3.2. *Concentrated Estimator and Probability Limit Results*

Define the concentrated extremum estimator $\widehat{\psi}_n(\pi)$ $(\in \Psi(\pi))$ of ψ for given $\pi \in \Pi$ by

$$Q_n(\widehat{\psi}_n(\pi), \pi) = \inf_{\psi \in \Psi(\pi)} Q_n(\psi, \pi) + o(n^{-1}). \tag{3.2}$$

Let $Q_n^c(\pi)$ denote the concentrated sample criterion function $Q_n(\widehat{\psi}_n(\pi), \pi)$. Define an extremum estimator $\widehat{\pi}_n$ $(\in \Pi)$ by

$$Q_n^c(\widehat{\pi}_n) = \inf_{\pi \in \Pi} Q_n^c(\pi) + o(n^{-1}). \tag{3.3}$$

We assume that the extremum estimator $\widehat{\theta}_n$ in (2.1) can be written as $\widehat{\theta}_n = (\widehat{\psi}_n(\widehat{\pi}_n), \widehat{\pi}_n)$.[29]

Next, we specify the limit of the sample criterion function $Q_n(\theta)$ along drifting sequences of true parameters $\{\gamma_n\} \in \Gamma(\gamma_0)$ whose limit is $\gamma_0 \in \Gamma$ and determine the probability limit of $\widehat{\theta}_n$.

ASSUMPTION B3: (i) *For some nonstochastic real-valued function* $Q(\theta; \gamma_0)$ *on* $\Theta \times \Gamma$, $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta; \gamma_0)| \to_p 0$ *under* $\{\gamma_n\} \in \Gamma(\gamma_0)$ $\forall \gamma_0 \in \Gamma$.

(ii) *When* $\beta_0 = 0$, *for every neighborhood* Ψ_0 $(\subset R^{d_\psi})$ *of* $\psi_0 = (\beta_0, \zeta_0)$, $\inf_{\pi \in \Pi}(\inf_{\psi \in \Psi(\pi)/\Psi_0} Q(\psi, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)) > 0$ $\forall \gamma_0 = (\psi_0, \pi_0, \phi_0) \in \Gamma$.

(iii) *When* $\beta_0 \neq 0$, *for every neighborhood* Θ_0 $(\subset \Theta)$ *of* $\theta_0 = (\beta_0, \zeta_0, \pi_0)$, $\inf_{\theta \in \Theta/\Theta_0} Q(\theta; \gamma_0) - Q(\theta_0; \gamma_0) > 0$ $\forall \gamma_0 = (\theta_0, \phi_0) \in \Gamma$.

Assumption B3(i) defines the (asymptotic) population criterion function $Q(\theta; \gamma_0)$. Assumption B3(ii) provides a condition for the identification of β and ζ despite the nonidentification of π when $\beta_0 = 0$. Uniformity over Π is required due to the nonidentification of π. A condition of this type also is used in Andrews (1993) for the uniform consistency of a family of estimators. Assumption B3(iii) is a standard identification condition for θ when $\beta_0 \neq 0$. A condition of this sort is verified for various extremum estimators in Newey and McFadden (1994).

A set of primitive sufficient conditions for Assumption B3(ii) and (iii) is given in Assumption B3* in Supplemental Appendix A.

LEMMA 3.1: *Suppose Assumptions* A *and* B3 *hold*. *Under* $\{\gamma_n\} \in \Gamma(\gamma_0)$, *where* $\gamma_0 = (\beta_0, \zeta_0, \pi_0, \phi_0)$, *there are two alternatives*:

(a) *When* $\beta_0 = 0$, $\sup_{\pi \in \Pi} \|\widehat{\psi}_n(\pi) - \psi_n\| \to_p 0$ *and* $\widehat{\psi}_n - \psi_n \to_p 0$.

(b) *When* $\beta_0 \neq 0$, $\widehat{\theta}_n - \theta_n \to_p 0$.

COMMENT: When $\beta_0 = 0$, the asymptotic behavior of $\widehat{\pi}_n$ is determined below.

[29]If (3.2) and (3.3) hold and $\widehat{\theta}_n = (\widehat{\psi}_n(\widehat{\pi}_n), \widehat{\pi}_n)$, then (2.1) automatically holds.

3.3. *Close to* $\beta = 0$ *Assumptions and Estimation Results*

The following Assumptions C1–C8 are used to determine the asymptotic distributions of estimators and test statistics under sequences of true parameters $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$ and to establish the consistency of $\widehat{\pi}_n$ under sequences $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| = \infty$. The "C" denotes that the sequences of parameters $\{\gamma_n\}$ considered are *close* to the point of nonidentification.

The first assumption, Assumption C1, requires that the criterion function $Q_n(\theta)$ has a stochastic quadratic expansion in ψ around the nonidentification point $\psi_{0,n} = (0, \zeta_n)$ uniformly in $\pi \in \Pi$. Assumptions C2 and C3 concern the behavior of the (generalized) first derivative in the expansion. Assumption C4 concerns the behavior of the (generalized) second derivative. Assumptions C5 and C7 arise because the quadratic expansion is about the nonidentification point $\psi_{0,n}$, rather than the true value ψ_n. Assumptions C6–C8 are used when determining the asymptotic behavior of $\widehat{\pi}_n$.

