SUPPLEMENT TO "UNIFORM INFERENCE IN AUTOREGRESSIVE MODELS": SUPPLEMENTARY APPENDIX

BY ANNA MIKUSHEVA

This supplementary appendix contains proofs of some results stated in the paper. In particular, it provides a proof of a statement about strong approximation, and proofs of Lemmas 11 and 12 about the asymptotic approximations for a scheme of series. It also proves results stated in Remarks 2, 3, and 4 for AR(1) processes with a linear time trend. Section S5 proves the validity of parametric and nonparametric grid bootstrap procedures for AR(p) processes with at most one root close to the unit circle. Section S7 contains an extensive Monte Carlo study of finite-sample properties of discussed methods. We keep the notation introduced in the paper.

S1. AN ARBITRARY VARIANCE

THIS SECTION CONTAINS the proof of the result stated in Section 2.3 of the paper.

Let $\widetilde{Y} = (\widetilde{y}_1, \ldots, \widetilde{y}_T)$ be a sample from an AR(1) process defined by an equation

$$(S1) \qquad \widetilde{y}_j = \widetilde{x}_j + c, \quad \widetilde{x}_j = \rho \widetilde{x}_{j-1} + \widetilde{\varepsilon}_j, \quad j = 0, \ldots, T, \quad \widetilde{x}_0 = 0.$$

ASSUMPTIONS A1—Repeated: *Let $(\widetilde{\varepsilon}_j, \mathcal{F}_j)$ be a martingale difference sequence with $E(\widetilde{\varepsilon}_j^2 | \mathcal{F}_{j-1}) = \sigma^2$ and $\sup_j E(|\widetilde{\varepsilon}_j|^r | \mathcal{F}_{j-1}) < \infty$ almost surely for some $2 < r \le 4$.*

Note that if the variance of error terms σ^2 is known, then the process $y_j = \widetilde{y}_j / \sigma$ is an AR(1) process with errors $\varepsilon_j = \widetilde{\varepsilon}_j / \sigma$ satisfying the set of Assumptions A and all inferences can be made using the three methods discussed in the paper.

Let $\widehat{e}_j = \widetilde{y}_j^\mu - \widehat{\rho}_{\text{OLS}} \widetilde{y}_{j-1}^\mu$ be the ordinary least squares (OLS) residuals. Let us define an estimator of σ^2 to be a sample variance of the OLS residuals: $\widehat{\sigma}^2 = \frac{1}{T} \sum_{j=1}^T \widehat{e}_j^2$. Despite the fact that the estimator $\widehat{\rho}_{\text{OLS}}$ of the autoregression (AR) coefficient is biased toward zero, the estimator $\widehat{\sigma}^2$ of the variance is uniformly consistent.

Let us define Studentized statistics

$$(\widetilde{S}, \widetilde{R}) = \left(\frac{1}{\sqrt{g(T, \rho)} \widehat{\sigma}^2} \sum_{j=1}^T \widetilde{y}_{j-1}^\mu \widetilde{\varepsilon}_j, \; \frac{1}{g(T, \rho) \widehat{\sigma}^2} \sum_{j=1}^T (\widetilde{y}_{j-1}^\mu)^2 \right).$$

LEMMA S1—Lemma 3 from the Paper: *Let us consider a model* (S1) *with error terms satisfying the set of Assumptions* A1. *Then, for every $\varepsilon > 0$,*

$$\lim_{T \to \infty} \sup_{\sigma > 0} \sup_{\rho \in \Theta_T} P\left\{ \left| \frac{\widehat{\sigma}^2}{\sigma^2} - 1 \right| > \varepsilon \right\} = 0.$$

1

Any statistic $\varphi(\widetilde{Y}, T, \rho) = \phi(\widetilde{S}, \widetilde{R}, T, \rho)$ for $\phi \in H$ is uniformly approximated by the corresponding statistic $\varphi_1 = \phi(S, R, T, \rho)$, where the pair (S, R) is defined for the process $y_j = \widetilde{y}_j / \sigma$. In particular, the three methods discussed in the paper could be used to make inferences.

PROOF: We note that $\widehat{e}_j - \widetilde{\varepsilon}_j^\mu = (\widehat{\rho}_{\mathrm{OLS}} - \rho)\widetilde{y}_{j-1}^\mu$. As a result,

$$\widehat{\sigma}^2 = \frac{1}{T}\sum_{j=1}^T (\widetilde{\varepsilon}_j^\mu)^2 + (\widehat{\rho}_{\mathrm{OLS}} - \rho)^2 \frac{1}{T}\sum_{j=1}^T (\widetilde{y}_{j-1}^\mu)^2$$

$$+ 2(\widehat{\rho}_{\mathrm{OLS}} - \rho)\frac{1}{T}\sum_{j=1}^T \widetilde{y}_{j-1}^\mu \widetilde{\varepsilon}_j$$

$$\frac{\widehat{\sigma}^2}{\sigma^2} - 1 = \left(\frac{1}{T}\sum_{j=1}^T (\varepsilon_j^\mu)^2 - 1\right) + 3\frac{1}{T}\frac{\left(\frac{1}{\sqrt{g(T,\rho)}}\sum_j y_{j-1}^\mu \varepsilon_j\right)^2}{\frac{1}{g(T,\rho)}\sum_j (y_{j-1}^\mu)^2}.$$

It is easy to see that all four terms converge to zero in probability uniformly over $\rho \in \Theta_T$ and uniformly over all values of $\sigma^2 > 0$.

From the definition of the class of functions H, we have

$$P\{|\phi(S, R, \rho) - \phi(\widetilde{S}, \widetilde{R}, \rho)| > x\}$$

$$\leq P\{|R| < C\} + P\{M_c(|S - \widetilde{S}| + |R - \widetilde{R}|) > x\}.$$

From the uniform approximation of R by R^N and Lemma 10, we know that R is uniformly separated from zero. It is easy to note that $\widetilde{S} - S = S(\frac{\sigma^2}{\widehat{\sigma}^2} - 1)$ and $\widetilde{R} - R = R(\frac{\sigma^2}{\widehat{\sigma}^2} - 1)$. By combining these facts with uniform consistency of the variance estimator, we obtain the statement of the lemma.   *Q.E.D.*

S2. ABOUT STRONG APPROXIMATION

LEMMA S2: *Let $(\varepsilon_j, \mathcal{F}_j)$ be a martingale difference sequence satisfying the set of Assumptions* A. *Let $S_j = \sum_{i=1}^j \varepsilon_i$ be partial sums. Then we can construct a sequence of processes $\eta_T(t) = \frac{1}{\sqrt{T}} S_{[tT]}$ and a sequence of Brownian motions w_T on a common probability space so that for every $\varepsilon > 0$, we have*

$$\sup_{0 \leq t \leq 1} |\eta_T(t) - w_T(t)| = o(T^{-1/2 + 1/r + \varepsilon}) \quad \text{a.s.}$$

PROOF: According to Lemma 6.2 from Park and Phillips (1999), conditions of the lemma imply the existence of an increasing sequence of stopping times $\{\tau_i\}_{i \geq 1}$ and a Brownian motion $w(\cdot)$ defined on the same probability space such that $\{S_j\} =^d \{w(\tau_j)\}$ and $\sup_{1 \leq j \leq T} |(\tau_j - j)/T^\delta| \to 0$ almost surely as $T \to \infty$, for any $\delta > 2/r$.

Similar to the proof of Theorem 2.2.4 in Csörgő and Révész (1981), it is easy to show that $\sup_{0 \le s \le T}(|w(\tau_{[s]}) - w(s)|)/(T^{\delta/2}\sqrt{\log T}) \to 0$ almost surely. This implies that for every $\varepsilon > 0$,

$$\sup_{0 \le t \le 1} \frac{|w(\tau_{[tT]})/\sqrt{T} - w_T(t)|}{T^{-1/2+1/r+\varepsilon}} \to 0 \quad \text{a.s.,}$$

where $w_T(t) = w(tT)/\sqrt{T}$. We define $\eta_T(t) = w(\tau_{[tT]})/\sqrt{T}$, which completes the proof of Lemma S2. $Q.E.D.$

S3. AR(1) MODEL WITH A LINEAR TIME TREND

This section shows that all results could be generalized to a model with a linear time trend. We prove statements in Remarks 2 and 3. Let us consider a process $y_j = a + bj + x_j$, where $x_j = \rho x_{j-1} + \varepsilon_j$. Then the modified test statistics are

$$(S^\tau, R^\tau) = \left(\frac{1}{\sqrt{g^\tau(T, \rho)}} \sum_{j=1}^T y_{j-1}^\tau (y_j - \rho y_{j-1}), \frac{1}{g^\tau(T, \rho)} \sum_{j=1}^T (y_{j-1}^\tau)^2 \right),$$

where y_t^τ denotes the detrended version of y_t: $y_{j-1}^\tau = y_{j-1} - \overline{y} - (\sum_{i=1}^T (y_{i-1} - \overline{y})i)(j - \frac{T+1}{2})/(\sum_{i=1}^T (i - \frac{T+1}{2})^2)$. The normalizing function is calculated as the mathematical expectation $g^\tau(T, \rho) = E_\rho \sum_{j=1}^T (y_{j-1}^\tau)^2$. Then the pair (S^τ, R^τ) is invariant with respect to the values of constants a and b.

Let $(S^{\tau,N}, R^{\tau,N})$ be the corresponding detrended version of the statistics generated in a model with normal errors.

LEMMA S3: *Assume that we have an* AR(1) *model with a linear trend and error terms satisfying the set of Assumptions* A. *Then for any function* $\phi \in H$ *we have that*

$$\lim_{T \to \infty} \sup_{\rho \in \Theta_T} \sup_x \big| P\{\phi(S^\tau, R^\tau, T, \rho) < x\}$$

$$- P\{\phi(S^{\tau,N}, R^{\tau,N}, T, \rho) < x\} \big| = 0.$$

PROOF: Our proof follows the framework suggested in Lemma 2. We start by checking conditions 2 and 3 of Lemma 2:

$$g^\tau(T, \rho) = E\left(\sum_{j=1}^T \left(y_{j-1}^\mu - \left(j - \frac{T+1}{2} \right) \frac{\sum_{i=1}^T y_{i-1}^\mu i}{\sum_{i=1}^T (i - \frac{T+1}{2})^2} \right)^2 \right)$$

$$= g(T, \rho) - E \frac{(\sum_{i=1}^T y_{i-1}^\mu i)^2}{\sum_{i=1}^T (i - \frac{T+1}{2})^2}.$$

It is easy to see that uniformly over $\mathcal{B}_T$ we have $\lim_{T \to \infty} \sup_{\rho \in \mathcal{B}_T} |g^\tau(T, \rho)/g(T, \rho) - 1| = 0$. We note that

$$S^\tau(T, \rho) = \sqrt{\frac{g(T, \rho)}{g^\tau(T, \rho)}} S(T, \rho)$$

$$- \left(\frac{\sum_{i=1}^T y_{i-1}^\mu i}{\sqrt{g^\tau(T, \rho)} \sqrt{\sum_{i=1}^T (i - \frac{T+1}{2})^2}} \right) \frac{\sum_{i=1}^T \varepsilon_i^\mu i}{\sqrt{\sum_{i=1}^T (i - \frac{T+1}{2})^2}}.$$

We can see that the term $\sum_{i=1}^T y_{i-1}^\mu i / (\sqrt{g^\tau(T, \rho)} \sqrt{\sum_{i=1}^T (i - \frac{T+1}{2})^2})$ converges to zero in probability uniformly over $\mathcal{B}_T$ by taking the mathematical expectation of its square and using Chebyshev's inequality. The term $\sum_{i=1}^T \varepsilon_i^\mu i / (\sum_{i=1}^T (i - \frac{T+1}{2})^2)^{1/2}$ is asymptotically normal. In the proof of Theorem 1, we showed that the distribution of $S(T, \rho)$ is asymptotically approximated by the standard normal distribution uniformly over $\mathcal{B}_T$. This implies that condition 2 of Lemma 2 is satisfied for the pair of statistics S^τ and $S^{\tau,N}$.

It is easy to see that

$$R^\tau(T, \rho) = \frac{g(T, \rho)}{g^\tau(T, \rho)} R(T, \rho) - \frac{1}{g^\tau(T, \rho)} \frac{(\sum_{i=1}^T y_{i-1}^\mu i)^2}{\sum_{i=1}^T (i - \frac{T+1}{2})^2}.$$

Since the second term converges to zero in probability uniformly over $\mathcal{B}_T$, we have that condition 3 of Lemma 2 is satisfied for statistics R^τ and $R^{\tau,N}$.

In the end, we are checking the closeness of the pairs (S^τ, R^τ) and $(S^{\tau,N}, R^{\tau,N})$ in proximity to the unit root. From discrete integration by parts, it is easy to see that

$$\left| \frac{1}{T^{3/2}} \sum_{j=1}^T \varepsilon_j j - \frac{1}{T^{3/2}} \sum_{j=1}^T e_{T,j} j \right|$$

$$= \left| \frac{1}{T} \sum_{j=1}^T \eta_T \left(\frac{j}{T} \right) - \frac{1}{T} \sum_{j=1}^T w_T \left(\frac{j}{T} \right) \right|$$

$$\leq \sup_{0 \leq t \leq 1} |\eta_T(t) - w_T(t)| \frac{1}{T} \sum_{j=1}^T 1 = o(T^{-1/2 + 1/r + \varepsilon}) \quad \text{a.s.}$$

By simple algebraic transformations, we have

$$\frac{1}{T} \sum_{j=1}^T y_{j-1}^\tau \varepsilon_j^\tau = \frac{1}{T} \sum_{j=1}^T y_{j-1} \varepsilon_j - \left(\frac{1}{T^{3/2}} \sum_{j=1}^T y_{j-1} \right) \frac{\sum_{j=1}^T \varepsilon_j}{\sqrt{T}}$$

$$- \left(\frac{1}{T^{5/2}} \sum_{j=1}^{T} y_{j-1} j - \frac{1}{T^{5/2}} \sum_{j=1}^{T} y_{j-1} \frac{T+1}{2} \right)$$

$$\times \left(\frac{1}{T^{5/2}} \sum_{j=1}^{T} \varepsilon_{j-1} j - \frac{1}{T^{5/2}} \sum_{j=1}^{T} \varepsilon_{j-1} \frac{T+1}{2} \right)$$

$$\times \frac{T^3}{\sum (i - \frac{T+1}{2})^2}.$$

By using statements (d) and (f) of Lemma 4, we can see that

$$\sup_{\rho \in \Theta_T} \frac{1}{(1+\rho)T+1} \left| \frac{1}{T} \sum_{j=1}^{T} y_{j-1}^{\tau} \varepsilon_j^{\tau} - \frac{1}{T} \sum_{j=1}^{T} z_{j-1}^{\tau} e_j^{\tau} \right| = o(T^{-1/2+1/r+\varepsilon}) \quad \text{a.s.}$$

Similarly,

$$\frac{1}{T^2} \sum_{j=1}^{T} (y_{j-1}^{\tau})^2 = \frac{1}{T^2} \sum_{j=1}^{T} (y_{j-1})^2 - \left(\frac{1}{T^{3/2}} \sum_{j=1}^{T} y_{j-1} \right)^2$$

$$- \left(\frac{1}{T^{5/2}} \sum_{j=1}^{T} y_{j-1} j - \frac{1}{T^{5/2}} \sum_{j=1}^{T} y_{j-1} \frac{T+1}{2} \right)^2$$

$$\times \frac{T^3}{\sum (i - \frac{T+1}{2})^2}.$$

From statements (e) and (f) of Lemma 4, we have

$$\sup_{\rho \in \Theta_T} \left| \frac{1}{T^2} \sum_{j=1}^{T} (y_{j-1}^{\tau})^2 - \frac{1}{T^2} \sum_{j=1}^{T} (z_{j-1}^{\tau})^2 \right| = o(T^{-1/2+1/r+\varepsilon}) \quad \text{a.s.}$$

Since we have $\sup_{\rho \in \mathcal{A}_T^{\pm}} (T^2/(g^{\tau}(T, \rho))) = O(T^{1-\alpha})$, condition 1 of Lemma 2 is satisfied for $\frac{3}{4} + \frac{1}{2r} < \alpha < 1$.   *Q.E.D.*

Let the local-to-unity statistics be

$$(S^{\tau,c}, R^{\tau,c}) = \left(\frac{1}{\sqrt{g^{\tau}(c)}} \int_0^1 J_c^{\tau}(x) \, dw(x), \frac{1}{g^{\tau}(c)} \int_0^1 (J_c^{\tau}(x))^2 \, dx \right),$$

where $J_c^{\tau}(x) = J_c(x) - \int_0^1 (4 - 6r) J_c(r) \, dr - x \int_0^1 (12r - 6) J_c(r) \, dr$ and $g^{\tau}(c) = E \int_0^1 (J_c^{\tau}(x))^2 \, dx$.

LEMMA S4: *Assume that we have an* AR(1) *model with a linear trend and the error terms satisfying the set of Assumptions* A. *Then for any function* $\phi \in H$, *we have that*

$$\lim_{T \to \infty} \sup_{\rho \in \Theta_T} \sup_x \big| P\{\phi(S^\tau, R^\tau, T, \rho) < x\}$$

$$- P\{\phi(S^{\tau, c(T,\rho)}, R^{\tau, c(T,\rho)}, T, \rho) < x\}\big| = 0,$$

where $c(T, \rho) = T \log(\rho)$.

