces the verification of Assumptions V1 (vector β) and V2 for the type 2 robust QLR test and CI because the asymptotic variance of $n^{1/2}(\widehat{\beta}_n(\pi_{H_0}) - \beta_n)$ is $(a(\pi_{H_0})'Y_0^{-1}a(\pi_{H_0}))^{-1}I_k$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b\}$ and $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$.

In this example,

$$\tau_\beta(\pi; \gamma_0, b) = -\frac{G_\beta(\pi; \gamma_0) - a(\pi)'Y_0^{-1}a(\pi_0)b}{a(\pi)'Y_0^{-1}a(\pi)}. \tag{13.20}$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $b \in R^k$, the asymptotic distribution of $A_n(\pi_{H_0})$ is

$$\begin{aligned} A(h, \pi_{H_0}) &= (a(\pi_{H_0})'Y_0^{-1}a(\pi_{H_0})\tau_\beta(\pi_{H_0}; \gamma_0, b)'\tau_\beta(\pi_{H_0}; \gamma_0, b)/k)^{1/2} \\ &= (-2\xi(\pi_{H_0}; \gamma_0, b)/k)^{1/2}. \end{aligned} \tag{13.21}$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $A_n(\pi_{H_0}) \sim (\chi^2_k/k)^{1/2}$ when $\pi_{H_0} = \pi_0$.

13.3. *Simplified Representation*

In this section, we simplify the expressions in (13.18) and (13.21) for the asymptotic distributions of $\mathrm{QLR}_n(\pi_{H_0})$ and $A_n(\pi_{H_0})$. We show that they correspond to the asymptotic distributions in Moreira (2003) and Andrews, Moreira, and Stock (2006) when $\Pi = R$. Above, we assume Π is compact because the general assumptions for nonlinear models used in AC1 rely on boundedness of the parameter space, as is common in the extremum estimator literature. In the linear model considered here that could be relaxed.

[83]By definition of $\widehat{\beta}_n(\pi)$, for the restriction $H_0 : \pi = \pi_{H_0}$, the restricted estimator $\widetilde{\beta}_n$ equals $\widehat{\beta}_n(\pi_{H_0})$. Also, for this restriction, some (lengthy) algebra shows that $\widetilde{\Sigma}_{\beta\beta,n}$ reduces to $\widetilde{J}^{-1}_{\beta\beta,n}\widetilde{V}_{\beta\beta,n}\widetilde{J}^{-1}_{\beta\beta,n}$, where $\widetilde{J}_{\beta\beta,n}$ and $\widetilde{V}_{\beta\beta,n}$ are the upper left $d_\beta \times d_\beta$ blocks of $\widetilde{J}_n$ and $\widetilde{V}_n$, respectively, and, in turn, $\widetilde{J}^{-1}_{\beta\beta,n}\widetilde{V}_{\beta\beta,n}\widetilde{J}^{-1}_{\beta\beta,n}$, reduces to the expression in (13.19) for $\widehat{\Sigma}_{\beta\beta,n}(\pi_{H_0})$.

Define two independent random variable S and T by

$$
\begin{aligned}
(13.22)\quad & S \sim N(c_\pi b, I_k) \quad \text{and} \quad T \sim N(d_\pi b, I_k), \quad \text{where} \\
& c_\pi = (\pi_0 - \pi_{H_0}) \cdot (a_\perp' Y_0 a_\perp)^{-1/2} \in R, \\
& d_\pi = a_0' Y_0^{-1} a \cdot (a' Y_0^{-1} a)^{-1/2} \in R, \\
& a_\perp = (1, -\pi_{H_0})', \quad a = (\pi_{H_0}, 1)', \quad \text{and} \quad a_0 = (\pi_0, 1)'.
\end{aligned}
$$

Now we show that under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $b \in R^k$, the distributions of $\mathrm{QLR}(h, \pi_{H_0})$ and $A(h, \pi_{H_0})$ in (13.18) and (13.21) satisfy

$$
\begin{aligned}
(13.23)\quad & \mathrm{QLR}(h, \pi_{H_0}) \sim \frac{1}{2}\big(Q_S - Q_T + \sqrt{(Q_S - Q_T)^2 + 4Q_{ST}^2}\big), \\
& A(h, \pi_{H_0}) \sim \sqrt{Q_T / k}, \quad \text{where} \\
& Q_S = S'S, \quad Q_T = T'T, \quad \text{and} \quad Q_{ST} = S'T.
\end{aligned}
$$

The result for $\mathrm{QLR}(h, \pi_{H_0})$ is analogous to the combination of (3.4) and Lemma 4 of Andrews, Moreira, and Stock (2006), but is obtained by a different route.