We now define a sequence of scalar constants $\{a_n(\gamma_n) : n \geq 1\}$ that provides the normalization required so that the (generalized) first derivative in the quadratic expansion in Assumption C1 is nondegenerate asymptotically.[30] These constants appear in the conditions on the remainder term of the approximation in Assumption C1. Define

$$(3.4) \qquad a_n(\gamma_n) = \begin{cases} n^{1/2}, & \text{if } \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \text{ and } \|b\| < \infty, \\ \|\beta_n\|^{-1}, & \text{if } \{\gamma_n\} \in \Gamma(\gamma_0, 0, b) \text{ and } \|b\| = \infty. \end{cases}$$

Note that $\|\beta_n\|^{-1} < n^{1/2}$ for n large when $\|b\| = \infty$, because $n^{1/2}\|\beta_n\| \to \infty$.[31] Hence, $a_n(\gamma_n) \leq n^{1/2}$ for n large.

ASSUMPTION C1: *Under* $\{\gamma_n = (\beta_n, \zeta_n, \pi_n, \phi_n)\} \in \Gamma(\gamma_0, 0, b)$, *for some* $\delta > 0$, $\forall \theta = (\psi, \pi) \in \Theta_\delta = \{\theta \in \Theta : \|\beta\| < \delta\}$, *the following statements hold*:

(i) *The sample criterion function* $Q_n(\psi, \pi)$ *has a quadratic expansion in* ψ *around* $\psi_{0,n} = (0, \zeta_n)$ *for given* π,

$$Q_n(\psi, \pi) = Q_n(\psi_{0,n}, \pi) + D_\psi Q_n(\psi_{0,n}, \pi)'(\psi - \psi_{0,n}) + \frac{1}{2}(\psi - \psi_{0,n})' D_{\psi\psi} Q_n(\psi_{0,n}, \pi)(\psi - \psi_{0,n}) + R_n(\psi, \pi),$$

where $D_\psi Q_n(\psi_{0,n}, \pi) \in R^{d_\psi}$ *is a stochastic generalized first partial-derivative vector, and* $D_{\psi\psi} Q_n(\psi_{0,n}, \pi) \in R^{d_\psi \times d_\psi}$ *is a generalized second partial-derivative matrix that is symmetric and may be stochastic or nonstochastic*.

(ii) *The remainder*, $R_n(\psi, \pi)$, *satisfies*

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

[30]See Lemma 9.1 in Supplemental Appendix B.

[31]The quantity $a_n(\gamma_n)$ actually depends on the entire sequence $\{\gamma_n\}$ because b depends on $\{\gamma_n\}$.

for all constants $\delta_n \to 0$,

(iii) $D_\zeta Q_n(\theta)$ *and* $D_{\zeta\zeta} Q_n(\theta)$ *do not depend on* π *when* $\beta = 0$, *where* $\theta = (\beta, \zeta, \pi) \in \Theta$, $D_\zeta Q_n(\theta)$ *denotes the last* d_ζ *elements of* $D_\psi Q_n(\theta)$, *and* $D_{\zeta\zeta} Q_n(\theta)$ *is the lower* $d_\zeta \times d_\zeta$ *block of* $D_{\psi\psi} Q_n(\theta)$.

Because the expansion in Assumption C1 is about the point of lack of identification $\psi_{0,n}$, rather than the true value ψ_n, the leading term $Q_n(\psi_{0,n}, \pi)$ does not depend on π by Assumption A. This is key. It implies that $\widehat{\theta}_n = (\widehat{\psi}_n, \widehat{\pi}_n)$ not only minimizes $Q_n(\psi, \pi)$, but also $Q_n(\psi, \pi) - Q_n(\psi_{0,n}, \pi)$. The latter has the quadratic expansion in Assumption C1 with linear and quadratic terms whose asymptotic properties one can determine using Assumptions C2–C5 below.

Sufficient conditions for Assumption C1 when $Q_n(\theta)$ is a sample average that is smooth in θ are given in Lemma 8.6 in Supplemental Appendix A. In this case, $D_\psi Q_n(\theta)$ and $D_{\psi\psi} Q_n(\theta)$ are the pointwise partial and second partial derivatives of $Q_n(\theta)$. For the nonsmooth sample average case, sufficient conditions are given in Lemma 8.7 in Supplemental Appendix A. In this case, $D_\psi Q_n(\theta)$ is a "stochastic derivative" of $Q_n(\theta)$, which typically equals the pointwise derivative for points where the latter exists, and $D_{\psi\psi} Q_n(\theta)$ is the (nonstochastic) second partial derivative of the expected value of $Q_n(\theta)$. This case covers quantile estimators and ML and LS estimators in continuous, but not smooth, threshold autoregressive models, as in Chan and Tsay (1998). Sufficient conditions for Assumption C1 when $Q_n(\theta)$ is a GMM or minimum distance (MD) criterion function, smooth or nonsmooth in θ, are given in AC3.

If $D_\psi Q_n(\theta)$ and $D_{\psi\psi} Q_n(\theta)$ are the pointwise partial and second partial derivatives of $Q_n(\theta)$, then Assumption C1(iii) is implied by Assumption A. Otherwise, in the presence of Assumption A, Assumption C1(iii) is not restrictive.

Note that Assumption C1 is compatible with semiparametric estimators.

The (generalized) first derivative of $Q_n(\theta)$ w.r.t. ψ is assumed to satisfy the following assumption.