PROOF: It is enough to show that

$$\lim_{T \to \infty} \sup_{\rho \in \Theta_T} \sup_x \big| P\{\phi(S^{\tau, N}, R^{\tau, N}, T, \rho) < x\}$$

$$- P\{\phi(S^{\tau, c(T,\rho)}, R^{\tau, c(T,\rho)}, T, \rho) < x\}\big| = 0.$$

We check that the conditions of Lemma 2 are satisfied. By simple algebraic manipulation, we have

$$J_c^\tau(x) = J_c^\mu(x) - 6(1/2 - x) \int_0^1 (1/2 - r) J_c^\mu(r) \, dr.$$

It is easy to see that $E(\int_0^1 (1/2 - r) J_c^\mu(r) \, dr)^2 \leq 1/c^2$. As a result, we have $\lim_{c \to -\infty} (1/g(c)) E(\int_0^1 (1/2 - r) J_c^\mu(r) \, dr)^2 = 0$ and $\lim_{c \to -\infty} |g(c)/g^\tau(c) - 1| = 0$.

By using Chebyshev's inequality, we can also note that $(1/\sqrt{g(c)}) \int_0^1 (1/2 - r) J_c^\mu(r) \, dr \to^p 0$ as $c \to -\infty$. This implies that

$$S^{\tau, c} = \sqrt{\frac{g(c)}{g^\tau(c)}} S^c$$

$$- \frac{6}{\sqrt{g^\tau(c)}} \int_0^1 \left(\frac{1}{2} - x\right) dw(x) \int_0^1 \left(\frac{1}{2} - r\right) J_c^\mu(r) \, dr$$

$$\Rightarrow N(0, 1)$$

and

$$R^{\tau, c} = \frac{g(c)}{g^\tau(c)} R^c - \frac{1}{g^\tau(c)} \left(6 \int_0^1 \left(\frac{1}{2} - r\right) J_c^\mu(r) \, dr\right)^2 \to^p 1$$

$$\text{as} \quad c \to -\infty.$$

As a result, conditions 2 and 3 of Lemma 2 are satisfied for the pairs $(S^{\tau, c(T,\rho)}, R^{\tau, c(T,\rho)})$ and $(S^{\tau, N}, R^{\tau, N})$.

Now we check condition 1 of Lemma 2 for the detrended pairs:

$$\frac{1}{T^{5/2}} \sum_{j=1}^{T} z_{j-1}^{\mu} \left(j - \frac{T+1}{2} \right) = \frac{1}{T} \sum_{j=1}^{T} \frac{z_{j-1}}{\sqrt{T}} \frac{(j - \frac{T+1}{2})}{T}$$

$$= \int_0^1 \left(\frac{[tT]}{T} - \frac{1}{2} \right) \int_0^t e^{c/T([Tt]-[Ts]-1)} I\left\{ s \leq \frac{[Tt]}{T} \right\} dw(s)\, dt$$

$$= \int_0^1 \int_s^1 \left(\frac{[tT]}{T} - \frac{1}{2} \right) e^{c/T([Tt]-[Ts]-1)} I\left\{ s \leq \frac{[Tt]}{T} \right\} dt\, dw(s).$$

Similarly, $\int_0^1 (t - 1/2) J_c^{\mu}(t)\, dt = \int_0^1 \int_s^1 (t - 1/2) e^{c(t-s)}\, dt\, dw(s)$. As a result,

$$E\left(\frac{1}{T^{5/2}} \sum_{j=1}^{T} z_{j-1}^{\mu} \left(j - \frac{T+1}{2} \right) - \int_0^1 \left(t - \frac{1}{2} \right) J_c^{\mu}(t)\, dt \right)^2$$

$$= \int_0^1 \left(\int_s^1 \left(\frac{[tT]}{T} - \frac{1}{2} \right) e^{c/T([Tt]-[Ts]-1)} I_{\{s \leq [Tt]/T\}}\, dt \right.$$

$$\left. - \int_s^1 \left(t - \frac{1}{2} \right) e^{c(t-s)}\, dt \right)^2 ds$$

$$\leq \text{const}(\log(\rho))^2.$$

Taking into account that $\sup_{\rho \in \mathcal{A}_T^+} (T^2 / (g^{\tau}(T, \rho))) = O(T^{1-\alpha})$ and $\lim_{t \to \infty} \sup_{\rho \in \mathcal{A}_T^+} ((T^2 g^{\tau}(c(T, \rho))) / (g^{\tau}(T, \rho))) = 1$, we have

$$\lim_{T \to \infty} \sup_{\rho \in \mathcal{A}_T} P\left\{ \left| \frac{1}{\sqrt{g^{\tau}(T, \rho)} T^{3/2}} \sum_{j=1}^{T} z_{j-1}^{\mu} j \right. \right.$$

$$\left. \left. - \frac{1}{\sqrt{g^{\tau}(c(T, \rho))}} \int_0^1 \left(t - \frac{1}{2} \right) J_c^{\mu}(t)\, dt \right| > x \right\} = 0.$$

It is easy to determine that

$$\lim_{T \to \infty} \sup_{\rho \in \mathcal{A}_T} P\left\{ \left| \frac{1}{\sqrt{g^{\tau}(T, \rho)} T^{1/2}} \sum_{j=1}^{T} e_{j-1}^{\mu} j \right. \right.$$

$$\left. \left. - \frac{1}{\sqrt{g^{\tau}(c(T, \rho))}} \int_0^1 \left(t - \frac{1}{2} \right) dw(t) \right| > x \right\} = 0.$$

We note that

$$S^{\tau,N}(T,\rho) = \sqrt{\frac{g(T,\rho)}{g^{\tau}(T,\rho)}} S^{N}(T,\rho)$$

$$- \left(\frac{1}{\sqrt{g^{\tau}(T,\rho)} T^{3/2}} \sum_{j=1}^{T} z_{j-1}^{\mu} \left(j - \frac{T+1}{2} \right) \right)$$

$$\times \left(\frac{1}{\sqrt{g^{\tau}(T,\rho)} T^{1/2}} \sum_{j=1}^{T} e_{j-1}^{\mu} \left(j - \frac{T+1}{2} \right) \right)$$

$$\times \left(\frac{T^3}{\sum_{j=1}^{T} (j - (T+1)/2)^2} \right),$$

$$S^{\tau,c(T,\rho)} = \sqrt{\frac{g(T,\rho)}{g^{\tau}(T,\rho)}} S^{c(T,\rho)}$$

$$- 6 \left(\frac{1}{\sqrt{g^{\tau}(c(T,\rho))}} \int_0^1 \left(t - \frac{1}{2} \right) J_c^{\mu}(t)\, dt \right)$$

$$\times \left(\frac{1}{\sqrt{g^{\tau}(c(T,\rho))}} \int_0^1 \left(t - \frac{1}{2} \right) dw(t) \right),$$

and

$$R^{\tau,N}(T,\rho) = \frac{g(T,\rho)}{g^{\tau}(T,\rho)} R^{N}(T,\rho)$$

$$- \left(\frac{1}{\sqrt{g^{\tau}(T,\rho)} T^{3/2}} \sum_{j=1}^{T} z_{j-1}^{\mu} \left(j - \frac{T+1}{2} \right) \right)^2$$

$$\times \left(\frac{T^3}{\sum_{j=1}^{T} (j - (T+1)/2)^2} \right)^2,$$

$$R^{\tau,c(T,\rho)} = \frac{g(T,\rho)}{g^{\tau}(T,\rho)} R^{c(T,\rho)} - \left(\frac{6}{\sqrt{g^{\tau}(c(T,\rho))}} \int_0^1 \left(t - \frac{1}{2} \right) J_c^{\mu}(t)\, dt \right)^2.$$

Since in Theorem 2 we proved that condition 1 is satisfied for the pairs $(S^{c(T,\rho)}, R^{c(T,\rho)})$ and (S^{N}, R^{N}), we have

$$\lim_{T \to \infty} \sup_{\rho \in \mathcal{A}_T} P\big\{ |S^{\tau,N}(T,\rho) - S^{\tau,c(T,\rho)}| + |R^{\tau,N}(T,\rho) - R^{\tau,c(T,\rho)}| > x \big\} = 0.$$

As a result, all conditions of Lemma 2 are satisfied.   *Q.E.D.*

S4. HANSEN'S BOOTSTRAP

This section proves some of the results stated in Section 5 of the paper. Lemma S5 below (stated in the paper as Lemma 11) shows that the normal approximation in the stationary region holds uniformly for arrays of random errors.

LEMMA S5—Lemma 11 from the Paper: *Let* $\{\varepsilon_{T,j}; j = 1, \ldots, T; T \in \mathbb{N}\}$ *be a triangular array of random variables such that for every* T, *variables* $\{\varepsilon_{T,j}\}_{j=1}^{T}$ *are independent and identically distributed with distribution* F_T. *Assume that* $y_{T,j} = \rho y_{T,j-1} + \varepsilon_{T,j}$. *Then, for any sequence* ρ_T *such that* $T(1 - \rho_T) \to \infty$, *we have*

$$\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} \sup_{|\rho| \leq \rho_T} \sup_{x} \left| P\left\{ \frac{1}{\sqrt{g(T,\rho)}} \sum_{j=1}^{T} y_{T,j-1}\varepsilon_{T,j} < x \right\} - \Phi(x) \right|$$
$$= 0$$

and, for every $\epsilon > 0$,

$$\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} \sup_{|\rho| \leq \rho_T} P\left\{ \left| \frac{1}{g(T,\rho)} \sum_{j=1}^{T} y_{T,j-1}^2 - 1 \right| > \epsilon \right\} = 0,$$

$$\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} \sup_{|\rho| \leq \rho_T} P\left\{ \left| \frac{1}{\sqrt{g(T,\rho)}\sqrt{T}} \sum_{j=1}^{T} y_{T,j-1} \right| > \epsilon \right\} = 0.$$

PROOF: This statement is a generalization of the main result of Giraitis and Phillips (2006) for arrays, where the distribution of error terms is allowed to be different for different sample sizes. First, we check the statement for variables that possesses a bounded fourth moment. Then we apply truncation methods to the case when variables may have infinite fourth moment.

Assume that $r = 4$. Let us define variables $X_j = (1/\sqrt{g(T,\rho)})y_{T,j-1}\varepsilon_{T,j}$ and $V_j^2 = \sum_{i=1}^{j} E(X_i^2|\mathcal{F}_{i-1}) = (1/g(T,\rho)) \sum_{j=1}^{T} y_{T,j-1}^2$. Then from the Corollary to Theorem 1 in Hall and Heyde (1981), it follows that

$$\sup_{x} \left| P\left\{ \frac{1}{\sqrt{g(T,\rho)}} \sum_{j=1}^{T} y_{T,j-1}\varepsilon_{T,j} < x \right\} - \Phi(x) \right|$$
$$\leq C\left(\sum_{j=1}^{T} E|X_j|^4 + E(V_T^2 - 1)^2 \right),$$

where C is an absolute constant.

We should note that

$$\sup_{|\rho|\le\rho_T}\sum_{j=1}^{T}E|X_j|^4 = \sup_{|\rho|\le\rho_T}C_1\frac{T}{g(T,\rho)^2}Ey_{T,j-1}^4$$

$$\le \sup_{|\rho|\le\rho_T}C_2\frac{T}{g(T,\rho)^2(1-\rho^2)^2} = \sup_{|\rho|\le\rho_T}\frac{CT}{\left(T-\frac{1-\rho^{2T}}{1-\rho^2}\right)^2}$$

$$\le \frac{CT}{\left(T-\frac{1-\rho_T^{2T}}{1-\rho_T^2}\right)^2} \le \frac{C}{\left(1-\frac{1}{T(1-\rho_T^2)}\right)^2}\frac{1}{T}$$

$$\to 0,$$

where C_1 is a constant depending on M, C_2 is a constant depending on M, and K, C is a constant depending on M, K and a sequence ρ_T.

Now let us estimate the second summand:

$$E(V_T^2-1)^2 = \frac{1}{g(T,\rho)^2}\left(\sum_{j=1}^{T}(y_{T,j-1}^2-Ey_{T,j-1}^2)\right)^2$$

$$= \frac{1}{g(T,\rho)^2}\left(\sum_{j=1}^{T}(y_{T,j-1}^2-Ey_{T,j-1}^2)^2\right.$$

$$\left.+ 2\sum_{j=1}^{T}\sum_{i=1}^{j}\rho^{2(j-i)}(y_{T,i-1}^2-Ey_{T,i-1}^2)^2\right)$$

$$\le \frac{1}{g(T,\rho)^2}\frac{C}{1-\rho^2}\left(\sum_{j=1}^{T}(y_{T,j-1}^2-Ey_{T,j-1}^2)^2\right)$$

$$\le \frac{1}{g(T,\rho)^2}\frac{C}{1-\rho^2}\left(\sum_{j=1}^{T}\sum_{k=1}^{j}\rho^{4(j-k)}E\varepsilon^4\right)$$

$$\le \frac{1}{g(T,\rho)^2}\frac{C}{1-\rho^2}\frac{T}{1-\rho^4}.$$

The last expression converges to zero uniformly over $\{|\rho|<\rho_T\}$. This completes the proof of asymptotic normality for the case when variables have a bounded fourth moment. It also proves the uniform convergence of $\frac{1}{g(T,\rho)}\times\sum_{j=1}^{T}y_{T,j-1}^2$ to 1 in probability. The last statement of the lemma can be checked by showing that

$$\lim_{T\to\infty}\sup_{|\rho|<\rho_T}E\left(\frac{1}{\sqrt{g(T,\rho)}\sqrt{T}}\sum_{j=1}^{T}y_{T,j-1}\right)^2 = 0$$

and applying Chebyshev's inequality.

The proof for the case when variables can have an infinite fourth moment follows from the truncation argument of the proof of Lemma 2.1 (part b) in Giraitis and Phillips (2006).   Q.E.D.

LEMMA S6—Lemma 12 from the Paper: *Let* $\{\varepsilon_{T,j}; j = 1, \ldots, T; T \in \mathbb{N}\}$ *be a triangular array of random variables such that for every* T, *the variables* $\{\varepsilon_{T,j}\}_{j=1}^T$ *are independent and identically distributed with cumulative distribution function* $F_T \in \mathcal{L}_r(K, M, \theta)$. *Then we can construct a process* $\eta_T(t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{[Tt]} \varepsilon_{T,j}$ *and Brownian motions* w_T *on a common probability space in such a way that for every* $\varepsilon > 0$, *we have*

$$\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} P\left\{ \sup_{0 \le t \le 1} |\eta_T(t) - w(t)| > \varepsilon T^{-\delta} \right\} = 0$$

for some $\delta > 0$.

PROOF: By Skorohod representation, for every T there exists an increasing sequence of stopping times $\tau_{T,1} \le \tau_{T,2} \le \cdots \le \tau_{T,T}$ such that:

1. $\{w(\tau_{T,j}) - w(\tau_{T,j-1})\}_{j=1}^T =^d \{\epsilon_{T,j}\}_{j=1}^T$;

2. $s_{T,j} = \tau_{T,j} - \tau_{T,j-1}$ are independent and identically distributed positive random variables with mean $E s_{T,j} = \mu_2(F_T)$ and $E|s_{T,j}|^{r/2} \le C_r \mu_r(F_T)$.

Let us define the process $\eta_T(t) = w(\tau_{T,[tT]}/T)$. Let a_T be a sequence of non-random positive numbers. Then

$$P\left\{ \sup_{0 \le t \le 1} |\eta_T(t) - w(t)| > \varepsilon T^{-\delta} \right\}$$

$$\le P\left\{ \sup_{0 \le t \le 1} \sup_{0 \le s \le a_T} |w(t+s) - w(t)| > \varepsilon T^{-\delta} \right\}$$

$$+ P\left\{ \sup_{0 \le t \le 1} \left| \frac{\tau_{[Tt]}}{T} - t \right| > a_T \right\}.$$

From Lemma 1.2.1 in Csörgő and Révész (1981), it follows that

$$P\left\{ \sup_{0 \le t \le 1} \sup_{0 \le s \le a_T} |w(t+s) - w(t)| > \varepsilon T^{-\delta} \right\} \le \frac{C}{a_T} \exp\left(-\frac{1}{3} \left(\frac{\varepsilon}{T^\delta \sqrt{a_T}} \right)^2 \right).$$

The right-hand side of the last inequality converges to zero for the sequence $a_T = T^{-\gamma}$ if $\gamma > \delta$. As a result, it is enough to prove that

$$\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} P\left\{ \sup_{0 \le t \le 1} \left| \frac{\tau_{T,[Tt]}}{T} - t \right| > T^{-\gamma} \right\} = 0$$

for some $\gamma > 0$. We can note that

$$\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} P\left\{ \sup_{0 \le t \le 1} \left| \frac{\tau_{T,[Tt]}}{T} - t \right| > T^{-\gamma} \right\}$$

$$\le \lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} P\left\{ \sup_{0 \le j \le T} |\tau_{T,j} - j\mu_2(F_T)| > T^{1-\gamma} \right\}$$

$$+ \lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} P\{ |\mu_2(F_T) - 1| > T^{-\gamma} \}.$$

The last term converges to 0 by definition of the class $\mathcal{L}_r(K, M, \theta)$ if $\gamma > \theta$.