Define $a^*(\pi) = Y_0^{-1/2} a(\pi)$, where as above $a(\pi) = (\pi, 1)' \in R^2$, and $a_\perp^*(\pi) = Y_0^{1/2} a_\perp(\pi)$, where $a_\perp(\pi) = (1, -\pi)' \in R^2$. Then $G_\beta(\pi; \gamma_0) = \eta' a^*(\pi)$ and $a(\pi)' Y_0^{-1} a(\pi_0) b = b a^*(\pi_0)' a^*(\pi)$. The chi-square process $\xi(\pi; \gamma_0, b)$ can be written as

$$
\begin{aligned}
(13.24)\quad & \xi(\pi; \gamma_0, b) = -\frac{a^*(\pi)' M' M a^*(\pi)}{2 a^*(\pi)' a^*(\pi)}, \quad \text{where} \\
& M = \eta' - b a^*(\pi_0)' \in R^{k \times 2}, \\
& \mathrm{vec}(M) \sim N(-a^*(\pi_0) \otimes b, I_{2k}),
\end{aligned}
$$

and η is defined in (13.18). Define a 2×2 orthogonal matrix

$$
\begin{aligned}
(13.25)\quad & L = [L_1, L_2] \\
& = \left[\frac{-a_\perp^*(\pi_{H_0})}{\sqrt{a_\perp^*(\pi_{H_0})' a_\perp^*(\pi_{H_0})}}, \frac{-a^*(\pi_{H_0})}{\sqrt{a^*(\pi_{H_0})' a^*(\pi_{H_0})}} \right], \quad \text{which yields} \\
& ML = [ML_1, ML_2] = [\eta' L_1 + c_\pi b, \eta' L_2 + d_\pi b] \\
& \quad \sim [S, T],
\end{aligned}
$$

where the distribution holds because $\eta' L_1, \eta' L_2 \sim N(0, I_k)$, $\eta' L_1$ and $\eta' L_2$ are independent, and $a^*(\pi_0)' a_\perp^*(\pi_{H_0}) = \pi_0 - \pi_{H_0}$. Using the expressions above, we

obtain

$$\xi(\pi_{H_0};\gamma_0,b) = -\frac{1}{2}(ML_2)'(ML_2) \sim -\frac{1}{2}T'T = -\frac{1}{2}Q_T, \tag{13.26}$$

$$\begin{aligned}\inf_{\pi\in R}\xi(\pi;\gamma_0,b) &= -\frac{1}{2}\lambda_{\max}(M'M) = -\frac{1}{2}\lambda_{\max}((ML)'(ML))\\ &\sim -\frac{1}{2}\lambda_{\max}([S,T]'[S,T])\\ &= -\frac{1}{4}\bigl(Q_S+Q_T+\sqrt{(Q_S-Q_T)^2+4Q_{ST}^2}\bigr).\end{aligned}$$

This implies the desired results in (13.23) because $\mathrm{QLR}(h,\pi_{H_0}) = 2(\xi(\pi_{H_0};\gamma_0, b) - \inf_{\pi\in R}\xi(\pi;\gamma_0,b))$ and $A(\pi_{H_0};\gamma_0,b) = (-2\xi(\pi_{H_0};\gamma_0,b)/k)^{1/2}$.

13.4. *Unrestricted ICS Statistic*

Next we provide an unrestricted ICS statistic using a LS estimator of β and show that the asymptotic distribution of this statistic is a function of S and T. In the numerical study, we compare the powers of the type 2 robust QLR tests with null-imposed and unrestricted ICS statistics.