ASSUMPTION C2: (i) $D_\psi Q_n(\theta)$ *takes the form* $D_\psi Q_n(\theta) = n^{-1} \sum_{i=1}^n m(W_i, \theta)$ *for some function* $m(W_i, \theta) \in R^{d_\psi}$ $\forall \theta \in \Theta_\delta$, *for any true parameter* $\gamma^* \in \Gamma$.

(ii) $E_{\gamma^*} m(W_i, \psi^*, \pi) = 0$ $\forall \pi \in \Pi$, $\forall i \geq 1$ *when the true parameter is* γ^* $\forall \gamma^* = (\psi^*, \pi^*, \phi^*) \in \Gamma$ *with* $\beta^* = 0$.[32]

Define an empirical process $\{G_n(\pi) : \pi \in \Pi\}$ by

$$G_n(\pi) = n^{-1/2} \sum_{i=1}^n \big(m(W_i, \psi_{0,n}, \pi) - E_{\gamma_n} m(W_i, \psi_{0,n}, \pi) \big). \tag{3.5}$$

[32]In some time series examples, $D_\psi Q_n(\theta)$ is of the form $n^{-1} \sum_{i=1}^n m_i(\theta)$, where $m_i(\theta)$ depends on $\{W_j : \forall 1 \leq j \leq i\}$. Assumption C2 can be relaxed to cover such cases without any changes to the results of the paper. In such cases, Assumption C3 below still can hold provided $\{m_i(\theta) : i \leq n\}$ satisfies a suitable "asymptotic weak dependence" condition, such as near-epoch dependence.

The recentered and rescaled (generalized) first derivative of $Q_n(\theta)$ w.r.t. ψ is assumed to satisfy an empirical process central limit theorem (CLT):

ASSUMPTION C3: *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $G_n(\cdot) \Rightarrow G(\cdot; \gamma_0)$, *where* $G(\cdot; \gamma_0)$ *is a mean zero Gaussian process indexed by* $\pi \in \Pi$ *with bounded continuous sample paths and some covariance kernel* $\Omega(\pi_1, \pi_2; \gamma_0)$ *for* $\pi_1, \pi_2 \in \Pi$.

Numerous empirical process results in the literature can be used to verify this assumption, including results in Pollard (1984, 1990), Andrews (1994), and van der Vaart and Wellner (1996).

The (generalized) second derivative of $Q_n(\theta)$ w.r.t. ψ is assumed to satisfy the following assumption.

ASSUMPTION C4: (i) *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $\sup_{\pi \in \Pi} \|D_{\psi\psi} Q_n(\psi_{0,n}, \pi) - H(\pi; \gamma_0)\| \to_p 0$ *for some nonstochastic symmetric* $d_\psi \times d_\psi$ *matrix-valued function* $H(\pi; \gamma_0)$ *on* $\Pi \times \Gamma$ *that is continuous on* Π $\forall \gamma_0 \in \Gamma$.

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

Define the $d_\psi \times d_\beta$ matrix of partial derivatives of the average population moment function w.r.t. the true β value, β^*, to be

$$K_n(\theta; \gamma^*) = n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \beta^{*\prime}} E_{\gamma^*} m(W_i, \theta). \tag{3.6}$$

The domain of the function $K_n(\theta; \gamma^*)$ is $\Theta_\delta \times \Gamma_0$, where $\Theta_\delta = \{\theta \in \Theta : \|\beta\| < \delta\}$, $\Gamma_0 = \{\gamma_a = (a\beta, \zeta, \pi, \phi) \in \Gamma : \gamma = (\beta, \zeta, \pi, \phi) \in \Gamma \text{ with } \|\beta\| < \delta \text{ and } a \in [0, 1]\}$, and $\delta > 0$ is as in Assumption B2(ii). The set Γ_0 is not empty by Assumptions B2(ii) and (iii).

ASSUMPTION C5: (i) $K_n(\theta; \gamma^*)$ *exists* $\forall (\theta, \gamma^*) \in \Theta_\delta \times \Gamma_0$, $\forall n \geq 1$.

(ii) *For some nonstochastic* $d_\psi \times d_\beta$ *matrix-valued function* $K(\psi_0, \pi; \gamma_0)$, $K_n(\overline{\psi}_n, \pi; \widetilde{\gamma}_n) \to K(\psi_0, \pi; \gamma_0)$ *uniformly over* $\pi \in \Pi$ *for all nonstochastic sequences* $\{\overline{\psi}_n\}$ *and* $\{\widetilde{\gamma}_n\}$ *such that* $\widetilde{\gamma}_n \in \Gamma$, $\widetilde{\gamma}_n \to \gamma_0 = (0, \zeta_0, \pi_0, \phi_0)$ *for some* $\gamma_0 \in \Gamma$, $(\overline{\psi}_n, \pi) \in \Theta$, *and* $\overline{\psi}_n \to \psi_0 = (0, \zeta_0)$.

(iii) $K(\psi_0, \pi; \gamma_0)$ *is continuous on* Π $\forall \gamma_0 \in \Gamma$ *with* $\beta_0 = 0$.

Assumption C5 is not restrictive. A set of primitive sufficient conditions for Assumption C5 is given in Supplemental Appendix A.