From the results of Montgomery-Smith (1993), there exists an absolute constant $c > 0$ such that

$$P\left\{ \sup_{1 \le j \le T} \left| \sum_{i=1}^{j} s_{T,i} - j\mu_2(F_T) \right| > T^{1-\gamma} \right\}$$

$$\le cP\left\{ \left| \sum_{i=1}^{T} s_{T,i} - T\mu_2(F_T) \right| > \frac{T^{1-\gamma}}{c} \right\}.$$

By applying Theorem 27 from Petrov (1975, Chap. 9), we have

$$P\left\{ \left| \sum_{i=1}^{T} s_{T,i} - T\mu_2(F_T) \right| > \frac{T^{1-\gamma}}{c} \right\}$$

$$\le C_r \mu_r(F_T(Y_T))(TT^{-(1-\gamma)r/2} + T^{-6(1-\gamma)}T^{7-r/2}).$$

If we choose $0 < \delta < \gamma < \min\{(r/2 - 1)/6, (r/2 - 1)2/r\}$, then we will achieve the required convergence.   Q.E.D.

The lemma below proves the result stated in Remark 4.

LEMMA S7: *Assume that we have an* AR(1) *model with a linear trend and the error terms satisfying the set of Assumptions* A. *Let* $y_{T,j} = \rho y_{T,j-1} + \varepsilon_{T,j}$, *where* $\varepsilon_{T,j}$ *are independent and identically distributed random variables with distribution function* $F^{\mathrm{res}}(x|\Sigma_T, \rho)$, *which is an empirical distribution function of residuals. Then for any function* $\phi \in H$, *we have that for almost all realizations of error term* Σ,

$$\lim_{T \to \infty} \sup_{\rho \in \Theta_T} \sup_{x} \left| P\{ \phi(S^\tau, R^\tau, T, \rho) < x \} \right.$$

$$\left. - P\{ \phi(S^{\tau,*}, R^{\tau,*}, T, \rho) < x | Y \} \right| = 0,$$

where the pair of statistics $(S^{\tau,*}, R^{\tau,*})$ is detrended statistics for the sample $Y^* = (y_{t,1}, \ldots, y_{T,T})$.

PROOF: The proof consists of two steps. First, we show that

(S2) $$\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} \sup_{\rho \in \Theta_T} \sup_x \left| P\{\phi(S^{\tau}, R^{\tau}, T, \rho) < x\} \right.$$

$$\left. - P\{\phi(S^{\tau,*}, R^{\tau,*}, T, \rho) < x\} \right| = 0.$$

In the second step, we check that for almost all realizations of error term Σ, there are constants $K(\Sigma)$ and $M(\Sigma)$ such that $F_T^{\text{res}}(\cdot|\Sigma, \rho) \in \mathcal{L}_r(K, M, \theta)$.

Assume that $F_T \in \mathcal{L}_r(K, M, \theta)$. According to Lemma S6 there exists an almost sure approximation of the partial sum process by a sequence of Brownian motions. Following the proof of Lemma S3 it is easy to prove that

$$\lim_{T \to \infty} \sup_{\rho \in \mathcal{A}_T} P\{|S^{\tau,N}(T, \rho) - S^{\tau,*}| + |R^{\tau,N}(T, \rho) - R^{\tau,*}| > x\} = 0,$$

that is, condition 1 of Lemma 2 is satisfied.

The only things that need to be proved are uniform convergence of the distribution of the statistic $S^{\tau,*}$ to the standard normal uniformly over $\mathcal{B}_T$ and uniform convergence in probability of the statistic $R^{\tau,*}$ to 1 over $\mathcal{B}_T$.

From the proof of Theorem 3, it follows that

$$\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_\gamma(K,M,\theta)} \sup_{\rho \in \Theta_T} \sup_x \left| P\{S^*(T, \rho) < x\} - \Phi(x) \right| = 0$$

and $\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_\gamma(K,M,\theta)} \sup_{\rho \in \Theta_T} P\{|R^*(T, \rho) - 1| > \varepsilon\} = 0$ for every $\varepsilon > 0$.

It is enough to show that

$$\lim_{T \to \infty} \sup_{\rho \in \Theta_T} P\left\{ \left| \frac{\sum_{i=1}^T y_{T,i-1}^\mu i}{\sqrt{g^\tau(T, \rho)}\sqrt{\sum_{i=1}^T (i - \frac{T+1}{2})^2}} \right| > x \right\} = 0$$

and

$$\lim_{T \to \infty} \sup_{\rho \in \Theta_T} \left| P\left\{ \frac{\sum_{i=1}^T \varepsilon_{T,i}^\mu i}{\sqrt{\sum_{i=1}^T (i - \frac{T+1}{2})^2}} > x \right\} - \Phi(x) \right| = 0.$$

The first can easily be checked by Chebyshev's inequality. For the proof of the second, one can check conditions of Theorem 1 in Hall and Heyde (1981). As a result, conditions 2 and 3 of Lemma 2 are satisfied. According to Lemma 2, the uniform approximation (S2) holds for the detrended statistics.

Now we turn to the second step of the proof. We check that the residual based bootstrap produces F_T that belongs to the $\mathcal{L}_r(K, M, \theta)$ class. We

define the destribution function $F_T^{\text{res}}(x) = \frac{1}{T}\sum_{j=1}^{T} I_{\{\widehat{e}_j \leq x\}}$, where $\widehat{e}_j$ are residuals from the regression of y_j on a constant, linear trend j and y_{j-1}. Then $\mu_r(F_T) = \frac{1}{T}\sum_{j=1}^{T}(\widehat{e}_j)^r$.

The first condition of the class is trivially satisfied. For the third condition, we have

$$\frac{1}{T}\sum_{j=1}^{T}|\widehat{e}_j|^r \leq C_r \frac{1}{T}\sum_{j=1}^{T}|\widehat{e}_j - \varepsilon_j^\tau|^r + C_r \frac{1}{T}\sum_{j=1}^{T}|\varepsilon_j^\tau|^r.$$

Let us consider each term separately. The second term is bounded almost surely due to the strong law of large numbers. We note that $\widehat{e}_j - \varepsilon_j^\tau = \sum_{i=1}^{T}\varepsilon_i y_{i-1}^\tau / (\sum_{i=1}^{T}(y_{i-1}^\tau)^2) y_{j-1}^\tau$,

$$\frac{1}{T}\sum_{j=1}^{T}|\widehat{e}_j - \varepsilon_j^\tau|^r = \frac{1}{T}\left|\frac{\sum_{j=1}^{T}\varepsilon_j y_{j-1}^\tau}{\sum_{j=1}^{T}(y_{j-1}^\tau)^2}\right|^r \sum_{j=1}^{T}|y_{j-1}^\tau|^r$$

$$= \text{const}\frac{1}{T}\frac{|\sum_{j=1}^{T}\varepsilon_j y_{j-1}^\tau|^r}{(\sum_{j=1}^{T}(y_{j-1}^\tau)^2)^{r/2}} = \text{const}\frac{1}{T}\left(\frac{S^\tau(T,\rho)}{\sqrt{R^\tau(T,\rho)}}\right)^r$$

$$= o_p(T^{-1+\varepsilon})$$

for every $\varepsilon > 0$.

Now, we check the second condition for the residual based bootstrap:

$$\frac{1}{T}\sum_{t=1}^{T}(\widehat{e}_t)^2 - 1 = \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t^\tau)^2 - 1 + 3\frac{1}{T}\frac{S^\tau(T,\rho)^2}{R^\tau(T,\rho)}.$$

The last expression converges almost surely to zero with a nontrivial speed since $E|\varepsilon_j|^r < \infty$ for $r > 2$.   Q.E.D.

S5. UNIFORM INFERENCES IN AR(p) MODELS

S5.1. *About* AR(p) *Models*

In this section we consider an AR(p) model with at most one root close to the unit circle. Let us consider an AR(p) model in augmented Dickey–Fuller (ADF) form,

$$\text{(S3)}\qquad y_t = \rho y_{t-1} + \sum_{j=1}^{p-1}\alpha_j \Delta y_{t-j} + \varepsilon_t,$$

where error terms satisfy the set of Assumptions C.

ASSUMPTIONS C: *Let $\{\varepsilon_t\}_{t=1}^{\infty}$ be independent and identically distributed error terms with zero mean $E\varepsilon_t = 0$, unit variance $E\varepsilon_t^2 = 1$, and finite fourth moments $E\varepsilon_t^4 < \infty$.*

We restrict ourselves to processes with at most one root close to the unit circle. The process (S3) could be described by the equation $a(L)y_t = \varepsilon_t$, where $a(L) = 1 - \rho L - \sum_{j=1}^{p-1} \alpha_j (1-L)L^j$. Let us represent the polynomial as $a(L) = (1 - \mu_1 L) \cdots (1 - \mu_p L)$, where $|\mu_1| \leq |\mu_2| \leq \cdots \leq |\mu_p| < 1$. Let us fix $0 < \delta < 1$. For every $\rho \in (0, 1)$, we define $\mathcal{R}_\rho$ to be a set of all possible values of the nuisance parameter $\alpha = (\alpha_1, \ldots, \alpha_{p-1})$ such that $|\mu_{p-1}| < \delta$.

The lemma below demonstrates some properties of an AR(p) process with at most one root close to the unit circle.

LEMMA S8: *Assume that $a(L) = (1 - \mu_1 L) \cdots (1 - \mu_p L)$, where $|\mu_1| \leq |\mu_2| \leq \cdots \leq |\mu_{p-1}| < \delta < 1$. Let $\frac{1-L}{a(L)} = \sum_{j=0}^{\infty} c_j L^j$. Then:*
   (a) *$\sum_{j=0}^{\infty} |c_j| < C_1(\delta)$;*
   (b) *For $\gamma_j = \sum_{i=0}^{\infty} c_i c_{i+j}$, we have $|\sum_{j=0}^{\infty} \gamma_j| < C_2(\delta)$;*
   (c) *For $\Gamma_{i,j,k} = \sum_{t=0}^{\infty} c_t c_{t+i} c_{t+j} c_{t+k}$, we have $|\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \Gamma_{i,j,k}| < C_3(\delta)$, where $C_i(\delta)$ are constants depending only on δ and p, but not on the value of the roots.*

PROOF: (a) We have

$$\frac{(1-L)}{(1-\mu_1 L) \cdots (1-\mu_p L)} = (1-L)\left(\sum_{j=0}^{\infty} \mu_1^j L^j\right) \cdots \left(\sum_{j=0}^{\infty} \mu_p^j L^j\right)$$

$$= (1-L)$$

$$\times \sum_{j=0}^{\infty} \left(\sum_{k_1,k_2,\ldots,k_p:\sum_i k_i = j} \mu_1^{k_1} \mu_2^{k_2} \cdots \mu_p^{k_p}\right) L^j.$$

As a result, we have

$$c_j = \sum_{k_1,k_2,\ldots,k_p:\sum_i k_i = j} \mu_1^{k_1} \mu_2^{k_2} \cdots \mu_p^{k_p}$$

$$- \sum_{k_1,k_2,\ldots,k_p:\sum_i k_i = j-1} \mu_1^{k_1} \mu_2^{k_2} \cdots \mu_p^{k_p}$$

$$= -(1 - \mu_p) \sum_{k_1,k_2,\ldots,k_p:\sum_i k_i = j-1} \mu_1^{k_1} \mu_2^{k_2} \cdots \mu_p^{k_p}$$

$$+ \sum_{k_1,k_2,\ldots,k_{p-1}:\sum_i k_i = j} \mu_1^{k_1} \mu_2^{k_2} \cdots \mu_{p-1}^{k_{p-1}},$$

$$\sum_{j=0}^{\infty} |c_j| \leq |1 - \mu_p| \left(\sum_{j=0}^{\infty} |\mu_1|^j \right) \cdots \left(\sum_{j=0}^{\infty} |\mu_p|^j \right)$$

$$+ \left(\sum_{j=0}^{\infty} |\mu_1|^j \right) \cdots \left(\sum_{j=0}^{\infty} |\mu_{p-1}|^j \right)$$

$$\leq \frac{1}{(1-\delta)^{p-1}} \left(\frac{|1 - \mu_p|}{1 - |\mu_p|} + 1 \right) \leq \mathrm{const}(\delta).$$

This ends the proof of part (a).

(b) We have

$$\left| \sum_{j=0}^{\infty} \gamma_j \right| \leq \sum_{j=0}^{\infty} |\gamma_j| \leq \sum_{j=0}^{\infty} \left| \sum_{i=0}^{\infty} c_i c_{i+j} \right|$$

$$\leq \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} |c_i||c_{i+j}| \leq \left(\sum_{i=0}^{\infty} |c_i| \right)^2 \leq C_2(\delta).$$

(c) The proof is totally similar to that of part (b). *Q.E.D.*

LEMMA S9: *Let us have two* AR(p) *processes* $y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \alpha_j \Delta y_{t-j} + \varepsilon_t$ *and* $z_t = \rho z_{t-1} + \sum_{j=1}^{p-1} \beta_j \Delta z_{t-j} + \varepsilon_t$, *where error terms* ε_j *are independent and identically distributed standard normal random variables. Then we have*:

(a) $\sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{\beta \in \mathcal{R}_\rho} ((E_\rho(y_t - z_t)^2)/\mathrm{Var}(y_t)) \leq C(\delta)\|\alpha - \beta\|^2$;

(b) $\sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{\beta \in \mathcal{R}_\rho} ((E_\rho(\Delta y_t - \Delta z_t)^2)/\mathrm{Var}(\Delta y_t)) \leq C(\delta)\|\alpha - \beta\|^2$;

(c) $\sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{\beta \in \mathcal{R}_\rho} |(\mathrm{Var}(y_t)/\mathrm{Var}(z_t)) - 1| \leq C(\delta)\|\alpha - \beta\|$;

(d) $\sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{\beta \in \mathcal{R}_\rho} |(\mathrm{Var}(\Delta y_t)/\mathrm{Var}(\Delta z_t)) - 1| \leq C(\delta)\|\alpha - \beta\|$;

(e) $\sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{\beta \in \mathcal{R}_\rho} ((E_\rho(y_t^2 - z_t^2)^2)/(\mathrm{Var}(y_t))^2) \leq C(\delta)\|\alpha - \beta\|^2$;

(f) $\sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{\beta \in \mathcal{R}_\rho} ((E_\rho(\Delta y_t \Delta y_{t-j} - \Delta z_t \Delta z_{t-j})^2)/(\mathrm{Var}(\Delta yz_t))^2) \leq C(\delta)\|\alpha - \beta\|^2$;

(g) $\sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{\beta \in \mathcal{R}_\rho} ((E_\rho(y_t \Delta y_{t-j} - z_t \Delta z_{t-j})^2)/(\mathrm{Var}(\Delta y_t))^2) \leq C(\delta)\|\alpha - \beta\|^2$.

Here we have $\|\alpha - \beta\| = \max_i |\alpha_i - \beta_i|$. *The constant* $C(\delta)$ *depends only on* δ *and the order of the process* p.

PROOF: (a) First of all, we can note that any complex root has a complex conjugate. Since we restrict ourselves to at most one root close to the unit circle, then if there is such a root it must be real.

Let us introduce polynomials $a(L) = 1 - \rho L - \sum_{j=1}^{p-1} \alpha_j (1 - L)L^{j-1}$ and $b(L) = 1 - \rho L - \sum_{j=1}^{p-1} \beta_j (1 - L)L^{j-1} = (1 - \mu_1 L) \cdots (1 - \mu_p L)$. Then we have

$a(L)y_t = \varepsilon_t$ and $b(L)z_t = \varepsilon_t$. We can note that

$$y_t - z_t = \left(1 - \frac{a(L)}{b(L)}\right)y_t = \frac{L(1-L)f(L)}{(1-\mu_1 L)(1-\mu_2 L)\cdots(1-\mu_p L)}y_t$$

$$= \frac{(1-L)f(L)}{(1-\mu_1 L)(1-\mu_2 L)\cdots(1-\mu_p L)}y_{t-1},$$

where $f(L) = f_0 + f_1 L + \cdots + f_{p-2}L^{p-2}$, $f_j = (\alpha_{j+1} - \beta_{j+1})$, and $\max_j |f_j| = \|\alpha - \beta\|$. We also assume that $|\mu_1| \leq |\mu_2| \leq \cdots \leq |\mu_{p-1}| < \delta < 1$.

Let

$$\frac{(1-L)f(L)}{(1-\mu_1 L)(1-\mu_2 L)\cdots(1-\mu_p L)} = \sum_{j=0}^{\infty} d_j L^j.$$

Then $y_t - z_t = \sum_{j=0}^{\infty} d_j y_{t-j-1}$. It is easy to show that

$$\mathrm{Var}(y_t - z_t) \leq \left(\sum_{i=1}^{\infty}|d_i|\right)^2 \mathrm{Var}(y_t).$$

We can note that $\sum_{j=0}^{\infty}|d_j| \leq (p-2)\|\alpha - \beta\|\sum_{j=0}^{\infty}|c_j|$, where c_j are defined as in Lemma S8. The statement of part (a) follows from the statement (a) of Lemma S8.

(b) The proof is absolutely similar to that of (a) and follows from the fact that $\Delta y_t - \Delta z_t = (1 - \frac{a(L)}{b(L)})\Delta y_t$.