Let $\widehat{\beta}_n = (Z'Z)^{-1}Z'Ye_2$ be the LS estimator of β based on the second reduced-form equation, where $Z = (Z_1,\dots,Z_n)' \in R^{n\times k}$, $Y = (Y_1, Y_2) \in R^{n\times 2}$, $Y_j = (y_{j,1},\dots,y_{j,n})' \in R^n$ for $j = 1$ and 2, and $e_2 = (0,1)'$. The asymptotic variance of $n^{1/2}(\widehat{\beta}_n - \beta_n)$ is $e_2'Y_0e_2I_k$. The unrestricted ICS statistic is

$$A_n = \left(\widehat{Y}_{n,22}^{-1}n\widehat{\beta}_n'\left(n^{-1}\sum_{i=1}^n Z_iZ_i'\right)\widehat{\beta}_n/k\right)^{1/2}, \quad \text{where} \tag{13.27}$$

$$\widehat{Y}_{n,22} = n^{-1}\sum_{i=1}^n (y_{2,i} - Z_i'\widehat{\beta}_n)^2.$$

Now we show that under $\Gamma(\gamma_0,0,b)$ with $b\in R^k$,

$$A_n \to_d A^*(\pi_{H_0};\gamma_0,b) \sim ((\varphi_1S+\varphi_2T)'(\varphi_1S+\varphi_2T)/k)^{1/2}, \quad \text{where} \tag{13.28}$$

$$\varphi = (\varphi_1,\varphi_2)' = D^{-1}e_2 \in R^2,$$

$$D = \bigl[(e_2'Y_0e_2)^{1/2}a_\perp(a_\perp'Y_0a_\perp)^{-1/2}, (e_2'Y_0e_2)^{1/2}Y_0^{-1}a(a'Y_0^{-1}a)^{-1/2}\bigr],$$

where $a_\perp$ and a are defined in (13.22). Define

$$S_n = (Z'Z)^{-1/2}Z'Ya_\perp'\cdot(a_\perp'Y_0a_\perp)^{-1/2}, \tag{13.29}$$

$$T_n = (Z'Z)^{-1/2}Z'YY_0^{-1}a\cdot(a'Y_0^{-1}a)^{-1/2}.$$

Note that

$$\begin{aligned}(13.30)\quad \varphi_1 S_n + \varphi_2 T_n &= [S_n : T_n]\varphi = (e_2' Y_0 e_2)^{-1/2}(Z'Z)^{-1/2} Z'YD\varphi \\ &= (e_2' Y_0 e_2)^{-1/2}(Z'Z/n)^{1/2} n^{1/2}\widehat{\beta}_n \\ &= (e_2' Y_0 e_2)^{-1/2} n^{1/2}\widehat{\beta}_n + o_p(1).\end{aligned}$$

Hence,

$$(13.31)\quad A_n = \left((\varphi_1 S_n + \varphi_2 T_n)'(\varphi_1 S_n + \varphi_2 T_n)/k\right)^{1/2} + o_p(1)$$

by (13.27) and (13.30). This implies the desired result because under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $S_n \to_d S$ and $T_n \to_d T$ by arguments analogous to those used to establish Lemma 4 of Andrews, Moreira, and Stock (2006).

13.5. *Simulation Design*

The model considered is the same as that in the numerical section in Andrews, Moreira, and Stock (2006). The parameters that characterize the distributions of the tests are $\lambda = b'b$, the number of IV's k, the correlation between