For simplicity, $K(\psi_0, \pi; \gamma_0)$ is abbreviated as $K(\pi; \gamma_0)$. Note that $(\overline{\psi}_n, \widetilde{\gamma}_n)$ in Assumption C5(ii) is in $\Theta_\delta \times \Gamma_0$ for n large.

Due to the expansion about $\psi_{0,n}$, rather than about the true value ψ_n, in Assumption C1, a bias is introduced in the first derivative $D_\psi Q_n(\psi_{0,n}, \pi)$: its mean is not zero. In consequence, its behavior differs between category I and

II sequences. With category I sequences, it converges (after suitable normalization) to the sum of the stochastic term $G(\pi)$ and the nonstochastic term $K(\pi;\gamma_0)b$ due to the bias, 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 $K(\pi;\gamma_0)\omega_0$, which is nonstochastic.

Specifically, Assumptions C2, C3, and C5 are used to show the key result

$$
\begin{aligned}
(3.7)\qquad & a_n(\gamma_n)D_\psi Q_n(\psi_{0,n},\pi) \\
&= \left[G_n(\pi) + (K_n(\psi_{0,n},\pi;\gamma_n) + o(1))n^{1/2}\beta_n\right]n^{-1/2}a_n(\gamma_n) \\
&\Rightarrow \begin{cases} G(\pi;\gamma_0) + K(\pi;\gamma_0)b, \\ \quad \text{if } n^{1/2}\beta_n \to b \in R^{d_\beta}, \\ K(\pi;\gamma_0)\omega_0, \\ \quad \text{if } \|n^{1/2}\beta_n\| \to \infty \text{ and } \beta_n/\|\beta_n\| \to \omega_0, \end{cases}
\end{aligned}
$$

where the convergence holds under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$.[33]

Next, we introduce the limits of the concentrated criterion function $Q_n^c(\pi) = Q_n(\widehat{\psi}_n(\pi),\pi)$ after suitable normalization. Define a "weighted noncentral chi-square" process $\{\xi(\pi;\gamma_0,b):\pi\in\Pi\}$ and a nonstochastic function $\{\eta(\pi;\gamma_0,\omega_0):\pi\in\Pi\}$ by

$$
\begin{aligned}
(3.8)\qquad \xi(\pi;\gamma_0,b) &= -\frac{1}{2}(G(\pi;\gamma_0)+K(\pi;\gamma_0)b)'H^{-1}(\pi;\gamma_0) \\
&\quad \times (G(\pi;\gamma_0)+K(\pi;\gamma_0)b), \\
\eta(\pi;\gamma_0,\omega_0) &= -\frac{1}{2}\omega_0'K(\pi;\gamma_0)'H^{-1}(\pi;\gamma_0)K(\pi;\gamma_0)\omega_0.
\end{aligned}
$$

Under Assumptions C3, C4, and C5(iii), $\{\xi(\pi;\gamma_0,b):\pi\in\Pi\}$ has bounded continuous sample paths almost surely (a.s.).

Let $Q_{0,n} = Q_n(\psi_{0,n},\pi)$, where $\psi_{0,n} = (0,\zeta_n)$ as in Assumption C1. Note that $Q_{0,n}$ does not depend on π by Assumption A.

LEMMA 3.2: *Suppose Assumptions A, B1–B3, and C1–C5 hold. Under $\{\gamma_n\}\in\Gamma(\gamma_0,0,b)$, there are two alternatives*:

(a) *When $\|b\|<\infty$, $n(Q_n^c(\cdot)-Q_{0,n}) \Rightarrow \xi(\cdot;\gamma_0,b)$.*

(b) *When $\|b\|=\infty$ and $\beta_n/\|\beta_n\|\to\omega_0$ for some $\omega_0\in R^{d_\beta}$ with $\|\omega_0\|=1$, $\|\beta_n\|^{-2}(Q_n^c(\pi)-Q_{0,n}) \to_p \eta(\pi;\gamma_0,\omega_0)$ uniformly over $\pi\in\Pi$.*

To obtain the asymptotic distribution of $\widehat{\pi}_n$ when $\beta_n = O(n^{-1/2})$ via the continuous mapping theorem, we use the following assumption.

[33]See Lemma 9.1 in Supplemental Appendix B.

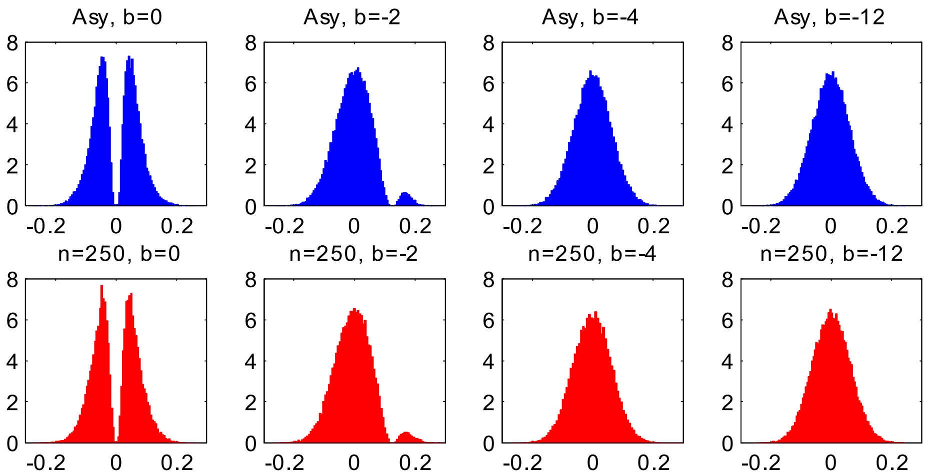

FIGURE 2.—Asymptotic and finite-sample ($n = 250$) densities of the estimator of β (centered at the true value) in the ARMA(1, 1) model when $\pi_0 = 0.4$.