(c) We have $\frac{a(L)}{b(L)}y_t = z_t$. Using the same reasoning as before, we get

$$\mathrm{Var}(z_t) \leq \left(\sum_i |f_i|\right)^2 \mathrm{Var}(y_t), \quad \text{where} \quad \frac{a(L)}{b(L)} = \sum_i f_i L^i.$$

It is easy to see that $f_i = d_i$ for $i \geq 1$ and $f_0 = d_0 + 1$, where d_i are defined in the proof of part (a) of Lemma S9. Then

$$\frac{\mathrm{Var}(z_t)}{\mathrm{Var}(y_t)} - 1 \leq \left(\sum_i |d_i| + 1\right)^2 - 1 \leq C(\delta)\|\alpha - \beta\|.$$

Similarly, $\frac{\mathrm{Var}(y_t)}{\mathrm{Var}(z_t)} - 1 \leq C(\delta)\|\alpha - \beta\|$, which gives us statement (c).

Proof of part (d) is analogous to that of part (c).

(e) It is easy to note that

$$E(y_t^2 - z_t^2)^2 \leq \sqrt{E(y_t - z_t)^4}\sqrt{E(y_t + z_t)^4}.$$

By reasoning similar to that in the proof of part (a), $E(y_t - z_t)^4 \leq (\sum_{i=0}^{\infty} |d_i|)^4 \times Ey_t^4$, where d_i are defined in the proof of part (a). We also have $E(y_t + z_t)^4 \leq (\sum_{i=0}^{\infty} |g_i|)^4 Ey_t^4$, where $\frac{a(L)+b(L)}{b(L)} = \sum_i g_i L^i$. It is easy to see that $g_i = d_i$ for $i \geq 1$ and $g_0 = d_0 + 2$. As a result,

$$E(y_t^2 - z_t^2)^2 \leq \left(\sum_{i=0}^{\infty} |d_i| \right)^2 \left(\sum_{i=0}^{\infty} |g_i| \right)^2 Ey_t^4 \leq \mathrm{const}(\delta) \|\alpha - \beta\|^2 Ey_t^4.$$

The only thing left to check is that the expression $Ey_t^4/(Ey_t^2)^2$ is bounded.

Let $y_t = \sum_i h_i \varepsilon_{t-i}$. Then $Ey_t^4 = \sum h_i^4 E\varepsilon_t^4 + (\sum_i h_i^2)^2 (E\varepsilon_t^2)^2 \leq (\sum_i h_i^2)^2 (E\varepsilon_t^4 + (E\varepsilon_t^2)^2) = (Ey_t^2)^2 (1 + \frac{E\varepsilon_t^4}{(E\varepsilon_t^2)^2})$. That finishes the proof of part (e).

The proofs of parts (f) and (g) are similar to the proof of part (e).   *Q.E.D.*

S5.2. *Estimation of the Nuisance Parameters*

Let us have a sample $Y = (y_1, \ldots, y_T)$ from the process (S3) with at most one root close to the unit circle. We should note that the parameter $\alpha = (\alpha_1, \ldots, \alpha_{p-1})$ is a nuisance parameter for the hypothesis $H_0 : \rho = \rho_0$. As a result, it is impossible to construct an exact confidence interval for the parameter ρ even if we deal with an AR(p) model with normal errors.

As part of a procedure to test that the sum of AR coefficient is equal to ρ, we calculate an estimate $\widehat{\alpha}(\rho)$ of the nuisance parameter α as the OLS estimate in a regression model with the null hypothesis imposed:

$$(S4) \qquad y_t - \rho y_{t-1} = \sum_{j=1}^{p-1} \alpha_j \Delta y_{t-j} + \varepsilon_t,$$

that is, we regress $y_t - \rho y_{t-1}$ on $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$.

LEMMA S10: *Assume that we have an* AR(p) *process defined by Equation* (S3) *with error terms satisfying the set of Assumptions* C. *Let us define* $Y_t(\rho) = y_t - \rho y_{t-1}$ *and* $X_t = (\Delta y_{t-1}, \ldots, \Delta y_{t-p+1})$. *Let* $\widehat{\alpha}(\rho)$ *be the OLS estimate of* α *in the regression of* $Y_t(\rho)$ *on* X_t. *Then* $\widehat{\alpha}(\rho)$ *is a uniformly consistent estimate of* α, *that is, the following convergence holds*:

$$(S5) \qquad \lim_{T \to \infty} \sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} P_\rho\{\|\alpha - \widehat{\alpha}\| > \epsilon\} = 0 \quad \textit{for every } \epsilon > 0.$$

PROOF: Let $X = (X_1', \ldots, X_T')'$ and $\Sigma_T = (\varepsilon_1, \ldots, \varepsilon_T)'$. Then

$$\widehat{\alpha} - \alpha = \left(\frac{1}{\sum_{t=1}^{T} \mathrm{Var}(\Delta y_t)} X'X \right)^{-1} \frac{1}{\sum_{t=1}^{T} \mathrm{Var}(\Delta y_t)} X'\Sigma_T.$$

We prove two statements: $1/(\sum_{t=1}^{T} \text{Var}(\Delta y_t)) X' \Sigma_T \to^p 0$ uniformly and $1/(\sum_{t=1}^{T} \text{Var}(\Delta y_t)) X'X \to^p A$ uniformly, where $\det(A)$ is uniformly bounded away from 0.

The first statement can be obtained by noting that $E[(1/\sum_{t=1}^{T} \text{Var}(\Delta y_t)) \sum_{t=1}^{T} \Delta y_{t-1} \varepsilon_t]^2 = 1/(\sum_{t=1}^{T} \text{Var}(\Delta y_t)) \to 0$ uniformly, since $\text{Var}(\Delta y_t) \geq \text{Var}(\varepsilon_t)$.

For the proof of the second statement, we note that $\Delta y_t = \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$, where the coefficients c_j are defined in Lemma S8. Then $E[\Delta y_t \Delta y_{t-j}] = \gamma_j = \sum_{t=0}^{\infty} c_t c_{t+j}$.

Let us consider the covariance of the form

$$\text{cov}_{t,s,j} = E(\Delta y_t \Delta y_{t-j} - E\Delta y_t \Delta y_{t-j})(\Delta y_s \Delta y_{s-j} - E\Delta y_s \Delta y_{s-j}).$$

It is easy to see that

$$E(\Delta y_t \Delta y_{t-j} \Delta y_s \Delta y_{s-j}) = \gamma_j^2 + \gamma_{|t-s|}^2 + \gamma_{|t-j-s|}\gamma_{|t-s+j|} + E\varepsilon^4 \Gamma_{j,t-s,t-s+j},$$

where γ_j and $\Gamma_{i,j,k}$ are defined in Lemma S8. Then

$$\text{cov}_{t,s,j} = \gamma_{|t-s|}^2 + \gamma_{|t-j-s|}\gamma_{|t-s+j|} + E\varepsilon^4 \sum_{i=0}^{\infty} c_{i-t} c_{i-t+j} c_{i-s} c_{i-s+j}.$$

After applying Lemma S8, it is easy to show that $\sum_{s=1}^{T}\sum_{t=1}^{T} |\text{cov}_{t,s,j}| \leq C(\delta)T$. As a final step we can note that

$$E\left[\frac{1}{\sum_{t=1}^{T} \text{Var}(\Delta y_t)} \sum_{t=1}^{T}(\Delta y_t \Delta y_{t-j} - E\Delta y_t \Delta y_{t-j})\right]^2 \leq \frac{\sum_{s=1}^{T}\sum_{t=1}^{T} |\text{cov}_{t,s,j}|}{(\sum_{t=1}^{T} \text{Var}(\Delta y_t))^2}$$

$$\leq \frac{\text{const}(\delta)}{T}.$$

This ends the proof of Lemma S10. *Q.E.D.*

S5.3. *Grid Bootstrap*

To perform a test that the sum of AR coefficient is equal to ρ, we calculate the conventional t-statistic $t(\rho, y_1, \ldots, y_T)$ for this hypothesis in the regression model (S3). We also calculate estimates $\widehat{\alpha}(\rho)$ of the nuisance parameters α as in Lemma S10. Then we compare the calculated conventional t-statistic $t(\rho, Y)$ with a critical value function $q(\rho, T, \widehat{\alpha}(\rho))$, which depends on the tested value ρ of the parameter of interest on the estimated nuisance parameter and on the sample size.

The confidence set for the parameter ρ is constructed as a set of values for which the corresponding hypothesis is accepted:

$$(\text{S6}) \qquad C(Y) = \big\{\rho : q_1(\rho, T, \widehat{\alpha}(\rho)) \leq t(\rho, Y) \leq q_2(\rho, T, \widehat{\alpha}(\rho))\big\}.$$

We consider two sets of critical value functions: the one obtained by parametric grid bootstrap, which is a generalization of Andrews' (1993) method, and that obtained by Hansen's (1999) nonparametric grid bootstrap. In the parametric grid bootstrap, critical value functions are quantiles of the distribution of the t-statistic $t(\rho, Z)$ in the model

$$(S7) \qquad z_t = \rho z_{t-1} + \sum_{j=1}^{p-1} \widehat{\alpha}_j(\rho) \Delta z_{t-j} + e_t,$$

where error terms e_t are independently normally distributed with zero mean and unit variance. In the nonparametric grid bootstrap, we simulate critical value functions as quantiles of the distribution of the t-statistic in the model (S7) with independent and identically distributed error terms distributed according to the empirical distribution of the demeaned residuals from regression (S4).

Below we prove the validity of both procedures. The proofs are based on the uniform approximation of the unknown distribution of the t-statistic $t(\rho, Y)$ provided by the distributions calculated via parametric and nonparametric grid bootstraps.

To formulate the results, let us introduce some notations. Let statistics S and R be defined by

$$S(Y, \rho, \alpha, T) = G(\rho, \alpha)^{-1/2} \widetilde{Y}' \varepsilon,$$
$$R(Y, \rho, \alpha, T) = G(\rho, \alpha)^{-1/2} \widetilde{Y}' \widetilde{Y} G(\rho, \alpha)^{-1/2},$$

where $\widetilde{Y}_t = (y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1})$, $\widetilde{Y} = (\widetilde{Y}_1', \ldots, \widetilde{Y}_T')'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$, and $G(\rho, \alpha) = \mathrm{diag}(\sum_{t=1}^{T} \mathrm{Var}(y_t), \sum_{t=1}^{T} \mathrm{Var}(\Delta y_t), \ldots, \sum_{t=1}^{T} \mathrm{Var}(\Delta y_t))$. Then the t-statistic for testing the hypothesis that the sum of AR coefficients equals ρ is

$$t(Y, \rho, \alpha, T) = l_1' R^{-1}(Y, \rho, \alpha, T) S(Y, \rho, \alpha, T) / \sqrt{l_1' R^{-1}(Y, \rho, \alpha, T) l_1},$$

where $l_1 = (1, 0, \ldots, 0)$.

S5.4. *Parametric Grid Bootstrap*

S5.4.1. *Parametric grid bootstrap for* AR(p) *processes with normal errors*

In the case of AR(1) processes with normal error terms, the parametric grid bootstrap (Andrews' method) provides an exact confidence interval for the autoregressive coefficient ρ. As mentioned before, the generalization of the method to AR(p) models is not exact, even if the error terms are normally distributed, because the approximating distribution employs an estimate of the nuisance parameter rather than the true value of the nuisance parameter. We

prove that the procedure provides a uniform approximation of the unknown distribution of the t-statistic in a model with normal errors as long as the estimate of the nuisance parameter is uniformly consistent.

THEOREM S1: *Let us have two* AR(p) *processes* $y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \alpha \Delta y_{t-j} + \varepsilon_t$ *and* $z_t = \rho z_{t-1} + \sum_{j=1}^{p-1} \widehat{\alpha}_j \Delta z_{t-j} + \varepsilon_t$, *where error terms* ε_j *are independent standard normal random variables. Assume that the parameter* $\widehat{\alpha}$ *uniformly converges to* α *as the sample size increases, that is, convergence* (S5) *holds. Then we have the following uniform approximations*:

(a) $\lim_{T \to \infty} \sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} P_\rho \{|S(Y, \rho, \alpha, T) - S(Z, \rho, \widehat{\alpha}, T)| > \epsilon\} = 0$;
(b) $\lim_{T \to \infty} \sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} P_\rho \{|R(Y, \rho, \alpha, T) - R(Z, \rho, \widehat{\alpha}, T)| > \epsilon\} = 0$;
(c) $\lim_{T \to \infty} \sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} P_\rho \{|t(Y, \rho, \alpha, T) - t(Z, \rho, \widehat{\alpha}, T)| > \epsilon\} = 0$.

PROOF: By combining (a), (b), (c), and (d) of Lemma S9 with Chebyshev's inequality we have

$$P_\rho \big\{ |S(y, \rho, \alpha, T) - S(z, \rho, \widehat{\alpha}, T)| > \epsilon \big\} \le \frac{C(\delta)}{\epsilon^2} \|\alpha - \widehat{\alpha}\|.$$

Parts (c), (d), (e), (f), and (g) of Lemma S9 combined with Chebyshev's inequality give

$$P_\rho \big\{ |R(y, \rho, \alpha, T) - R(z, \rho, \widehat{\alpha}, T)| > \epsilon \big\} \le \frac{C(\delta)}{\epsilon^2} \|\alpha - \widehat{\alpha}\|.$$

Using the uniform consistency of the nuisance parameter estimate (S5), we get statements (a) and (b) of Theorem S1.

Statement (c) follows from parts (a) and (b), continuity of the ratio function, and the fact that statistic R is uniformly separated from 0. *Q.E.D.*

S5.4.2. *Parametric grid bootstrap. Approximation in the near unity region*

To prove that the parametric grid bootstrap is an asymptotically valid procedure for constructing confidence sets in models with nonnormal errors, we employ the same idea as in Section 2 of the paper. We divide the set of values of ρ into two subsets. One of the two subsets is increasing, while the second subset is contracting toward the unit root with a speed slower than $1/T$ as the sample size T increases. Over the first subset, the standard normal distribution provides the uniform approximation of the unknown distribution of the t-statistic. We obtain an approximation over the second set via constructing an AR(p) process with the same AR coefficients and normal errors, such that the t-statistic for this process is uniformly close to the initial t-statistic. This allows us to state that the distribution of the t-statistic for an AR(p) process is uniformly approximated by the distribution of the t-statistic for an AR(p)

process with the same AR coefficients, but with normal errors. Given that the parametric grid bootstrap works for models with normal errors, we obtain the validity of the procedure for models with nonnormal error terms.

LEMMA S11: *Assume that* $Y = (y_1, \ldots, y_T)$ *is a sample from an* $\mathrm{AR}(p)$ *process described by Equation* (S3) *with error terms satisfying the set of Assumptions* C. *Let* z_t *be an* $\mathrm{AR}(p)$ *process with normal errors*:

$$z_t = \rho z_{t-1} + \sum_{j=1}^{k} \alpha_j \Delta z_{t-j} + e_t, \quad e_t \sim i.i.d. \; N(0,1), \quad t = 1, \ldots, T.$$

Then there exists a completion of the initial probability space and the realization of process z_t *on this probability space such that, for every* $\delta > 0$, *we have*:

(a) $\sup_{\rho \in (0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{j=1,\ldots,T} |y_j/\sqrt{T} - z_j/\sqrt{T}| = o(T^{-1/4+\delta})$ *a.s.*,

(b) $\sup_{\rho \in (0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{j=1,\ldots,T} |y_{T,j}/\sqrt{T}| = O(1)$ *a.s.*,

(c) $\sup_{\rho \in (0,1)} \sup_{\alpha \in \mathcal{R}_\rho} |(1/T)\sum_{j=1}^{T} y_{j-1}\varepsilon_j - (1/T)\sum_{j=1}^{T} z_{j-1}e_j| = o(T^{-1/4+\delta})$ *a.s.*,

(d) $\sup_{\rho \in (0,1)} \sup_{\alpha \in \mathcal{R}_\rho} |(1/T^2)\sum_{j=1}^{T} y_{j-1}^2 - (1/T^2)\sum_{j=1}^{T} z_{j-1}^2| = o(T^{-1/4+\delta})$ *a.s.*

The statistic $S(Y, \rho, \alpha, T)$ *is a* p *dimensional vector. Let* $S_1(Y, \rho, \alpha, T) = l_1'S(Y, \rho, \alpha, T)$ *be the first component, while* $S_{(2\cdots p)}(Y, \rho, \alpha, T)$ *is the* $(p-1)$ *dimensional vector consisting of the last* $p - 1$ *components of the vector* $S(Y, \rho, \alpha, T)$. *Then*:

(e) $\sup_{\rho \in \mathcal{A}_T} \sup_{\alpha \in \mathcal{R}_\rho} \sup_x |S_1(Y, \rho, \alpha, T) - S_1(Z, \rho, \alpha, T)| = o(1)$ *a.s.*,

(f) $\sup_{\rho \in \mathcal{A}_T} \sup_{\alpha \in \mathcal{R}_\rho} \sup_x |P\{S_{(2\cdots p)}(Y, \rho, \alpha, T) > x\} - P\{\xi > x\}| = o(1)$, *where* $\xi \sim N(0, \Gamma)$.