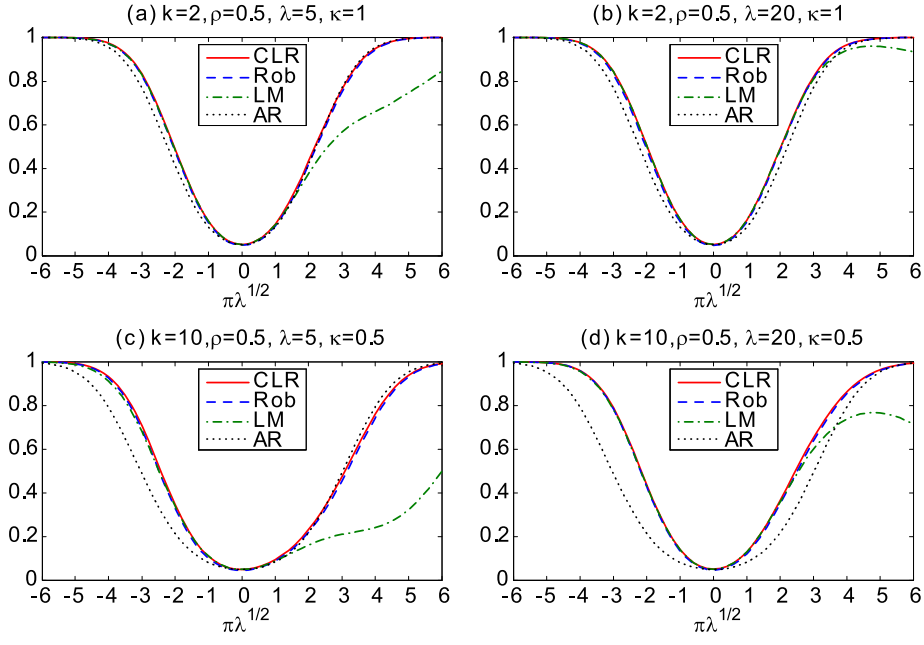

FIGURE S-26.—Power functions for the CLR, robust QLR, LM, and AR tests for the structural parameter π in the linear IV model, $k = 2, 10$, $\rho = 0.5$, $\lambda = 5, 20$. The ICS statistic for the robust QLR test is the null-imposed Wald statistic.

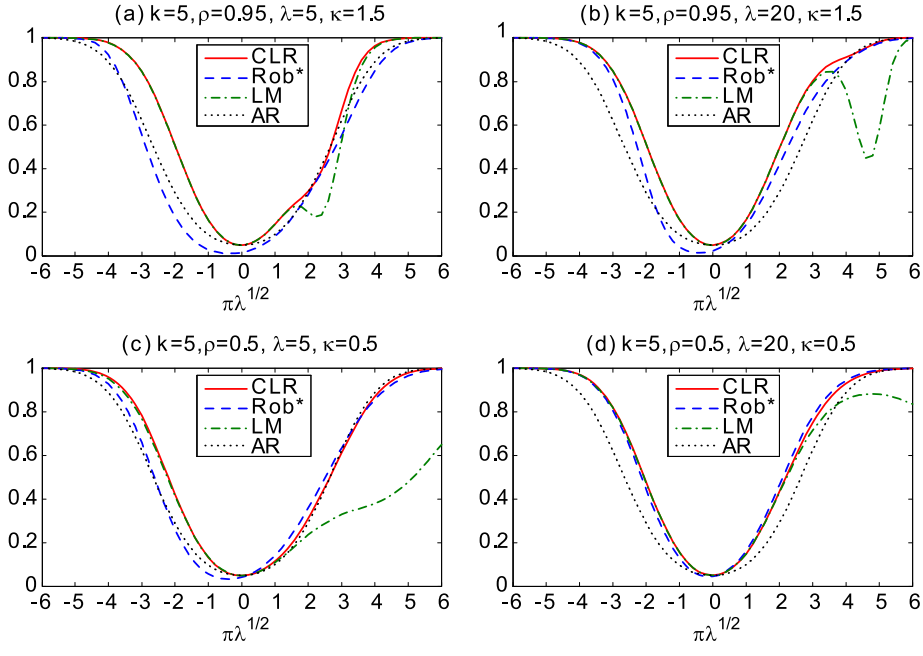

FIGURE S-27.—Power functions for the CLR, robust QLR, LM, and AR tests for the structural parameter π in the linear IV model, $k = 5$, $\rho = 0.95, 0.5$, $\lambda = 5, 20$. The ICS statistic for the robust QLR test is the unrestricted Wald statistic.

the reduced-form errors ρ, and $\pi_{H_0} - \pi_0$. The significance level of the tests is 5% and the parameter space for π is R. All results are based on 50,000 simulation repetitions.

We plot the power functions of the CLR, LM, and Anderson–Rubin (denoted AR in Figures S-26–S-28) tests together with the power function of the type 2 robust QLR test. For the robust test, we consider both the null-imposed ICS statistic $A_n(\pi_{H_0})$ and the unrestricted ICS statistic A_n.

For the type 2 robust test, the LF critical value is obtained over discrete values of λ from 0 to 40 with a grid of 1. The transition function $s(x)$ equals $\exp(-2x)$ and the constant D equals 0. The choices of $s(x)$ and D were determined via some experimentation to be good choices in terms of yielding null rejection probabilities that are relatively close to the nominal size 5% across different values of λ. Given $s(x)$ and D, the choice of κ was determined by maximizing average power against the alternatives plotted in the figures. The choice set of κ runs from 0 to 3 with a grid 0.5. A wide range of κ values yields similar average power.

The conditional critical values for the CLR test are based on tables in the Supplemental Appendix of Andrews, Moreira, and Stock (2006) and are computed with linear interpolation.