ASSUMPTION C7: *The nonstochastic function* $\eta(\pi; \gamma_0, \omega_0)$ *is uniquely minimized over* $\pi \in \Pi$ *at* π_0 $\forall \gamma_0 \in \Gamma$ *with* $\beta_0 = 0$.

In Assumption C7, the minimizing value π_0 is nonrandom. In some examples, such as the ARMA(1, 1) example, Assumption C7 can be verified directly. In other examples, Assumption C7 can be verified using the Cauchy–Schwarz inequality or a matrix version of it (see Tripathi (1999)) when $K(\pi; \gamma_0)$ and $H(\pi; \gamma_0)$ take proper forms. For example, see the verification of Assumption C7 for the nonlinear regression example in Supplemental Appendix E and the verification of Assumption C7 for GMM estimators in AC3.

Lemma 9.3 in Supplemental Appendix B shows that when $\pi = \pi_0$, $K(\pi; \gamma_0) = -H(\pi; \gamma_0)S_\beta'$, where $S_\beta = [I_{d_\beta} : 0] \in R^{d_\beta \times d_\psi}$, whereas this relationship does not hold for $\pi \neq \pi_0$ in general.

LEMMA 3.3: *Suppose Assumptions* A, B1–B3, C1–C5, *and* C7 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, (a) $\widehat{\pi}_n - \pi_n \to_p 0$ *and* (b) $\widehat{\psi}_n - \psi_n \to_p 0$.

The following assumption is used when obtaining a key rate of convergence result for $\widehat{\psi}_n$ for sequences $\{\gamma_n\}$ for which $\beta_n \to 0$ and $n^{1/2}\|\beta_n\| \to \infty$.

ASSUMPTION C8: *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$, $\frac{\partial}{\partial \psi'} E_{\gamma_n} D_\psi Q_n(\psi, \pi_n)|_{\psi = \psi_n} \to H(\pi_0; \gamma_0)$.

By Assumption C4(i), $H(\pi; \gamma_0)$ is the probability limit of $D_{\psi\psi} Q_n(\psi_{0,n}, \pi_n)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$. When $Q_n(\theta)$ is a twice differentiable sample average,

For $v \in r(\Theta)$, we define a restricted estimator $\widetilde{\theta}_n(v)$ of θ subject to the restriction that $r(\theta) = v$. By definition,

$$\begin{aligned}(4.8)\qquad &\widetilde{\theta}_n(v) \in \Theta, \quad r(\widetilde{\theta}_n(v)) = v, \quad \text{and} \\ &Q_n(\widetilde{\theta}_n(v)) = \inf_{\theta \in \Theta : r(\theta) = v} Q_n(\theta) + o(n^{-1}).\end{aligned}$$

For testing $H_0 : r(\theta) = v$, the QLR test statistic is

$$(4.9)\qquad \mathrm{QLR}_n(v) = 2n\big(Q_n(\widetilde{\theta}_n(v)) - Q_n(\widehat{\theta}_n)\big)/\widehat{s}_n,$$

where $\widehat{s}_n$ is a real-valued scaling factor that is employed in some cases to yield a QLR statistic that has an asymptotic $\chi^2_{d_r}$ null distribution under strong identification; see Assumptions RQ2 and RQ3 below.

4.5. *QLR Assumptions*

If $r(\theta)$ includes restrictions on π (i.e., $d_{r_2} > 0$), then not all values $\pi \in \Pi$ are consistent with the restriction $r_2(\pi) = v_2$. For $v_2 \in r_2(\Theta)$, the set of π values that are consistent with $r_2(\pi) = v_2$ is denoted by

$$(4.10)\qquad \Pi_r(v_2) = \{\pi \in \Pi : r_2(\pi) = v_2 \text{ for some } \theta = (\psi, \pi) \in \Theta\}.$$

If $d_{r_2} = 0$, then by definition $\Pi_r(v_2) = \Pi\ \forall v_2 \in r_2(\Theta)$.

We assume that $r(\theta)$ satisfies the following assumption.

ASSUMPTION RQ1: (i) $r(\theta)$ *is continuously differentiable on* Θ.
(ii) $r_\theta(\theta)$ *is full row rank* $d_r\ \forall \theta \in \Theta$.
(iii) $r(\theta)$ *satisfies* (4.7).
(iv) $d_H(\Pi_r(v_2), \Pi_r(v_{0,2})) \to 0$ *as* $v_2 \to v_{0,2}\ \forall v_{0,2} \in r_2(\Theta^*)$.
(v) $Q(\psi, \pi; \gamma_0)$ *is continuous in* ψ *at* ψ_0 *uniformly over* $\pi \in \Pi$ (*i.e.*, $\sup_{\pi \in \Pi} |Q(\psi, \pi; \gamma_0) - Q(\psi_0, \pi; \gamma_0)| \to 0$ *as* $\psi \to \psi_0$) $\forall \gamma_0 \in \Gamma$ *with* $\beta_0 = 0$.
(vi) $Q(\theta; \gamma_0)$ *is continuous in* θ *at* $\theta_0\ \forall \gamma_0 \in \Gamma$ *with* $\beta_0 \neq 0$.