The statistic $R(Y, \rho, \alpha, T)$ *is a* $p \times p$ *dimensional matrix. Let* $R_{11}(Y, \rho, \alpha, T) = l_1'R(Y, \rho, \alpha, T)l_1$ *be the left upper corner element of* $R(Y, \rho, \alpha, T)$. *Let* $R_{1,(2\cdots p)}(Y, \rho, \alpha, T)$ *be the* $(p-1)$ *dimensional vector consisting of the elements of the first column of the matrix, excluding the first element. We denote by* $R_{(2\cdots p),(2\cdots p)}(Y, \rho, \alpha, T)$ *the right lower square* $(p-1) \times (p-1)$ *submatrix of* $R(Y, \rho, \alpha, T)$. *That is*,

$$R(Y, \rho, \alpha, T) = \begin{pmatrix} R_{11} & R_{1,(2\cdots p)}' \\ R_{1,(2\cdots p)} & R_{(2\cdots p),(2\cdots p)} \end{pmatrix}.$$

Then:

(g) $\sup_{\rho \in \mathcal{A}_T} \sup_{\alpha \in \mathcal{R}_\rho} \sup_x |R_{11}(Y, \rho, \alpha, T) - R_{11}(Z, \rho, \alpha, T)| = o(1)$ *a.s.*,

(h) $\sup_{\rho \in \mathcal{A}_T} \sup_{\alpha \in \mathcal{R}_\rho} P\{|R_{1,(2\cdots p)}(Y, \rho, \alpha, T)| > x\} = o(1)$ *for any* $x > 0$,

(i) $\sup_{\rho \in \mathcal{A}_T} \sup_{\alpha \in \mathcal{R}_\rho} P\{|R_{(2\cdots p),(2\cdots p)}(Y, \rho, \alpha, T) - \Gamma| > x\} = o(1)$ *for any* $x > 0$.

Finally:

(k) $\lim_{T\to\infty} \sup_{\rho\in\mathcal{A}_T} \sup_{\alpha\in\mathcal{R}_\rho} \sup_x |P\{t(Y, \rho, \alpha, T) > x\} - P\{t(Z, \rho, \alpha, T) > x\}| = 0$.

Here a set $\mathcal{A}_T$ of parameters ρ is defined by $\mathcal{A}_T = \{\rho \in (0, 1) : |1 - \rho| < T^{-1+\varepsilon}\}$ for a sufficiently small $\varepsilon > 0$. All limits are taken as T increases to infinity.

PROOF: In the proof, the word "uniformly" always mean "uniformly over $\rho \in \mathcal{A}_T$ and $\alpha \in \mathcal{R}_\rho$."

(a) We can find a probability space with a realization of the partial sum process $\eta_T(t)$ and a sequence of Brownian processes $w_T(t)$ on it such that $\sup_{0\le t\le 1} |\eta_T(t) - w_T(t)| = O(T^{-1/4+\delta})$ almost surely. As before, we define the realization of error terms to be the normalized increments of the corresponding processes:

$$\frac{\varepsilon_j}{\sqrt{T}} = \eta_T\left(\frac{j}{T}\right) - \eta_T\left(\frac{j-1}{T}\right), \quad \frac{e_j}{\sqrt{T}} = w_T\left(\frac{j}{T}\right) - w_T\left(\frac{j-1}{T}\right).$$

Let us define a sequence of numbers d_j by the equality $\frac{1}{a(L)} = \sum_{j=0}^{\infty} d_j L^j$. Then the sequence $c_j = d_j - d_{j-1}$ is the same as in Lemma S8. We have

$$y_t = \sum_{j=0}^{t} d_j \varepsilon_{t-j} = \sum_{j=0}^{t} (d_j - d_{j-1}) \eta_T\left(\frac{t-j}{T}\right) = \sum_{j=0}^{t} c_j \eta_T\left(\frac{t-j}{T}\right)$$

and

$$z_t = \sum_{j=0}^{t} d_j e_{t-j} = \sum_{j=0}^{t} (d_j - d_{j-1}) w_T\left(\frac{t-j}{T}\right) = \sum_{j=0}^{t} c_j w_T\left(\frac{t-j}{T}\right).$$

Then by using statement (a) from Lemma S8, we get

$$\sup_{\rho\in(0,1)} \sup_{\alpha\in\mathcal{R}_\rho} \sup_{j=1,\ldots,T} \left| \frac{y_{T,j}}{\sqrt{T}} - \frac{z_{T,j}}{\sqrt{T}} \right|$$

$$\le \left(\sup_{\rho\in(0,1)} \sup_{\alpha\in\mathcal{R}_\rho} \sum_{j=0}^{\infty} |c_j| \right) \sup_{0\le t\le 1} |\eta_T(t) - w_T(t)| = O(T^{-1/4+\delta}) \quad \text{a.s.}$$

This ends the proof of the part (a).

(b) We have

$$\sup_{\rho\in(0,1)} \sup_{\alpha\in\mathcal{R}_\rho} \sup_{j=1,\ldots,T} \left| \frac{y_{T,j}}{\sqrt{T}} \right| \le \left(\sup_{\rho\in(0,1)} \sup_{\alpha\in\mathcal{R}_\rho} \sum_{j=0}^{\infty} |c_j| \right) \sup_{0\le t\le 1} |w_T(t)|$$

$$= O(1) \quad \text{a.s.}$$

(c) We have

$$\frac{1}{T}\sum_{j=1}^{T}y_{j-1}\varepsilon_j - \frac{1}{T}\sum_{j=1}^{T}z_{j-1}e_j$$

$$= \sum_{j=1}^{T}\frac{\Delta y_{j-1}}{\sqrt{T}}\eta_T\left(\frac{j}{T}\right) - \sum_{j=1}^{T}\frac{\Delta z_{j-1}}{\sqrt{T}}w_T\left(\frac{j}{T}\right)$$

$$- \left(\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\eta_T\left(\frac{j}{T}\right)\varepsilon_{T,j} - \frac{1}{\sqrt{T}}\sum_{j=1}^{T}w_T\left(\frac{j}{T}\right)e_{T,j}\right)$$

$$+ \left(\frac{y_{T,T}(\rho)}{\sqrt{T}}\eta_T(1) - \frac{z_{T,T}(\rho)}{\sqrt{T}}w_T(1)\right)$$

Let us consider the first term:

$$\sup_{\rho\in(0,1)}\sup_{\alpha\in\mathcal{R}_\rho}\left|\sum_{j=1}^{T}\frac{\Delta y_{j-1}}{\sqrt{T}}\eta_T\left(\frac{j}{T}\right) - \sum_{j=1}^{T}\frac{\Delta z_{j-1}}{\sqrt{T}}w_T\left(\frac{j}{T}\right)\right|$$

$$= \sup_{\rho\in(0,1)}\sup_{\alpha\in\mathcal{R}_\rho}\left|\sum_{j=1}^{T}\frac{\sum_i c_i\varepsilon_{j-i}}{\sqrt{T}}\eta_T\left(\frac{j}{T}\right) - \sum_{j=1}^{T}\frac{\sum_i c_i e_{j-i}}{\sqrt{T}}w_T\left(\frac{j}{T}\right)\right|$$

$$\leq \left(\sup_{\rho\in(0,1)}\sup_{\alpha\in\mathcal{R}_\rho}\sum_i|c_i|\right)\max_i\left|\sum_{j=1}^{T}\frac{\varepsilon_{j-i}}{\sqrt{T}}\eta_T\left(\frac{j}{T}\right) - \sum_{j=1}^{T}\frac{e_{j-i}}{\sqrt{T}}w_T\left(\frac{j}{T}\right)\right|$$

$$= o(T^{-1/4+\delta}) \quad \text{a.s.}$$

According to part (c) of Lemma 3, the following asymptotic equality has place:

$$\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\eta_T\left(\frac{j}{T}\right)\varepsilon_{T,j} - \frac{1}{\sqrt{T}}\sum_{j=1}^{T}w_T\left(\frac{j}{T}\right)e_{T,j}\right| = o(T^{-1/4+\delta}) \quad \text{a.s.}$$

From the parts (a) and (b) of Lemma S11, it is easy to determine that

$$\sup_{\rho\in(0,1)}\sup_{\alpha\in\mathcal{R}_\rho}\left|\frac{y_{T,T}(\rho)}{\sqrt{T}}\eta_T(1) - \frac{z_{T,T}(\rho)}{\sqrt{T}}w_T(1)\right| = o(T^{-1/4+\delta}) \quad \text{a.s.}$$

The last three limits give us statement (c).

(d) The statement easily follows from parts (a) and (b):

$$\sup_{\rho\in(0,1)}\sup_{\alpha\in\mathcal{R}_\rho}\left|\frac{1}{T^2}\sum_{j=1}^{T}y_j^2 - \frac{1}{T^2}\sum_{j=1}^{T}z_j^2\right|$$

$$\leq \sup_{\rho \in (0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_j \left| \frac{y_j}{\sqrt{T}} - \frac{z_j}{\sqrt{T}} \right| \left(\sup_j \left| \frac{y_j}{\sqrt{T}} \right| + \sup_j \left| \frac{z_j}{\sqrt{T}} \right| \right)$$

$$= o(T^{-1/4+\delta}) \quad \text{a.s.}$$

(e) The statistic $S(Y, \rho, \alpha, T)$ is a p dimensional vector, the first component of which, $l_1' S(Y, \rho, \alpha, T) = \sum_{t=1}^T y_{t-1} \varepsilon_t / \sqrt{\sum_{t=1}^T \mathrm{Var}(y_t)}$ may have nonstandard behavior.

We note that $\sum_{t=1}^T \mathrm{Var}(y_t) = T \sum_{j=0}^\infty d_j^2$, where $y_t = \sum_{j=1}^\infty d_j \varepsilon_{t-j}$. It is easy to see that $\sum_{j=0}^\infty d_j^2 \geq C(\delta)\frac{1}{1-\rho}$. Then

$$\frac{1}{\sqrt{\sum_{t=1}^T \mathrm{Var}(y_t)}} \left(\sum_{t=1}^T y_{t-1}\varepsilon_t - \sum_{t=1}^T z_{t-1}e_t \right)$$

$$\leq C(\delta)\sqrt{T(1-\rho)} \left(\frac{1}{T} \sum_{t=1}^T y_{t-1}\varepsilon_t - \frac{1}{T} \sum_{t=1}^T z_{t-1}e_t \right)$$

$$= \sqrt{T(1-\rho)} o(T^{-1/4+\delta}) = o(1) \quad \text{a.s.}$$

uniformly over the set $\mathcal{A}_T$. This gives us that there exist realizations of processes y_t and z_t on the same probability space such that $l_1' S(Y, \rho, \alpha, T)$ and $l_1' S(Z, \rho, \alpha, T)$ are almost surely uniformly close to each other over the set $\mathcal{A}_T$.

(f) We note that

$$\Delta y_t = -(1 - \mu_p) y_{t-1} + Y_t,$$

where $(1 - \mu_1 L) \cdots (1 - \mu_{p-1} L) Y_t = \varepsilon_t$ is a stationary process with all roots strictly outside the $1/\delta$ circle. It is easy to see that

$$E\left(\frac{(1 - \mu_p) \sum_t^T y_{t-j}\varepsilon_t}{\sqrt{\sum_t^T \mathrm{Var}(y_t)}} \right)^2 = (1 - \mu_p)^2 \to 0 \quad \text{uniformly over } \mathcal{A}_T.$$

As a result, we have that the sequence $((1 - \mu_p) \sum_t^T y_{t-j}\varepsilon_t)/\sqrt{\sum_t^T \mathrm{Var}(y_t)}$ converges in probability to zero uniformly over $\mathcal{A}_T$.

We need to check that

$$\Gamma^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^T (Y_{t-1}\varepsilon_t, \ldots, Y_{t-p+1}\varepsilon_t)' \Rightarrow N(0, I_{p-1}),$$

where $\Gamma = E[(Y_{t-1}, \ldots, Y_{t-p+1})(Y_{t-1}, \ldots, Y_{t-p+1})']$ and the convergence is taking place uniformly over all possible processes with roots outside the $1/\delta$ circle.

The statement follows from the central limit theorem for martingale differences.

(g) From the definition of the statistics, we have $R_{11}(Y, \rho, \alpha, T) = 1/(\sum_{j=1}^{T} \text{Var}(y_j)) \sum_{j=1}^{T} y_{j-1}^2$ and $R_{11}(Z, \rho, \alpha, T) = 1/(\sum_{j=1}^{T} \text{Var}(z_j)) \sum_{j=1}^{T} z_{j-1}^2$.

From statement (d) of Lemma S11, we obtain, uniformly over $\mathcal{A}_T$,

$$|R_{11}(y, \rho, \alpha, T) - R_{11}(z, \rho, \alpha, T)|$$

$$\leq T(1-\rho) \left| \frac{1}{T^2} \sum_{j=1}^{T} y_{j-1}^2 - \frac{1}{T^2} \sum_{j=1}^{T} z_{j-1}^2 \right|$$

$$= T(1-\rho) o(T^{-1/4+\varepsilon}) = o(T^{-1/4+2\varepsilon}) = o(1) \quad \text{a.s.}$$

The last inequality holds if $\varepsilon < 1/4$.

(h) Since $\Delta y_t = -(1 - \mu_p) y_t + Y_t$, where the process Y_t is defined by the equation $(1 - \mu_1 L) \cdots (1 - \mu_{p-1} L) Y_t = \varepsilon_t$, then

$$\sum_{j=1}^{T} y_j \Delta y_{j-1} = (1 - \mu_p) \sum_{j=1}^{T} y_j^2 + \sum_{j=1}^{T} y_j Y_j.$$

We have

$$\frac{1}{T^{3/2}} (1 - \mu_p) \sum_{j=1}^{T} y_j^2$$

$$= (1 - \mu_p) \sqrt{T} \frac{1}{T^2} \sum_{j=1}^{T} y_j^2 \leq (1 - \mu_p) \sqrt{T} \left(\max_{j=1,\ldots,T} \left| \frac{y_j}{\sqrt{T}} \right| \right)^2$$

$$= O((1 - \mu_p) \sqrt{T}) = O(T^{-1/2+\varepsilon}) \quad \text{a.s. uniformly over } \mathcal{A}_T.$$

We note that $\sum_{t=1}^{T} \text{Var}(y_t) \geq CT \frac{1}{1-\rho}$; and $\sum_{t=1}^{T} \text{Var}(\Delta y_t) \geq T$. As a result,

$$\frac{T^{3/2}}{\sqrt{\sum_{t=1}^{T} \text{Var}(y_t)} \sqrt{\sum_{t=1}^{T} \text{Var}(\Delta y_t)}} \leq \sqrt{T(1-\rho)}$$

and

(S8) $$\frac{1}{\sqrt{\sum_{t=1}^{T} \text{Var}(y_t)} \sqrt{\sum_{t=1}^{T} \text{Var}(\Delta y_t)}} (1 - \mu_p) \sum_{j=1}^{T} y_j^2 = O(T^{-1/2+3/2\varepsilon}) \quad \text{a.s.}$$

Now let us turn to the second term. First of all, we note that $y_t = \sum_j d_j \varepsilon_{t-j}$ and $Y_t = \sum_j d_j^* \varepsilon_{t-j}$, where

$$
|d_j| = \left| \sum_{k_1 + \cdots + k_p = j} \mu_1^{k_1} \cdots \mu_p^{k_p} \right| \leq \sum_{l=0}^{j} l^k \delta^l |\mu_p|^{j-l}
$$

$$
\leq |\mu_p|^j \sum_{l=0}^{\infty} l^k \left(\frac{\delta}{|\mu_p|} \right)^l \leq C |\mu_p|^j.
$$

The constant C depends on δ but not on μ_p or other roots. The inequality holds if $|\mu_1| \leq \cdots \leq |\mu_{p-1}| < \delta$, and $\rho \in \mathcal{A}_T$ for sufficiently large T such that $|1 - \mu_p| < (1 - \rho)/(1 - \delta)^{k-1} < 1 - (\delta + \epsilon)$ for some fixed $\epsilon > 0$.

Similarly, $|d_j^*| = |\sum_{k_1 + \cdots + k_{p-1} = j} \mu_1^{k_1} \cdots \mu_{p-1}^{k_{p-1}}| \leq C \delta^j$.

We have that

$$
E(y_T Y_T y_{T-j} Y_{T-j})
$$

$$
= E \varepsilon_t^4 \sum_i d_i d_{i+j} d_i^* d_{i+j}^* + \left(\sum_i d_i d_i^* \right)^2
$$

$$
+ \left(\sum_i d_i d_{i+j}^* \right) \left(\sum_i d_i^* d_{i+j} \right) + \left(\sum_i d_i^* d_{i+j}^* \right) \left(\sum_i d_i d_{i+j} \right).
$$

By using inequalities for d_j and d_j^*, we can get $|E(y_T Y_T y_{T-j} Y_{T-j})| \leq C_1 + C_2((|\mu_p|^j \delta^j)/(1 - |\mu_p|))$. As a result,

$$
E \left(\sum_{t=1}^{T} y_t Y_t \right)^2 \leq T \sum_{j=1}^{T} |E(y_T Y_T y_{T-j} Y_{T-j})| \leq C_1 T^2 + C_3 \frac{T}{1 - |\mu_p|}.
$$

By using Chebyshev's lemma, we have, uniformly over $\mathcal{A}_T$,

$$
\sup_{\rho \in \mathcal{A}_T} \sup_{\alpha \in \mathcal{R}_\rho} P \left\{ \left| \frac{1}{\sqrt{\sum_{t=1}^{T} \mathrm{Var}(y_t)} \sqrt{\sum_{t=1}^{T} \mathrm{Var}(\Delta y_t)}} \sum_{j=1}^{T} y_j Y_j \right| > x \right\}
$$

$$
\leq \frac{1 - \rho}{T^2} \left(C_1 T^2 + C_3 \frac{T}{1 - |\mu_p|} \right) = O(T^{-1+\varepsilon}).
$$

By joining the last inequality with (S8), we obtain statement (h) of the lemma.