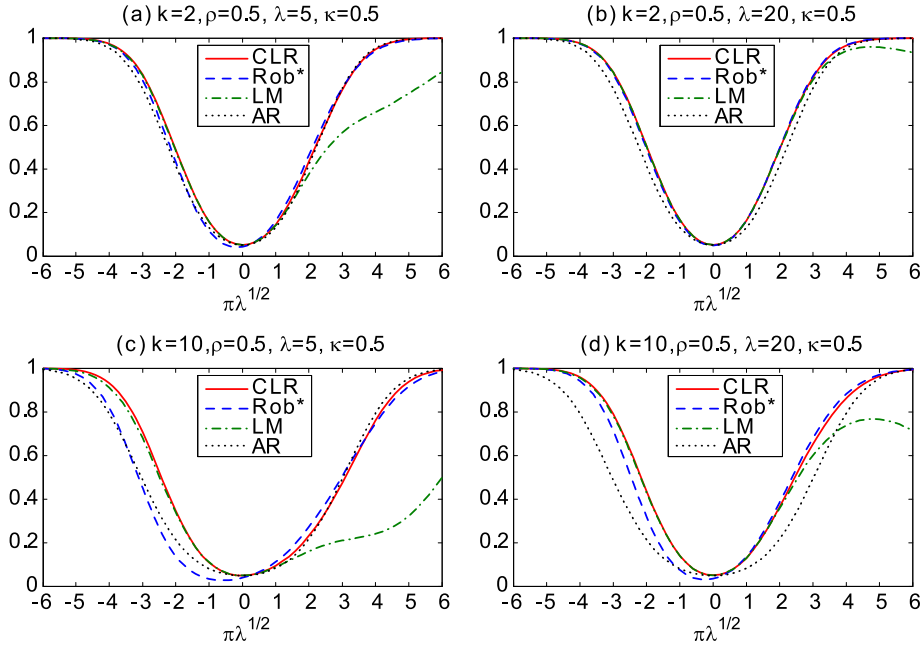

FIGURE S-28.—Power functions for the CLR, robust QLR, LM, and AR tests for the structural parameter π in the linear IV model, $k = 2, 10$, $\rho = 0.5$, $\lambda = 5, 20$. The ICS statistic for the robust QLR test is the unrestricted Wald statistic.

13.6. *Results*

The results are given in Figure 8 of AC1, as well as Figures S-26–S-32. Figure S-26 shows that the robust QLR test based on the NI-ICS statistic has

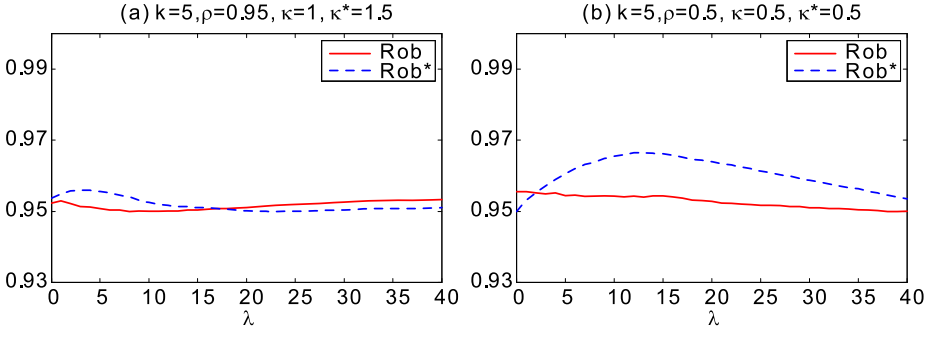

FIGURE S-29.—Coverage probabilities of robust QLR CI's for the structural parameter π in the linear IV model, $k = 5$, $\rho = 0.95, 0.5$. The ICS statistics for Rob and Rob* are the null-imposed and unrestricted Wald statistics.