In Assumption RQ1(iv), d_H denotes the Hausdorff distance. Assumption RQ1(i) and (ii) are standard. Assumption RQ1(iv) is easy to verify in most cases. Assumption RQ1(v) and (vi) are not restrictive.

Even under strong identification, it is known that the QLR statistic has an asymptotic $\chi^2_{d_r}$ null distribution only under additional assumptions to those used for Wald and Lagrange multiplier (LM) statistics.[42] The following assumptions correspond to these additional conditions.

[42]The reason is that the weight matrices of the Wald and LM statistics can be designed specifically to achieve an asymptotic $\chi^2_{d_r}$ null distribution, whereas with the QLR statistic, no weight matrix appears and at most one has a real-valued scaling factor $\widehat{s}_n$ with which to make adjustments.

ASSUMPTION RQ2: *Either* (i) $V(\gamma_0) = s(\gamma_0)J(\gamma_0)$ *for some nonrandom scalar constant* $s(\gamma_0)$ $\forall \gamma_0 \in \Gamma$, *or* (ii) $V(\gamma_0)$ *and* $J(\gamma_0)$ *are block diagonal* (*possibly after reordering their rows and columns*), *the restrictions* $r(\theta)$ *only involve parameters that correspond to one block of* $V(\gamma_0)$ *and* $J(\gamma_0)$—*call them* $V_{11}(\gamma_0)$ *and* $J_{11}(\gamma_0)$—*and for this block*, $V_{11}(\gamma_0) = s(\gamma_0)J_{11}(\gamma_0)$ *for some nonrandom scalar constant* $s(\gamma_0)$ $\forall \gamma_0 \in \Gamma$.

ASSUMPTION RQ3: *The scalar statistic* $\widehat{s}_n$ *satisfies* $\widehat{s}_n \to_p s(\gamma_0)$ *under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ *and under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$.

For example, Assumptions RQ2(i) and RQ3 hold with $s(\gamma_0) = \widehat{s}_n = 1$ for a correctly specified log-likelihood criterion function, a GMM criterion function with asymptotically optimal weight matrix, and an empirical likelihood criterion function. For a homoskedastic nonlinear regression model, Assumptions RQ2(i) and RQ3 hold with $s(\gamma_0)$ equal to the error variance σ^2 and $\widehat{s}_n$ equal to a consistent estimator of σ^2, such as the sample variance based on the residuals.

4.6. *Asymptotic Distribution of the QLR Statistic*

Now we determine the asymptotic distribution of the QLR statistic under the sequences $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ and $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ when the null hypotheses are true, that is, when $v = v_n = r(\theta_n)$ for $\gamma_n = (\theta_n, \phi_n)$ $\forall n \geq 1$. These results are needed to obtain asymptotic size results for QLR-based CS's. The results for the QLR statistic rely on results for the restricted estimator $\widetilde{\theta}_n(v_n)$. These results are complicated by the fact that not all values $\pi \in \Pi$ are necessarily consistent with the restrictions $r((\psi_n, \pi)) = v_n$. For brevity, results for the restricted estimators are stated in Supplemental Appendix B.

We use the notational simplifications

$$\text{(4.11)} \qquad \text{QLR}_n = \text{QLR}_n(v_n) \quad \text{and} \quad \widetilde{\theta}_n = \widetilde{\theta}_n(v_n), \quad \text{where} \\ v_n = r(\theta_n) \quad \text{and} \quad \gamma_n = (\theta_n, \phi_n).$$

The matrix $r_\theta(\theta)$ of partial derivatives of $r(\theta)$ can be written as

$$\text{(4.12)} \qquad r_\theta(\theta) = \frac{\partial}{\partial \theta'} r(\theta) = \begin{bmatrix} r_{1,\psi}(\psi) & 0_{d_{r_1} \times d_\pi} \\ 0_{d_{r_2} \times d_\psi} & r_{2,\pi}(\pi) \end{bmatrix},$$

where $r_{1,\psi}(\psi) = (\partial/\partial\psi')r_1(\psi) \in R^{d_{r_1} \times d_\psi}$ and $r_{2,\pi}(\pi) = (\partial/\partial\pi')r_2(\pi) \in R^{d_{r_2} \times d_\pi}$.

For notational simplicity, let $\Pi_{r,0} = \Pi_r(v_{0,2})$, where $v_{0,2} = r_2(\pi_0)$ and $\gamma_0 = (\theta_0, \phi_0) \in \Gamma$. That is, $\Pi_{r,0}$ is the set of values π that are compatible with the restrictions on π when γ_0 is the true parameter value.