(i) We use the fact that $\Delta y_t = -(1 - \mu_p)y_{t-1} + Y_t$, where Y_t is defined in the proof of statement (f):

$$(S9) \qquad \frac{1}{T}\sum_{t=1}^{T}\Delta y_t \Delta y_{t-j} = \frac{(1-\mu_p)^2}{T}\sum_{t=1}^{T} y_t y_{t-j}$$

$$- \frac{(1-\mu_p)}{T}\sum_{t=1}^{T}(y_{t-1}Y_{t-j} + y_{t-j-1}Y_t) + \frac{1}{T}\sum_{t=1}^{T} Y_t Y_{t-j}.$$

As in the proof of part (h) we can show that $((1 - \mu_p)/T^{3/2})\sum_{t=1}^{T} y_t y_{t-j} = O(T^{-1/2+\varepsilon})$ almost surely uniformly over $\mathcal{A}_T$. This gives us $((1 - \mu_p)^2/T) \times \sum_{t=1}^{T} y_t y_{t-j} = O(T^{-1+2\varepsilon}) = o(1)$ almost surely uniformly over $\mathcal{A}_T$.

As in the proof of part (h), we can show that $1/(\sqrt{\mathrm{Var}(\Delta y_t)}\sqrt{\mathrm{Var}(y_t)}) \times \sum_{t=1}^{T}(y_{t-1}Y_{t-j} + y_{t-j-1}Y_t)$ converges in probability to zero uniformly over $\mathcal{A}_T$. Given that $(\sqrt{\mathrm{Var}(\Delta y_t)}\sqrt{\mathrm{Var}(y_t)})/T(1 - \mu_p) = o(1)$ uniformly, we conclude that the second term in (S9) uniformly converges to zero in probability.

The only thing left is to prove that $\frac{1}{T}\sum_{t=1}^{T} Y_t Y_{t-j}$ uniformly converges in probability to $E(Y_t Y_{t-j})$. For this statement, we show that $E(\frac{1}{T}\sum_{t=1}^{T} Y_t Y_{t-j} - E(Y_t Y_{t-j}))^2$ converges uniformly to zero and then we use Chebyshev's inequality.

We already showed that $Y_t = \sum_{j=0}^{\infty} d_j^* \varepsilon_{t-j}$ with $|d_j^*| \le C\delta^j$:

$$E(Y_0 Y_{0-j} Y_s Y_{s-j}) = E\varepsilon^4 \sum_i d_{i+j}^* d_i^* d_{s+j}^* d_{s+i}^* + \left(\sum_i d_i^* d_{i+j}^*\right)^2$$

$$+ \left(\sum_i d_i^* d_{i+s}^*\right)^2 + \left(\sum_i d_i^* d_{i+s+j}^*\right)\left(\sum_i d_i^* d_{i+s-j}^*\right),$$

$$|\mathrm{cov}(Y_0 Y_{0-j}, Y_s Y_{s-j})| \le E\varepsilon^4 \left|\sum_i d_{i+j}^* d_i^* d_{s+j}^* d_{s+i}^*\right|$$

$$+ \left(\sum_i d_i^* d_{i+s}^*\right)^2 + \left|\sum_i d_i^* d_{i+s+j}^*\right|\left|\sum_i d_i^* d_{i+s-j}^*\right|$$

$$\le C\delta^{2s}.$$

As a result,

$$E\left(\frac{1}{T}\sum_{t=1}^{T} Y_t Y_{t-j} - E(Y_t Y_{t-j})\right)^2$$

$$\le \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} |\mathrm{cov}(Y_t Y_{t-j}, Y_s Y_{s-j})| \le \frac{C}{T},$$

where C depends only on δ and p.

(k) Since matrix $R(Y, \rho, \alpha, T)$ is asymptotically uniformly close in probability to the matrix,

$$\widetilde{R}(Z, \rho, \alpha, T) = \begin{pmatrix} \frac{1}{\sum_{t=1}^{T} \mathrm{Var}(z_t)} \sum_{t=1}^{T} z_t^2 & \mathbf{0}' \\ \mathbf{0} & \Gamma \end{pmatrix},$$

where $\mathbf{0}$ is a zero $p \times 1$ vector. Then $R^{-1}(Y, \rho, \alpha, T)$ is asymptotically uniformly close in probability to the matrix

$$\widetilde{R}^{-1}(Z, \rho, \alpha, T) = \begin{pmatrix} \frac{1}{\sum_{t=1}^{T} z_t^2} \sum_{t=1}^{T} \mathrm{Var}(z_t) & \mathbf{0}' \\ \mathbf{0} & \Gamma^{-1} \end{pmatrix},$$

where we used that $\sum_{t=1}^{T} z_t^2 / (\sum_{t=1}^{T} \mathrm{Var}(z_t))$ is uniformly separated from zero in the sense of Lemma 13. Given the fact that $S_{1,(2\cdots p)}(Y, \rho, \alpha, T)$ is asymptotically uniformly normally distributed, we have that

$$t(Y, \rho, \alpha, T) = \frac{l_1' R^{-1}(Y, \rho, \alpha, T) S(Y, \rho, \alpha, T)}{\sqrt{l_1' R^{-1}(Y, \rho, \alpha, T) l_1}}$$

is uniformly close in probability to $(R_{11}^{-1}(Z, \rho, \alpha, T) S_{11}(Z, \rho, \alpha, T)) / (R_{11}^{-1}(Z, \rho, \alpha, T))^{1/2}$. The last expression is equal to $\sum_{t=1}^{T} z_{t-1} e_t / \sqrt{\sum_{t=1}^{T} z_t^2}$ and it is asymptotically uniformly close in probability to $t(Z, \rho, \alpha, T)$. This ends the proof of Lemma S11.   *Q.E.D.*

S5.4.3. *Parametric grid bootstrap. Approximation in the stationary region*

LEMMA S12: *Assume that $Y = (y_1, \ldots, y_T)$ is a sample from an* $\mathrm{AR}(p)$ *process defined by Equation* (S3) *with error terms satisfying the set of Assumptions* C. *Let us define a set* $\mathcal{B}_T = \{\rho \in (0, 1) : 1 - \rho > CT^{-1+\varepsilon}\}$ *for arbitrarily small* $\varepsilon > 0$. *Let* Y *be the correlation matrix for a random vector* $X_t = (y_{y-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1})$. *Then:*

(a) $\lim_{T \to \infty} \sup_{\rho \in \mathcal{B}_T} \sup_{\alpha \in \mathcal{R}_\rho} P\{|R(Y, \rho, \alpha, T) - Y| > \epsilon\} = 0$ *for every* $\epsilon > 0$;

(b) $\lim_{T \to \infty} \sup_{\rho \in \mathcal{B}_T} \sup_{\alpha \in \mathcal{R}_\rho} \sup_x |P\{a'S(Y, \rho, \alpha, T) < x\} - \Phi(x)| = 0$, *for any* p *dimensional vector* a *such that* $a'Ya = 1$;

(c) $\lim_{T \to \infty} \sup_{\rho \in \mathcal{A}_T} \sup_{\alpha \in \mathcal{R}_\rho} \sup_x |P\{t(Y, \rho, \alpha, T) < x\} - \Phi(x)| = 0$.

PROOF: The proof is totally analogous to that of Lemmas 2.1 and 2.2 from Giraitis and Phillips (2006). Since we assumed the existence of a finite fourth moment, we do not need to use the truncation argument.

(a) As before, we use that $y_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j}$ and $|d_j| \leq C|\mu_p|^j$, where C depends only on δ and p. We have

$$|\operatorname{cov}(y_0^2, y_s^2)| \leq E\varepsilon_t^4 \sum_{i=0}^{\infty} d_i^2 d_{i+s}^2 + \left(\sum_{i=0}^{\infty} d_i d_{i+s}\right)^2 \leq \frac{C}{(1-\mu_p)^2} |\mu_p|^{2s}.$$

As a result,

$$\frac{1}{T^2 \operatorname{Var}^2(y_t)} E\left(\sum_{t=1}^{T} y_t^2 - T\operatorname{Var}(y_t)\right)^2$$

$$\leq \frac{1}{T\operatorname{Var}^2(y_t)} \sum_{s}^{T} |\operatorname{cov}(y_0^2, y_s^2)| \leq C\frac{1}{T\operatorname{Var}^2(y_t)(1-\mu_p)^3}$$

$$\leq \frac{C}{T(1-\mu_p)} \to 0.$$

This gives us

(S10) $$\limsup_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} \sup_{\alpha\in\mathcal{R}_\rho} P\left\{\left|\frac{1}{T\operatorname{Var}(y_t)} \sum_{t=1}^{T} y_t^2 - 1\right| > \epsilon\right\} = 0 \quad \text{for every } \epsilon > 0.$$

Similarly, since $\Delta y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ and $|c_j| \leq C\delta^j$, we have

$$|\operatorname{cov}(y_t \Delta y_{t-j}, y_{t+s}\Delta y_{t+s-j})|$$

$$\leq E\varepsilon_t^4 \sum_{i=0}^{\infty} c_i d_{i+j} c_{i+s} d_{i+j+s} + \left(\sum_{i=0}^{\infty} c_i c_{i+s}\right)\left(\sum_{i=0}^{\infty} d_i d_{i+s}\right)$$

$$+ \left(\sum_{i=0}^{\infty} c_i d_{i+j+s}\right)\left(\sum_{i=0}^{\infty} c_{i+s} d_{i+j}\right)$$

$$\leq \frac{C}{(1-|\mu_p|)} |\mu_p|^s \delta^s,$$

$$\frac{1}{T^2 \operatorname{Var}(y_t)\operatorname{Var}(\Delta y_t)} E\left(\sum_{t=1}^{T} y_t\Delta y_{t-j} - T\operatorname{cov}(y_t, \Delta y_{t-j})\right)^2$$

$$\leq \frac{1}{T\operatorname{Var}(y_t)\operatorname{Var}(\Delta y_t)} \sum_{s}^{T} |\operatorname{cov}(y_t\Delta y_{t-j}, y_{t+s}\Delta y_{t+s-j})|$$

$$\leq C\frac{1}{T\operatorname{Var}(y_t)(1-|\mu_p|)} \leq \frac{C}{T} \to 0.$$

This gives us

$$(S11) \quad \lim_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} \sup_{\alpha\in\mathcal{R}_\rho} P\left\{ \left| \frac{1}{T\sqrt{\mathrm{Var}(y_t)}\sqrt{\mathrm{Var}(\Delta y_t)}} \sum_{t=1}^{T} y_t\Delta y_{t-j} \right. \right.$$

$$\left. \left. - \mathrm{corr}(y_t,\Delta y_{t-j}) \right| > \epsilon \right\} = 0.$$

In the proof of step (b) of Lemma S10 we showed that

$$(S12) \quad \lim_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} \sup_{\alpha\in\mathcal{R}_\rho} P\left\{ \left| \frac{1}{T} \sum_{t=1}^{T} \Delta y_t\Delta y_{t-j} - \mathrm{cov}(\Delta y_t,\Delta y_{t-j}) \right| > \epsilon \right\} = 0$$

for every $\epsilon > 0$.

Putting together equations (S10), (S11), and (S12) we obtain statement (a) of the lemma.

(b) Let a be a p dimensional vector such that $a'Ya = 1$. We consider the sequence of random variables

$$(S13) \quad \xi_{t,T} = \frac{1}{\sqrt{T}} a'\,\mathrm{diag}(\mathrm{Var}(y_t), \mathrm{Var}(\Delta y_{t-1}), \ldots, \mathrm{Var}(\Delta y_{t-p+1}))^{-1/2} X_t\varepsilon_t.$$

In order to prove that $\sum_{t=1}^{T} \xi_{t,T}$ converges to $N(0,1)$ as the sample size increases, we need to check three conditions:

(1) $E(\xi_{t,T}|\mathcal{F}_{t-1}) = 0$,

(2) $\sum_{t=1}^{T} E(\xi_{t,T}^2|\mathcal{F}_{t-1})$ converges uniformly in probability to 1,

(3) $\sum_{t=1}^{T} E(\xi_{t,T}^2 I_{|\xi_{t,T}|>\epsilon}|\mathcal{F}_{t-1})$ converges uniformly in probability to 0.

The first condition is trivially satisfied since $(\xi_t, \mathcal{F}(\{\varepsilon_i\}_{i=-\infty}^{t}))$ is a martingale difference sequence. For the second condition, we note that $\sum_{t=1}^{T} E(\xi_{t,T}^2| \mathcal{F}_{t-1}) = a'R(y,\rho,\alpha,T)a$ that, according the part (a) of Lemma S12, converges to 1.

We check the third condition:

$$E\left(\sum_{t=1}^{T} E(\xi_{t,T}^2 I_{|\xi_{t,T}|>\epsilon}|\mathcal{F}_{t-1}) \right)$$

$$\leq \epsilon^{-2} E\left(\sum_{t=1}^{T} E(\xi_{t,T}^4|\mathcal{F}_{t-1}) \right)$$

$$= \epsilon^{-2} \frac{1}{T^2} \sum_{t=1}^{T} E\left(a'\,\mathrm{diag}(\mathrm{Var}(y_t), \right.$$

$$\left. \text{Var}(\Delta y_{t-1}), \ldots, \text{Var}(\Delta y_{t-p+1}))^{-1/2} X_t \right)^4$$

$$= \frac{\epsilon^{-2}}{T} E\left(a_1 \frac{y_{t-1}}{\sqrt{\text{Var}(y_{t-1})}} + a_2 \frac{\Delta y_{t-1}}{\sqrt{\text{Var}(\Delta y_{t-1})}} \right.$$

$$\left. + \cdots + a_p \frac{\Delta y_{t-p+1}}{\sqrt{\text{Var}(\Delta y_{t-p+1})}} \right)^4.$$

It is enough to show that

$$E\left(\frac{y_{t-1}}{\sqrt{\text{Var}(y_{t-1})}} \right)^4 < C \quad \text{and} \quad E\left(\frac{\Delta y_{t-1}}{\sqrt{\text{Var}(\Delta y_{t-1})}} \right)^4 < C,$$

which can be shown easily.

(c) By applying parts (a) and (b) of the lemma with $a = Y^{-1} l_1 / \sqrt{l_1' Y^{-1} l_1}$ to the formula

$$t(y, \rho, \alpha, T) = \frac{l_1' R^{-1}(y, \rho, \alpha, T) S(y, \rho, \alpha, T)}{\sqrt{l_1' R^{-1}(y, \rho, \alpha, T) l_1}},$$

we get the statement (c).

Q.E.D.

S5.4.4. *Parametric grid bootstrap. Main theorem*

The validity of the parametric bootstrap procedure is stated in the theorem below.

THEOREM S2: *Assume that the process y_t is an* AR(p) *process defined by Equation* (S3) *with error terms satisfying the set of Assumptions* C. *Let z_t be an* AR(p) *process with normal errors defined by Equation* (S7), *where $\widehat{\alpha}(\rho)$ denotes the OLS estimates in a regression model* (S4). *Then the distribution of the t-statistic based on the process y_t could be uniformly approximated by the distribution of the t-statistic based on the process z_t:*

$$\lim_{T \to \infty} \sup_{\rho \in (0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_x \left| P\{t(Y, \rho, \alpha, T) > x\} - P\{t(Z, \rho, \widehat{\alpha}, T) > x\} \right| = 0.$$

As a result, the set defined by (S6) *with $q_i(\rho, T, \widehat{\alpha}(\rho))$, $i = 1, 2$, being quantiles of the distribution of $t(Z, \rho, \widehat{\alpha}, T)$, is an asymptotic confidence set for ρ.*

PROOF: Let a process ξ_t be defined as an $\text{AR}(p)$ process with the same coefficients as y_t with normal errors:

$$\xi_t = \rho \xi_{t-1} + \sum_{j=1}^{k} \alpha_j \Delta \xi_{t-j} + e_t, \quad e_t \sim \text{i.i.d. } N(0, 1), \quad t = 1, \dots, T.$$

It follows from Lemmas S11 and S12 that

$$\lim_{T \to \infty} \sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_x \left| P\{t(y, \rho, \alpha, T) < x\} - P\{t(\xi, \rho, \alpha, T) < x\} \right| = 0.$$

Theorem S1 states that

$$\lim_{T \to \infty} \sup_{\rho \in [0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_x \left| P\{t(\xi, \rho, \alpha, T) < x\} \right.$$

$$\left. - P\{t(z, \rho, \widehat{\alpha}(\rho), T) < x\} \right| = 0$$

as long as $\widehat{\alpha}(\rho)$ is a uniformly consistent estimator of α. The uniform consistency was obtained in Lemma S10. This ends the proof of the theorem.

Q.E.D.

S5.5. *Nonparametric Grid Bootstrap*

The nonparametric grid bootstrap procedure uses an approximation of the unknown distribution of the t-statistic $t(Y, \rho, \alpha, T)$ by the distribution of the t-statistic $t(Z, \rho, \widehat{\alpha}, T)$, where z_t is an $\text{AR}(p)$ process defined by (S7) with error terms having distribution F_T. We consider F_T to be an empirical distribution function of the residuals from the regression (S4). The distribution function $F_T(\Sigma, \rho_0, \rho, \alpha)$ depends on the realization of error terms of the process y_t, on the coefficients ρ and α of the process y_t, and on the null value ρ_0 tested.