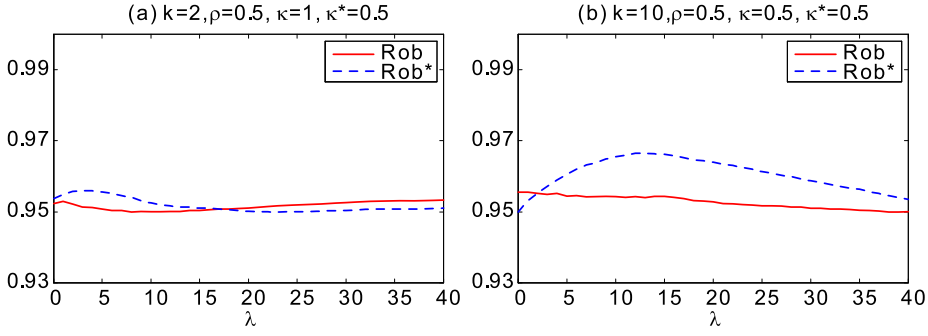

FIGURE S-30.—Coverage probabilities of robust QLR CI's for the structural parameter π in the linear IV model, $k = 2, 10$, $\rho = 0.5$. The ICS statistics for Rob and Rob* are the null-imposed and unrestricted Wald statistics.

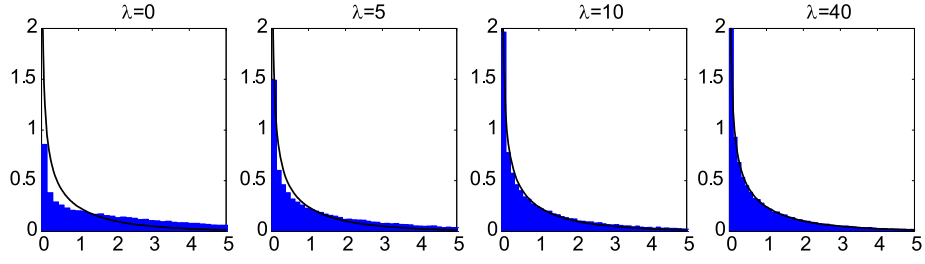

FIGURE S-31.—Asymptotic densities of the QLR statistic for the structural parameter π in the linear IV model when $k = 5$, $\rho = 0.5$ and the χ^2_1 density (black line).

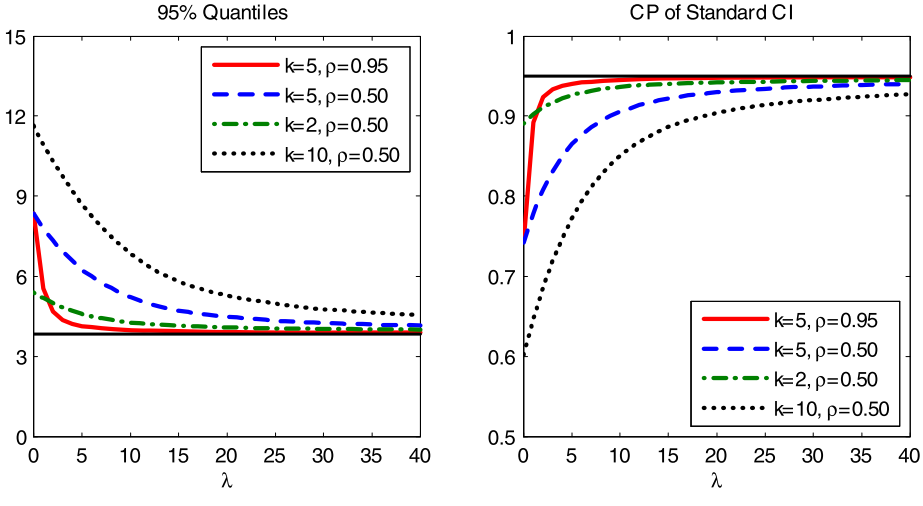

FIGURE S-32.—Asymptotic 95% quantiles of the QLR statistic and asymptotic coverage probabilities of standard CI's concerning the structural parameter π in the linear IV model.

power that is essentially equal to that of the CLR test. Figures S-27 and S-28 show that the type 2 robust test based on the unrestricted ICS statistic has lower power than the CLR test.

Figures S-29 and S-30 show the coverage probabilities of the two robust QLR tests as a function of λ, which measures the strength of the IV's. The robust test based on the NI-ICS statistic is close to being asymptotically similar. The robust test based on the unrestricted ICS statistic overcovers in some scenarios.

Figure S-31 graphs the density of the QLR statistic under the null hypothesis and compares it to a chi-square distribution with 1 degree of freedom, χ_1^2 (which is its distribution under strong identification). It is clear that for weak IV's (i.e., small λ), the χ_1^2 distribution does not provide a good approximation in the upper tail to the actual asymptotic distribution.