Next, we introduce the limit under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $\|b\| < \infty$ of the restricted concentrated criterion function after suitable normalization. For

$\pi \in \Pi$, define

$$\text{(4.13)} \quad \xi_r(\pi; \gamma_0, b) = \xi(\pi; \gamma_0, b) + \frac{1}{2}\tau(\pi; \gamma_0, b)' P_\psi(\pi; \gamma_0)' \times H(\pi; \gamma_0) P_\psi(\pi; \gamma_0) \tau(\pi; \gamma_0, b), \quad \text{where}$$

$$P_\psi(\pi; \gamma_0) = H^{-1}(\pi; \gamma_0) r_{1,\psi}(\psi_0)' (r_{1,\psi}(\psi_0) \times H^{-1}(\pi; \gamma_0) r_{1,\psi}(\psi_0)')^{-1} r_{1,\psi}(\psi_0)$$

and $\tau(\pi; \gamma_0, b)$ is defined in (3.9). The $d_\psi \times d_\psi$ matrix $P_\psi(\pi; \gamma_0)$ is an oblique projection matrix that projects onto the space spanned by the rows of $r_{1,\psi}(\psi_0)$.

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

THEOREM 4.2: *Suppose Assumptions* A, B1–B3, C1–C5, RQ1, *and* RQ3 *hold*. *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ *with* $\|b\| < \infty$, $\text{QLR}_n \to_d 2(\inf_{\pi \in \Pi_{r,0}} \xi_r(\pi; \gamma_0, b) - \inf_{\pi \in \Pi} \xi(\pi; \gamma_0, b))/s(\gamma_0)$.

COMMENTS: (i) Using Theorem 4.2, Figure 5 provides the asymptotic and finite-sample ($n = 250$) densities of the QLR statistic for tests concerning the MA parameter π in the ARMA(1, 1) model for $\pi_0 = 0.4$ and $b = 0, -2, -4$, and -12. The black line in Figure 5 is the χ^2_1 density, which is the strong-

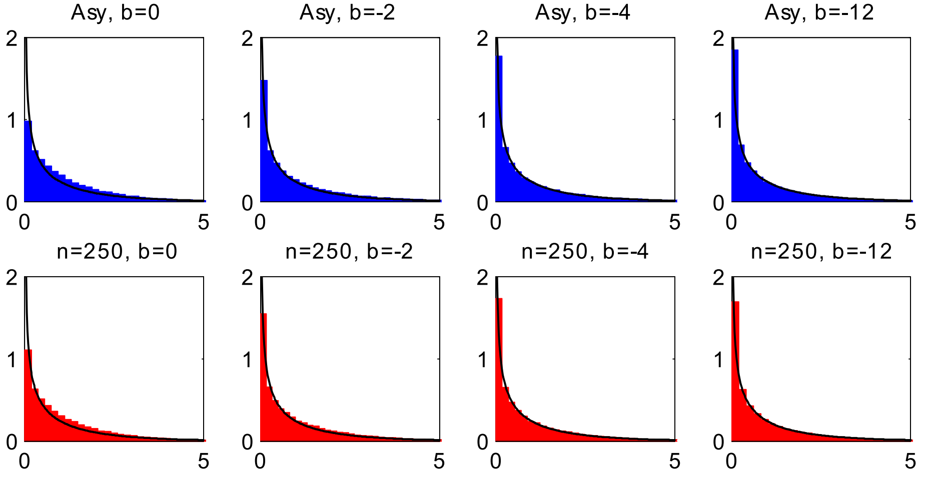

FIGURE 5.—Asymptotic and finite-sample ($n = 250$) densities of the QLR statistic for the MA parameter π in the ARMA(1, 1) model when $\pi_0 = 0.4$ and with χ^2_1 density (black line).

identification asymptotic density of the QLR statistic. Figure 5 indicates that the QLR statistic is well approximated by a χ_1^2 distribution, even under weak identification. This suggests that the QLR statistic yields tests and CI's that are substantially less sensitive to weak identification than t-based tests and CI's are.

(ii) Figure 4(b) provides graphs of the 0.95 asymptotic quantiles of the QLR statistic for π as a function of $|b|$. For small to medium $|b|$ values, the graphs exceed the 0.95 quantile under strong identification (given by the horizontal black line). Thus, tests and CI's based on the standard critical values (from the χ_1^2 distribution) have incorrect asymptotic size. For the QLR statistic the exceedance is much smaller than for the $|t|$ statistic. For the QLR statistic, for $\pi_0 = 0.8$ and $b = 0$, the quantile is roughly 4.4, whereas for strong identification it is roughly 3.8.

(iii) The proof of Theorem 4.2 requires an extension of the arg max theorem (e.g., see Lemma 3.2.1 of van der Vaart and Wellner (1996, p. 286)) to the case where the maximum is taken over a sample-size dependent sequence of sets.[43] See Lemma 9.10 in Supplemental Appendix B. This lemma may be of use in other contexts.

(iv) Assumption RQ1(iii) rules out the case where any single restriction depends on both ψ and π, but, in some cases, a reparametrization can be used to obtain results for such restrictions. Suppose $d_\pi = d_\beta$. Consider restrictions of the form $r(\theta) = (r_1(\psi), \pi + \beta)$. In this case, the asymptotic distribution of the QLR statistic in Theorems 4.2 and 4.3 (below) is the same as its distribution when $r(\theta) = (r_1(\psi), \pi)$. We use this result in the ARMA$(1, 1)$ example to obtain CI's for the AR parameter, which equals $\pi + \beta$.[44]

(v) The proof of Theorem 4.2 can be altered easily to yield some results for the QLR test under sequences of alternative hypothesis distributions, which yield asymptotic power results for QLR-based tests. Suppose the restrictions $r(\theta)$ depend only on π, that is, $d_{r_1} = 0$ and $r(\theta) = r_2(\pi)$. The sequence of true values of $r_2(\pi)$ satisfies $r_2(\pi_n) \to r_2(\pi_0) = v_{0,2}$ as $n \to \infty$. Now suppose the null hypothesis value of $r_2(\pi)$ is $v_{0,2}^{\text{null}}$, where $v_{0,2}^{\text{null}} \neq v_{0,2}$. Then the asymptotic distribution of QLR_n for this null hypothesis under the alternative hypothesis distributions $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ is given by the expression in Theorem 4.2, but with $\Pi_{r,0} = \Pi_r(v_{0,2})$ replaced by $\Pi_r(v_{0,2}^{\text{null}})$. This covers both local and fixed alternatives.