The validity of Hansen's grid bootstrap is proven in the same way as we proved it for $\text{AR}(1)$, given the validity of Andrews' method.

THEOREM S3: *Assume that the process y_t is an $\text{AR}(p)$ process defined by Equation (S3) with error terms satisfying the set of Assumptions* C. *Let z_t be an $\text{AR}(p)$ process defined by Equation (S7), where $\widehat{\alpha}(\rho)$ denotes the OLS estimates in a regression model* (S4). *Assume that the errors e_t of the process z_t are independent and identically distributed with the distribution function F_T.*

(1) *We have*

$$\lim_{T \to \infty} \sup_{\rho \in (0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{F_T \in \mathcal{L}_4(K,M,\theta)} \sup_x \left| P\{t(Y, \rho, \alpha, T) > x\} \right.$$

$$\left. - P\{t(Z, \rho, \widehat{\alpha}, T) > x\} \right| = 0.$$

(2) *For almost all realizations of error terms* $\Sigma = \{\varepsilon_1, \ldots, \varepsilon_j, \ldots\}$, *let there exist constants* $K(\Sigma) > 0$, $M(\Sigma) > 0$, *and* $\delta > 0$ *such that, for all* $\rho \in \Theta_T$,

$$F_T(\Sigma, \rho, \rho, \alpha) \in \mathcal{L}_4(K, M, \theta).$$

Then

$$\lim_{T \to \infty} \sup_{\rho \in (0,1)} \sup_{\alpha \in \mathcal{R}_\rho} \sup_x \left| P_\rho\{t(Y, \rho, \alpha, T) > x\} \right.$$

$$\left. - P_\rho^*\{t(Z, \rho, \widehat{\alpha}, T) > x | \Sigma\} \right| = 0 \quad a.s.$$

That is, the bootstrap provides a uniform asymptotic approximation for almost all realizations of error terms.

Let $C(Y)$ *be a set defined by Equation* (S6) *with* $q_i(\rho, T, \widehat{\alpha}(\rho)) = q_i(\rho, T, \widehat{\alpha}(\rho)|Y)$, $i = 1, 2$, *being quantiles of the distribution of the statistic* $t(Z, \rho, \widehat{\alpha}, T)$, *given the realization of* Y. *Then the set* $C(Y)$ *is an asymptotic confidence set.*

(3) *Let* F_T^{err} *be an empirical distribution function of the residuals from the regression* (S4). *Then, for almost all realizations of error term* Σ, *there exist constants* $K(\Sigma) > 0$, $M(\Sigma) > 0$, *and* $\delta > 0$ *such that* $F_T^{\mathrm{err}} \in \mathcal{L}_4(K, M, \theta)$.

PROOF: According to Lemma 15, there exist realizations of a partial sum process and a sequence of Brownian motions such that

$$\lim_{T \to \infty} \sup_{F_T \in \mathcal{L}_r(K, M, \theta)} P\left\{ \sup_{0 \le t \le 1} |\eta_T(t) - w(t)| > \varepsilon T^{-\delta} \right\} = 0.$$

In the part (k) of Lemma S10, we proved that having such realizations of the processes leads to a uniform approximation in the near unity region.

In the proof of part (a) of Lemma S12, we showed that for any element ξ of the matrix $R(y, \rho, \alpha, T) - Y$ we have that $E(\xi)^2 \le \frac{C}{T(1-\rho)}$, where C is a constant that depends only on p, δ, M, and K. This implies that for every sequence of sets $\mathcal{B}_T = [-\rho_T, \rho_T]$ such that $T(1 - \rho_T) \to \infty$, we have

$$\lim_{T \to \infty} \sup_{\rho \in \mathcal{B}_T} \sup_{\alpha \in \mathcal{R}_\rho} \sup_{F_T \in \mathcal{L}_4(K, M, \theta)} P\{|R(y, \rho, \alpha, T) - Y| > x\} = 0$$

for any $x > 0$.

Let $\xi_{t,T}$ be defined by Equation (S13). Then according to the corollary to Theorem 1 of Heyde and Hall (1981), we have

$$\sup_x \left| P\left\{ \sum_{t=1}^T \xi_{t,T} > x \right\} - \Phi(x) \right|$$

$$\le C\left(\sum_{t=1}^T E(\xi_{t,T})^4 + E(a'R(y, \rho, \alpha, T)a - 1)^2 \right).$$

In the proof of part (b) of Lemma S11, we showed that the first term is less than C/T, where C depends only on p, δ, M, and K. As a result, we have convergence of the distribution of $a'S(y, \rho, \alpha, T)$ to $N(0, 1)$ uniformly over $\mathcal{B}_T$ and uniformly over $F_T \in \mathcal{L}_4(K, M, \theta)$. This finishes the proof of part (1).

The proof of part (2) is exactly the same as the proof of Theorem 3.

(3) Let X_t be defined as in Lemma S9. Then $\widehat{e}_t = \varepsilon_t + (\alpha - \widehat{\alpha}(\rho))'X_t$. We have

$$
\mu_2(F_T^{\text{err}}) - 1 = \left(\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 - 1 \right) + 2(\alpha - \widehat{\alpha}(\rho))' \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t X_t
$$

$$
+ (\alpha - \widehat{\alpha}(\rho))' \frac{1}{T} \sum_{t=1}^{T} X_t X_t' (\alpha - \widehat{\alpha}(\rho)).
$$

From Lemma S9, we know that $\widehat{\alpha}(\rho)$ is a uniformly consistent estimate of α. According to law of large numbers $\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 - 1 \to 0$ a.s., $\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t X_t \to 0$ a.s., and $\sum_{t=1}^{T} X_t X_t'$ is bounded almost surely. As a result, we have convergence of $\mu_2(F_T^{\text{err}}) - 1$ to zero almost surely. The third condition of the class $\mathcal{L}_4(K, M, \theta)$ can be checked in a similar way. *Q.E.D.*

S6. SUBSAMPLING

In this section we clarify some technical details of the proof of subsampling invalidity (Theorem 4 of the paper).

First, we note that local-to-unity asymptotic results (Phillips (1987) were established for processes starting from zero, whereas for subsampling we need to make a different assumption about initial condition.[1] If $|\rho| < 1$, the initial variable z_0 is normally distributed with mean $a/(1 - \rho)$ (here a is the value of the intercept) and variance $1/(1 - \rho^2)$. When $\rho = 1$, the initial value is an arbitrary constant. Lemma S13 below follows the line of reasoning proposed by Elliott (1999) and Elliott and Stock (2001).

LEMMA S13: *Let $u_j = \rho u_{j-1} + e_j$, $u_0 = 0$ with errors e_j being independent and identically distributed standard normal. Let us define $z_j = \rho^j(\xi/\sqrt{1 - \rho^2}) + u_j$, where $\xi \sim N(0, 1)$ is distributed independently of $\{e_j\}_{j=1}^{\infty}$. Let*

$$
t^{\mu} = \frac{\sum_{j=1}^{T} z_{j-1}^{\mu} e_j}{\sqrt{\sum_{j=1}^{T} (z_{j-1}^{\mu})^2}}.
$$

[1] I thank Don Andrews and Patrik Guggenberger for pointing this out.

We consider $\rho = 1 + c/T$ for some $c < 0$. Let $K_c(s) = J_c(s) + (e^{cs}/\sqrt{-2c})\xi$, where J_c is an Ornstein–Uhlenbeck process independent on ξ. Let

$$K_c^{\mu}(s) = K_c(s) - \int_0^1 K_c(t)\, dt = J_c^{\mu}(s) + \frac{\xi}{\sqrt{-2c}}\left(e^{cs} - \frac{1 - e^c}{-c}\right)$$

stay for the demeaned version of K_c. Then

$$t^{\mu} \Rightarrow \frac{\int_0^1 K_c^{\mu}(t)\, dw(t)}{\sqrt{\int_0^1 (K_c^{\mu}(t))^2\, dt}} \quad as \quad T \to \infty.$$

PROOF: All asymptotic convergence statements below hold simultaneously. It is easy to see that

$$\frac{1}{T}\sum_{j=1}^{T} z_{j-1} e_j = \frac{1}{T}\sum_{j=1}^{T} u_{j-1} e_j + \frac{\xi}{\sqrt{1 - \rho^2}}\frac{1}{T}\sum_{j=1}^{T}\rho^{j-1} e_j$$

$$\Rightarrow \quad \int_0^1 J_c(t)\, dw(t) + \frac{\xi}{\sqrt{-2c}}\int_0^1 e^{cs}\, dw(s) = \int_0^1 K_c(s)\, dw(s).$$

For the denominator, we have

$$\frac{1}{T^2}\sum_{j=1}^{T} z_{j-1}^2 = \frac{1}{T^2}\sum_{j=1}^{T}\left(u_{j-1} + \rho^{j-1}\frac{\xi}{\sqrt{1 - \rho^2}}\right)^2$$

$$= \frac{1}{T^2}\sum_{j=1}^{T} u_{j-1}^2 + 2\frac{\xi}{\sqrt{1 - \rho^2}}\frac{1}{T^2}\sum_{j=1}^{T}\rho^{j-1} u_{j-1}$$

$$+ \left(\frac{\xi}{\sqrt{1 - \rho^2}}\right)^2 \frac{1}{T^2}\sum_{j=1}^{T}\rho^{2(j-1)}.$$

We know that $(1/T^2)\sum_{j=1}^{T} u_{j-1}^2 \Rightarrow \int_0^1 J_c^2(t)\, dt$ (Phillips (1987)). We notice that $1/(T(1 - \rho^2)) \to 1/(-2c)$ and $(1/T)\sum_{j=1}^{T}\rho^{2(j-1)} = (1 - \rho^{2T})/(T(1 - \rho^2)) \to (1 - e^{2c})/(-2c)$. The next observation is

$$\frac{1}{T^{3/2}}\sum_{j=1}^{T}\rho^{j-1} u_{j-1} = \frac{1}{T^{3/2}}\sum_{i=1}^{T} e_i\left(\sum_{j=i}^{T}\rho^j \rho^{j-i}\right) = \frac{1}{T^{3/2}}\sum_{i=1}^{T} e_i \rho^i \frac{1 - \rho^{2(T-i)}}{1 - \rho^2}$$

$$= \frac{1}{(1 - \rho^2)T^{3/2}}\sum_{i=1}^{T} e_i \rho^i - \frac{\rho^T}{(1 - \rho^2)T^{3/2}} u_T$$

$$\Rightarrow \frac{1}{-2c}\left(\int_0^1 e^{cs}\,dw(s) - e^c J_c(1)\right).$$

As a result,

$$\frac{1}{T^2}\sum_{j=1}^{T} z_{j-1}^2 \Rightarrow \int_0^1 J_c^2(t)\,dt + 2\frac{\xi}{\sqrt{-2c}}\frac{1}{-2c}\left(\int_0^1 e^{cs}\,dw(s) - e^c J_c(1)\right)$$

$$+ \left(\frac{\xi}{\sqrt{-2c}}\right)^2 \frac{1 - e^{2c}}{-2c}.$$

We notice that

$$\int_0^1 K_c^2(s)\,ds$$

$$= \int_0^1 \left(J_c(s) + \frac{\xi}{\sqrt{-2c}}e^{cs}\right)ds$$

$$= \int_0^1 J_c^2(s)\,ds + 2\frac{\xi}{\sqrt{-2c}}\int_0^1 (J_c(s)e^{cs})\,ds + \left(\frac{\xi}{\sqrt{-2c}}\right)^2 \int_0^1 e^{2cs}\,ds.$$

Consider in more detail the integral

$$\int_0^1 (J_c(s)e^{cs})\,ds$$

$$= \int_0^1 \left(\int_0^s e^{c(s-t)}\,dw(t)\,e^{cs}\right)ds = \int_0^1 e^{ct}\left(\int_s^1 e^{2c(s-t)}\,ds\right)dw(t)$$

$$= \frac{1}{-2c}\int_0^1 e^{ct}(1 - e^{2c(1-s)})\,dw(t) = \frac{1}{-2c}\left(\int_0^1 e^{cs}\,dw(s) - e^c J_c(1)\right).$$

So we have

$$\frac{1}{T^2}\sum_{j=1}^{T} z_{j-1}^2 \Rightarrow \int_0^1 K_c^2(t)\,dt.$$

Now let us move to a model with demeaning. What will change in our results? For the numerator, we have

$$\frac{1}{T}\sum_{j=1}^{T} z_{j-1}^{\mu} e_j = \frac{1}{T}\sum_{j=1}^{T} u_{j-1}^{\mu} e_j + u_0\frac{1}{T}\sum_{j=1}^{T}\left(\rho^{j-1} - \frac{1 - \rho^T}{T(1-\rho)}\right)e_j$$

$$\Rightarrow \quad \int_0^1 J_c^\mu(t)\,dw(t) + \frac{\xi}{\sqrt{-2c}}\left(\int_0^1 e^{cs}\,dw(s) - \frac{1-e^c}{-c}w(1)\right)$$

$$= \int_0^1 K_c^\mu(t)\,dw(t).$$

We handle the denominator in a similar way:

$$\frac{1}{T^2}\sum_{j=1}^{T}(z_{j-1}^\mu)^2 = \frac{1}{T^2}\sum_{j=1}^{T}\left(u_{j-1}^\mu + \left(\rho^{j-1} - \frac{1-\rho^T}{T(1-\rho)}\right)\frac{\xi}{\sqrt{1-\rho^2}}\right)^2$$

$$= \frac{1}{T^2}\sum_{j=1}^{T}(u_{j-1}^\mu)^2 + 2\frac{\xi}{\sqrt{1-\rho^2}}\frac{1}{T^2}\sum_{j=1}^{T}\rho^{j-1}u_{j-1}^\mu$$

$$+ \left(\frac{\xi}{\sqrt{1-\rho^2}}\right)^2\frac{1}{T^2}\sum_{j=1}^{T}\left(\rho^{j-1} - \frac{1-\rho^T}{T(1-\rho)}\right)^2.$$

Similarly to above, we can determine that

$$\frac{1}{T^2}\sum_{j=1}^{T}(z_{j-1}^\mu)^2 \Rightarrow \int_0^1 (K_c^\mu(t))^2\,dt.$$

Finally we get

$$t^\mu \Rightarrow \frac{\int_0^1 K_c^\mu(t)\,dw(t)}{\sqrt{\int_0^1 (K_c^\mu(t))^2\,dt}}. \qquad\qquad Q.E.D.$$

The quantiles of the distribution of $t_K^c = (\int_0^1 K_c^\mu(t)\,dw(t))/\sqrt{\int_0^1 (K_c^\mu(t))^2\,dt}$ have not been reported in literature, so we have to simulate critical values. We also show that for at least one $c < 0$, if we use an equitailed interval based on the distribution of t_K^c, whereas the true variable is normal, then the coverage will be smaller than declared.

We simulated quantiles and coverage for $-c = 0.05, 0.1, 0.5, 1, 2, 4, 10, 15,$ 20, 25. The simulations are based on samples of size $T = 300$. We performed 5,000 simulations. The results are reported in Figures S1 and S2.

The second technicality we address in this appendix is related to the rate of mixing coefficients decay for summands in empirical cumulative distribution functions.
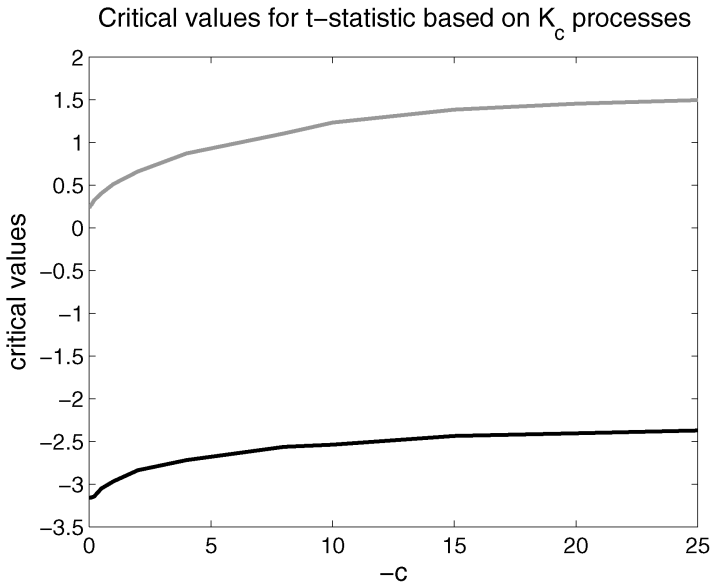
Critical values for t–statistic based on K_c processes

FIGURE S1.—The 2.5% and 97.7% quantiles of statistic t_K^c. Quantiles are based on simulated t-statistics for AR(1) processes with a constant and stationary initial distribution for values of the AR parameter $\rho = 1 + c/T$ local to unity. Number of simulations 5,000; sample size $T = 300$; normal errors.
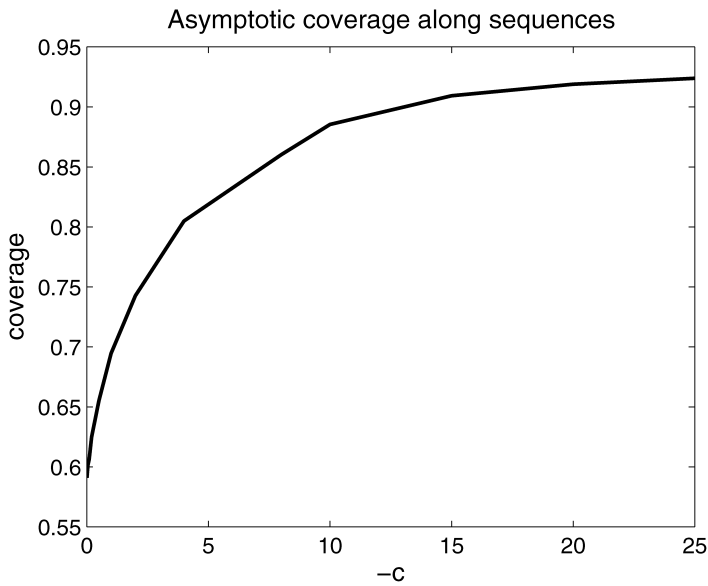
Asymptotic coverage along sequences

FIGURE S2.—Coverage of equitailed intervals based on the distribution of t_K^c, whereas the true distribution is standard normal. Based on simulated quantiles as in Figure S1.