The first set of graphs in Figure S-32 shows that the 95% quantiles of the asymptotic distribution of the QLR statistic increase noticeably as λ decreases to 0. The second set of graphs in Figure S-32 show that the standard QLR test, which uses the 95% quantile from the χ_1^2 distribution, undercovers noticeably with weak IV's. The asymptotic size of the standard QLR test varies from 60% to 90%, depending on the parameter configuration.

REFERENCES

ANDREWS, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257. [9,13,74]

——— (1993): "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856. [3,20]

——— (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. McFadden. Amsterdam: North-Holland. [16,123]

——— (1999): "Estimation When a Parameter Is on a Boundary," *Econometrica*, 67, 1341–1383. [4,7,23,39,40,44,46,52]

——— (2001): "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734. [4,7,14,18]

——— (2002): "Generalized Method of Moments Estimation When a Parameter Is on a Boundary," *Journal of Business & Economic Statistics*, 20, 530–544. [44]

ANDREWS, D. W. K., AND X. CHENG (2011a): "Maximum Likelihood Estimation and Inference With Weak, Semi-Strong, and Strong Identification," Cowles Foundation Discussion Paper No. 1024, Yale University. [2,118,120,124]

——— (2011b): "GMM Estimation and Inference With Weak, Semi-Strong, and Strong Identification," Cowles Foundation Discussion Paper No. 1028, Yale University. [2]

ANDREWS, D. W. K., AND P. GUGGENBERGER (2010): "Asymptotic Size and a Problem With Subsampling and With the *m* out of *n* Bootstrap," *Econometric Theory*, 26, 426–468. [19]

ANDREWS, D. W. K., AND W. PLOBERGER (1996): "Testing for Serial Correlation Against an ARMA(1, 1) Process," *Journal of the American Statistical Association*, 91, 1331–1342. [80,88]

ANDREWS, D. W. K., M. J. MOREIRA, AND J. H. STOCK (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752. [130,131,136,137,139,140]

——— (2008): "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression With Weak Instruments," *Journal of Econometrics*, 146, 241–254. [130]

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. [16]

CHERNOFF, H. (1954): "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 54, 573–578. [4,7,23,39]

CHERNOZHUKOV, V., C. HANSEN, AND M. JANSSON (2009): "Admissible Invariant Similar Tests for Instrumental Variables Regression," *Econometric Theory*, 25, 808–818. [130]

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. [16]

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Applications*. New York: Academic Press. [87]

HAN, S. (2009): "Identification and Inference in a Bivariate Probit Model With Weak Instruments," Unpublished Manuscript, Department of Economics, Yale University. [2,18]

HAN, S., AND E. VYTLACIL (2009): "Identification in a Bivariate Probit Model With Endogenous Regressors," Unpublished Manuscript, Department of Economics, Yale University. [18]

HANSEN, B. E. (1996): "Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays," *Econometric Theory*, 12, 347–359. [16]

KIM, J., AND D. POLLARD (1990): "Cube Root Asymptotics," *The Annals of Statistics*, 17, 191–219. [12,61]

KLEIBERGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803. [130]

MAGNUS, J. R., AND H. NEUDECKER (1979): "The Communication Matrix: Some Properties and Applications," *The Annals of Statistics*, 7, 381–394. [133]

MOREIRA, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048. [2,130,136]

——— (2009): "Tests With Correct Size When Instruments Can Be Arbitrarily Weak," *Journal of Econometrics*, 152, 131–140. [130]

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057. [4]

POLLARD, D. (1984): *Convergence of Stochastic Processes*. New York: Springer-Verlag. [2]

——— (1985): "New Ways to Prove Central Limit Theorems," *Econometric Theory*, 1, 295–313. [14,16]

——— (1990): *Empirical Processes: Theory and Applications*. CBMS Conference Series in Statistics and Probability, Vol. 2. Hayward, CA: Institute of Mathematical Statistics. [2]

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. [130]

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM With Weak Instruments," *Econometrica*, 68, 1055–1096. [118,128]

TRIPATHI, G. T. (1999): "A Matrix Extension of the Cauchy–Schwarz Inequality," *Economics Letters*, 63, 1–3. [17,125]

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press. [124]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer. [2,6,14,25,42,47-49]

Cowles Foundation, Yale University, 30 Hillhouse Ave., New Haven, CT 06520-8281, U.S.A.; donald.andrews@yale.edu

and

Dept. of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, U.S.A.; xu.cheng@sas.upenn.edu.

Manuscript received July, 2010; final revision received January, 2012.