(vi) The proof of Theorem 4.2 makes use of the approach of Chernoff (1954).

[43]The arg max/min theorem provides the asymptotic distribution of a maximizer/minimizer of a stochastic process that converges weakly to some limit process.

[44]See Section 9.4.4 of Supplemental Appendix B for more details.

Next, we give results for the QLR statistic under sequences $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$. Define

$$\lambda_{\mathrm{QLR}}(\gamma_0) = G^*(\gamma_0)' J^{-1}(\gamma_0) P_\theta(\gamma_0)' J(\gamma_0) \times P_\theta(\gamma_0) J^{-1}(\gamma_0) G^*(\gamma_0), \quad \text{where}$$
$$P_\theta(\gamma_0) = J^{-1}(\gamma_0) r_\theta(\theta_0)' (r_\theta(\theta_0) J^{-1}(\gamma_0) r_\theta(\theta_0)')^{-1} r_\theta(\theta_0) \tag{4.14}$$

and $J(\gamma_0)$ and $G^*(\gamma_0)$ are defined in Assumptions D2 and D3. The matrix $P_\theta(\gamma_0)$ is an oblique projection matrix that projects onto the space spanned by the rows of $r_\theta(\theta_0)$.

THEOREM 4.3: *Suppose Assumptions* A, B1–B3, C1–C5, C7, C8, D1–D3, RQ1, *and* RQ3 *hold. Under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $\mathrm{QLR}_n \to_d \lambda_{\mathrm{QLR}}(\gamma_0)/s(\gamma_0)$.

COMMENT: When Assumption RQ2 holds, by Theorem 4.3 and some calculations, under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$,

$$\mathrm{QLR}_n \to_d \lambda_{\mathrm{QLR}}(\gamma_0)/s(\gamma_0) \sim \chi^2_{d_r}. \tag{4.15}$$

4.7. *Asymptotic Size of Standard t and QLR Confidence Sets*

Now, we establish the asymptotic size of standard CS's obtained by inverting t and QLR statistics using Lemma 2.1 and Theorems 4.1–4.3. The standard nominal $1-\alpha$ symmetric two-sided t, upper one-sided t, lower one-sided t, and QLR CS's take the form in (2.4) with $\mathcal{T}_n(v) = |T_n(v)|$, $T_n(v)$, $-T_n(v)$, and $\mathrm{QLR}_n(v)$, respectively, and $c_{n,1-\alpha}(v) = z_{1-\alpha/2}$, $z_{1-\alpha}$, $z_{1-\alpha}$, and $\chi^2_{d_r,1-\alpha}$, where $T_n(v)$ is defined in (4.2), $\mathrm{QLR}_n(v)$ is defined in (4.9), $z_{1-\alpha}$ is the $1-\alpha$ quantile of a standard normal distribution, and $\chi^2_{d_r,1-\alpha}$ is the $1-\alpha$ quantile of the $\chi^2_{d_r}$ distribution.

For $h = (b, \gamma_0)$ with $\|b\| < \infty$ and H as in (2.8), define

$$T(h) = \begin{cases} T_\psi(\pi^*(\gamma_0, b); \gamma_0, b), & \text{if } d^*_\pi = 0, \\ T_\pi(\pi^*(\gamma_0, b); \gamma_0, b), & \text{if } d^*_\pi = 1, \end{cases}$$
$$\mathrm{QLR}(h) = 2\Big(\inf_{\pi \in \Pi_{r,0}} \xi_r(\pi; \gamma_0, b) - \inf_{\pi \in \Pi} \xi(\pi; \gamma_0, b)\Big)\Big/ s(\gamma_0). \tag{4.16}$$

As defined, $T(h)$ is the asymptotic distribution of T_n under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $\|b\| < \infty$ given in Theorem 4.1(a) or (b) depending on the rank of $r_\pi(\theta)$, which is denoted by d^*_π. Only one of the cases applies for any particular parameter of interest $r(\theta)$ and it is known which applies. Here, $\mathrm{QLR}(h)$ is the asymptotic distribution of QLR_n under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for $\|b\| < \infty$ given in Theorem 4.2.

Let $c_{|t|,1-\alpha}(h)$, $c_{t,1-\alpha}(h)$, $c_{-t,1-\alpha}(h)$, and $c_{\mathrm{QLR},1-\alpha}(h)$ denote the $1-\alpha$ quantiles of $|T(h)|$, $T(h)$, $-T(h)$, and $\mathrm{QLR}(h)$, respectively, for $h \in H$.

TRIPATHI, G. T. (1999): "A Matrix Extension of the Cauchy–Schwarz Inequality," *Economics Letters*, 63, 1–3. [2174]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer. [2170,2185]

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.