LEMMA S14: *Given the assumptions made in Section* 6 *of the paper, we have*

$$\lim_{T \to \infty} \sup_{x} \left| L_{T,b}(x) - P \left\{ \frac{\int_0^1 K_c^\mu(s) \, dw(s)}{\sqrt{\int_0^1 (K_c^\mu(s))^2 \, ds}} < x \right\} \right| = 0 \quad in\ probability.$$

PROOF: We follow the lines of the proof of Theorem 3.1 of Romano and Wolf (2001), substituting their statistic for the corresponding *t*-statistic. The only thing we need to check is that

$$\frac{1}{T} \sum_{h=1}^{T-b_T} \alpha_{T,b_T}(h) \to 0 \quad \text{as} \quad T \to \infty,$$

where $\alpha_{T,b_T}(h)$ are strong mixing coefficients for an array of variables $\{t_j(b_T)\}_{j=1}^{T-b_T}$. The α-mixing coefficient $\alpha_{T,b_T}(h)$ does not exceed the α-mixing coefficients for a set of subsamples $\{z_1, \ldots, z_b\}$ and $\{z_{h+1}, \ldots, z_{h+b}\}$, where z_j is a Gaussian AR(1) process with AR coefficient $\rho = 1 + c/b_T$. The latter is not bigger than the α-mixing coefficient $\alpha_z(h - b_T)$ for the process z.

We use a statement proved below that $\alpha_z(h) \leq \rho^h$. Then

$$\frac{1}{T} \sum_{h=1}^{T-b_T} \alpha_{T,b_T}(h) \leq \frac{1}{T} \sum_{h=1}^{T-b_T} \min\{1, \rho_T^{h-b_T}\} = C \frac{1}{(1 - \rho_T)T} \to 0$$

as $T \to \infty$.

The last statement holds since $(1 - \rho_T)T \to \infty$.   *Q.E.D.*

LEMMA S15: *Let $z_t = \rho z_{t-1} + u_t$ be a stationary Gaussian* AR(1) *process. Then*

$$\alpha_z(h) \leq \rho^h.$$

PROOF: From the definition of mixing coefficients, we have $\alpha_z(h) \leq \rho_z(h)$. Here the ρ-mixing coefficient $\rho_z(h)$ is the maximum correlation between the variables that are measurable with respect to the two σ algebras. According to Kolmogorov and Rozanov (1960), it is enough to restrict attention to linear functions of the variables $\{z_j\}_{j \leq t}$ and $\{z_j\}_{j \geq t+n}$.

According to Ibragimov (1970) (see formula (4.2)),

$$\rho(n) = \sup_{\varphi, \psi} |(\varphi e^{in\lambda}, \overline{\psi})_f| = \sup_{\varphi, \psi} \left| \int_{-\pi}^{\pi} \varphi(\lambda) \psi(\lambda) e^{in\lambda} f(\lambda) \, d\lambda \right|,$$

where $f(\lambda) = \sum_{k=-\infty}^{\infty} e^{ik\lambda} \rho^{|k|}$ is a spectral density function, and φ and ψ are polynomials of $e^{i\lambda}$ with the condition $\|\varphi\|_f = \|\psi\|_f = 1$. Here we use $(\varphi, \psi)_f = \int_{-\pi}^{\pi} \varphi(\lambda) \overline{\psi}(\lambda) f(\lambda) \, d\lambda$.

Let $\varphi(\lambda) = \sum_{k=0}^{L} a_k e^{ik\lambda}$ and $\psi(\lambda) = \sum_{j=0}^{K} b_j e^{ij\lambda}$. Then

$$\frac{|\int_{-\pi}^{\pi} \varphi(\lambda)\psi(\lambda)e^{in\lambda} f(\lambda)\,d\lambda|}{\|\varphi\|_f \|\psi\|_f}$$

$$= \rho^n \left| \frac{\sum_{k,j} a_k b_j \rho^{k+j}}{\sqrt{(\sum_{k,k'} a_k a_{k'} \rho^{|k-k'|})(\sum_{j,j'} b_j b_{j'} \rho^{|j-j'|})}} \right|.$$

Let us define matrices $A = (\rho^{|i-j|})_{i,j}$ and $B = (\rho^{i+j-2})_{i,j}$. Then

$$\rho(n) = \rho^n \sup_{a,b} \left| \frac{a'Bb}{\sqrt{a'Aa}\sqrt{b'Ab}} \right|,$$

where $a = (a_0, a_1, \ldots, a_M)$, $b = (b_0, \ldots, b_M)$, and $M = \max\{L, K\}$. We also define a matrix

$$L = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

Then $LAL' = \text{diag}(1-\rho^2, 1-\rho^2, \ldots, 1-\rho^2) = \widetilde{A}$ and $LBL' = \text{diag}(1-\rho^2, 0, \ldots, 0) = \widetilde{B}$. Let $\widetilde{a} = (L')^{-1}a$ and $\widetilde{b} = (L')^{-1}b$. Then

$$\rho(n) = \rho^n \sup_{a,b} \left| \frac{a'Bb}{\sqrt{a'Aa}\sqrt{b'Ab}} \right| = \rho^n \sup_{\widetilde{a},\widetilde{b}} \frac{|\widetilde{a}'\widetilde{B}\widetilde{b}|}{\sqrt{\widetilde{a}'\widetilde{A}\widetilde{a}}\sqrt{\widetilde{b}'\widetilde{A}\widetilde{b}}}$$

$$= \rho^n \sup_{\widetilde{a},\widetilde{b}} \frac{|\widetilde{a}_0 \widetilde{b}_0|}{\sqrt{(\sum \widetilde{a}_k^2)(\sum \widetilde{b}_k^2)}} = \rho^n. \qquad\qquad Q.E.D.$$

S7. SIMULATIONS

We performed a small simulation study to assess the extend to which asymptotic results are reflected in finite samples. The study is intended to fulfill the goals listed below:

- Check finite-sample performance of the three procedures, the validity of which was proven in the paper.
- Explore sensitivity of the described methods to nonsymmetry or heavy-tailedness of the distribution of error terms.
- Compare the accuracy of the three methods.
- Assess the size distortion of subsampling; that is, whether it is as extreme as predicted by the asymptotic results of Andrews and Guggenberger (2007).

- Examine how coverage properties of subsampling intervals depend on block size and determine for what range of AR coefficients it is safe to use subsampling.

We start with the first group of objectives concerning the three methods for which we provided proofs. We simulate an AR(1) model with a linear trend since this is the setup where the distortions are most pronounced. We used normal errors, errors having centered χ^2 distributions with 4 and 8 degrees of freedom, and errors following the ARCH(1) process with parameters 0.3 and 0.85. Those specifications are taken from Andrews (1993). We used sample size $T = 120$ as a typical one for macroeconomic time series. We performed simulations for ρ equals to 0.3, 0.5, 0.7, 0.8, 0.9, 0.95, 0.99 and 1. This range of values of ρ covers some values in the stationary region, in close proximity to the unit root, as well as in the intermediate range. The number of simulations is equal to 1000. Some of the results are reported in Tables I and II.

All three methods achieved 95% coverage for an AR(1) model with linear trend and normal errors for all values of ρ that we checked (we did not report these results in the tables). Table I is intended to show that all methods seems to be robust toward asymmetry and heavy-tailedness of the distribution of error terms. We should also note that there is no strong leader among the three methods. In Table II we allowed conditional heteroscedasticity. Strictly speaking our proofs do not allow for heteroscedasticity. We can see that the methods failed in this setup, and the coverage may fall as low as 70%.

Now we turn to subsampling. According to our results reported in Section 6 of the paper, the subsampling procedure fails to provide asymptotically correct confidence sets. According to Andrews and Guggenberger (2007), the asymptotic coverage is as low as 26% for an AR(1) with a linear time trend. We would like to know the extent to which these asymptotic results are reflected in finite samples.

TABLE I

COVERAGE OF INTERVALS FOR THE AR COEFFICIENT IN AN AR(1) MODEL WITH A LINEAR TREND ($y_t = a + bt + x_t$, $x_t = \rho x_{t-1} + \varepsilon_t$; SAMPLE SIZE 120)

ρ	$\varepsilon_i \sim \chi_4^2 - 4$			$\varepsilon_i \sim \chi_8^2 - 8$		
	Andrews (1993)	Stock (1991)	Hansen (1999)	Andrews (1993)	Stock (1991)	Hansen (1999)
0.3	0.95	0.97	0.97	0.96	0.96	0.96
0.5	0.96	0.96	0.96	0.96	0.96	0.96
0.7	0.97	0.96	0.96	0.96	0.95	0.95
0.8	0.97	0.96	0.96	0.96	0.95	0.95
0.9	0.97	0.96	0.96	0.97	0.96	0.96
0.95	0.97	0.96	0.96	0.96	0.97	0.96
0.99	0.95	0.95	0.95	0.96	0.96	0.95
1	0.95	0.96	0.96	0.96	0.95	0.96

TABLE II

COVERAGE OF INTERVALS FOR THE AR COEFFICIENT IN AN AR(1) MODEL WITH A LINEAR
TREND ($y_t = a + bt + x_t$, $x_t = \rho x_{t-1} + \varepsilon_t$; SAMPLE SIZE 120)

ρ	$\varepsilon_i \sim$ ARCH(0.3)			$\varepsilon_i \sim$ ARCH(0.85)		
	Andrews (1993)	Stock (1991)	Hansen (1999)	Andrews (1993)	Stock (1991)	Hansen (1999)
0.3	0.89	0.88	0.90	0.70	0.72	0.73
0.5	0.92	0.89	0.90	0.72	0.72	0.73
0.7	0.93	0.90	0.91	0.73	0.76	0.77
0.8	0.93	0.91	0.92	0.77	0.79	0.80
0.9	0.94	0.92	0.92	0.83	0.84	0.85
0.95	0.95	0.94	0.94	0.85	0.88	0.87
0.99	0.96	0.95	0.95	0.90	0.90	0.92
1	0.95	0.95	0.95	0.91	0.91	0.93

According to the proof of Theorem 4, bad coverage is expected for intermediate values of ρ. Romano and Wolf (2001) provided some simulations regarding the coverage of subsampling intervals, but for a very restricted set of values of $\rho \in \{0.6, 0.9, 0.95, 0.99, 1\}$. We repeated their exercise for a wider range of ρ's and for several different sample sizes $T = 120, 240, 480, 960$. For each sample size we tried several block sizes. For $T = 120$ and 240, we used the same set of block sizes as used by Romano and Wolf. For $T = 480$ and 960, we used block sizes b that approximately follow the rule proposed by Romano and Wolf: $b = cT^{1/2}$, $c \in [0.5, 3]$. For all simulations we used a model with normal homoscedastic errors only. All results are summarized on Figure S3.

First of all, we should note that subsampling yields undercoverage for quite a wide range of ρ's. However, the amount of undercoverage is not as extreme as predicted by Andrews and Guggenberger (2007). One more interesting aspect could be noted: the size property of the procedure becomes worse as the sample size increases! According to the intuition of Theorem 4, the size distortion becomes pronounced when T/b_T is large, which can only be true for large sample sizes. As for the right choice of block size, there is no clear leader: for different ranges of ρ and for different T, different block sizes serve better.

One main conclusion of our simulation study is that we do not recommend the use of the subsampling procedure in empirical studies to make inferences about the persistence of a time series.

Dept. of Economics, Massachusetts Instiyute of Technology, 50 Memoriol Drive, Cambridge, MA 02142, U.S.A.; amikushe@mit.edu.
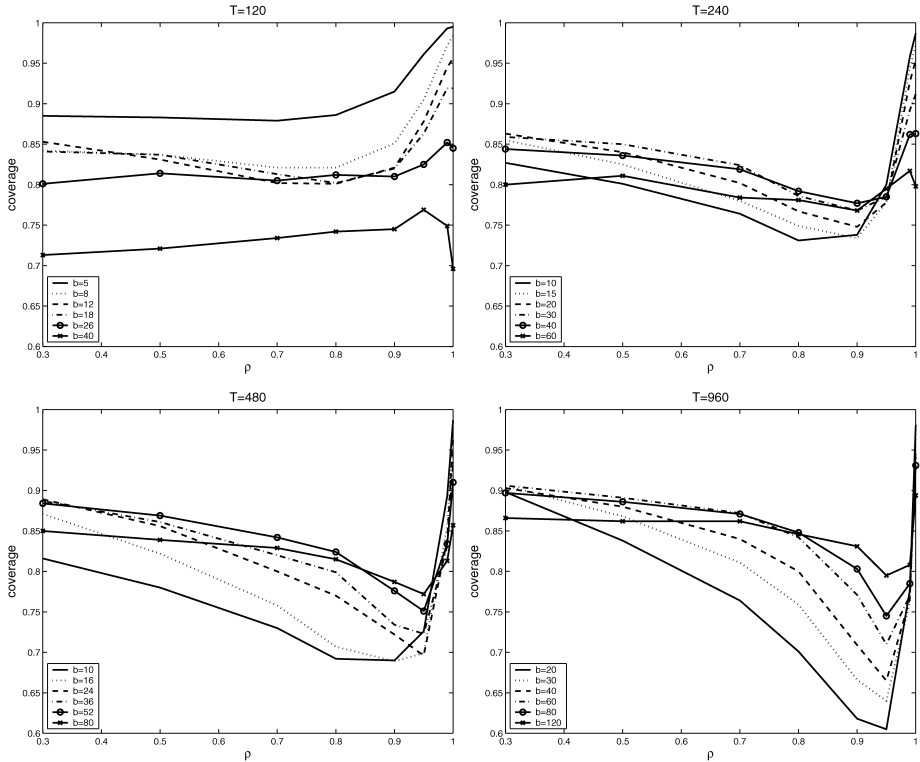
FIGURE S3.——Coverage of equitailed subsampling intervals with nominal level 95% for an AR(1) model with a linear time trend and normal errors. Number of simulations 1,000.

REFERENCES

ANDREWS, D. W. K. (1993): "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61, 139–165. [20,42,43]

ANDREWS, D. W. K., AND P. GUGGENBERGER (2007): "Hybrid and Size-Corrected Subsample Methods," Discussion Paper 1606, Cowles Foundation. [41-43]

CSÖRGŐ, M., AND P. RÉVÉSZ (1981): *Strong Approximations in Probability and Statistics*. New York: Academic Press. [3,11]

ELLIOTT, G. (1999): "Efficient Tests for an Autoregressive Unit Root when the Initial Observation Is Drawn from Its Unconditional Distribution," *International Economic Review*, 40, 767–783. [35]

ELLIOTT, G., AND J. H. STOCK (2001): "Confidence Intervals for Autoregressive Coefficients Near One," *Journal of Econometrics*, 103, 155–181. [35]

HALL, P., AND C. C. HEYDE (1981): "Rates of Convergence in the Martingale Central Limit Theorem," *Annals of Probability*, 9, 395–404. [9,13]

HANSEN, B. E. (1999): "The Grid Bootstrap and the Autoregressive Model," *Review of Economics and Statistics*, 81, 594–607. [20,42,43]

GIRAITIS, L., AND P. C. B. PHILLIPS (2006): "Uniform Limit Theory for Stationary Autoregression," *Journal of Time Series Analysis*, 27, 51–60. [9,11,29]

IBRAGIMOV, I. A. (1970): "On the Spectrum of Stationary Gaussian Sequences Satisfying the Strong Mixing Condition. II," *Theory Probability and Its Applications*, 15, 23–36. [40]

KOLMOGOROV, A. N., AND Y. A. ROZANOV (1960): "On Strong Mixing Conditions for Stationary Gaussian Processes," *Theory Probability and Its Applications*, 5, 204–208. [40]

MONTGOMERY-SMITH, S. J. (1993): "Comparison of Sums of Independent Identically Distributed Random Vectors," *Probability and Mathematical Statistics*, 14, 281–285. [12]

PARK, J. Y., AND P. C. B. PHILLIPS (1999): "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory*, 15, 269–298. [2]

PETROV, V. V. (1975): *Sums of Independent Random Variables*. New York: Springer-Verlag. [12]

PHILLIPS, P. C. B. (1987): "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547. [35,36]

ROMANO, J. P., AND M. WOLF (2001): "Subsampling Intervals in Autoregressive Models with Linear Time Trend," *Econometrica*, 69, 1283–1314. [40,43]

STOCK, J. (1991): "Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435–459. [42,43]