Econometrica Supplementary Material

SUPPLEMENT TO "FIXED-SMOOTHING ASYMPTOTICS IN A TWO-STEP GENERALIZED METHOD OF MOMENTS FRAMEWORK"

(*Econometrica*, Vol. 82, No. 6, November 2014, 2327–2370)

BY YIXIAO SUN

THIS SUPPLEMENT HAS TWO PARTS. Appendix B provides proofs of the technical lemmas and other results in the paper. Appendix C provides some practical guidance on implementing the testing procedures proposed in the paper. It also applies the proposed test procedures to make inference in a stochastic volatility model.

APPENDIX B: PROOFS

B.1. *Proofs of Technical Lemmas*

PROOF OF LEMMA A.2: Since (ξ_{1T}, η_{1T}) converges weakly, we can apply Lemma A.1 with $s(T) = T$. It follows from the condition in (A.1) that $Ef(\xi_{1T}, \eta_{1T}) - Ef(\xi_{2T}, \eta_{2T}) \to 0$ for any $f \in \mathcal{BC}$. But $Ef(\xi_{1T}, \eta_{1T}) \to Ef(\xi, \eta)$ and so $Ef(\xi_{2T}, \eta_{2T}) \to Ef(\xi, \eta)$ for any $f \in \mathcal{BC}$. That is, (ξ_{2T}, η_{2T}) also converges weakly to (ξ, η). Using the same proof for proving the continuous mapping theorem, we have $Ef(g(\xi_{1T}, \eta_{1T})) - Ef(g(\xi_{2T}, \eta_{2T})) \to 0$ for any $f \in \mathcal{BC}$. Therefore, $g(\xi_{1T}, \eta_{1T}) \overset{a}{\sim} g(\xi_{2T}, \eta_{2T})$. *Q.E.D.*

PROOF OF LEMMA A.3: Let $\varepsilon > 0$ and $\xi \in \mathbb{R}$. Under condition (iii), we can find a $\delta := \delta(\varepsilon) > 0$ such that for some integer $T_{\min} > 0$,

$$P(\xi - \delta \le \xi_T < \xi + \delta) \le \varepsilon$$

for all $T \ge T_{\min}$. Here $T_{\min}$ does not depend on δ or ε. Under condition (iv), we can find a $J_{\min} := J_{\min}(\varepsilon)$ that does not depend on T such that

$$P\big(|\eta_{T,J}| > \delta\big) \le \varepsilon$$

for all $J \ge J_{\min}$ and all T. From condition (ii), we can find a $J'_{\min} \ge J_{\min}$ and a $T'_{\min} \ge T_{\min}$ such that

$$\big|P\big(\xi^*_{T,J} < \xi\big) - P(\xi_T < \xi)\big| \le \varepsilon$$

for all $J \ge J'_{\min}$ and all $T \ge T'_{\min}$. It follows from condition (i) that for any finite $J_0 \ge J'_{\min}$, there exists a $T''_{\min}(J_0) \ge T'_{\min} \ge T_{\min}$ such that

$$\big|P(\xi_{T,J_0} < \xi + \delta) - P\big(\xi^*_{T,J_0} < \xi + \delta\big)\big| \le \varepsilon,$$

$$\big|P(\xi_{T,J_0} < \xi - \delta) - P\big(\xi^*_{T,J_0} < \xi - \delta\big)\big| \le \varepsilon$$

 DOI: 10.3982/ECTA11684

for $T \geq T''_{\min}(J_0)$.

When $T \geq T''_{\min}(J_0)$, we have

$$\begin{aligned} P(\omega_T \leq \xi) &= P(\xi_{T,J_0} + \eta_{T,J_0} \leq \xi) \leq P(\xi_{T,J_0} \leq \xi + \delta) + P\big(|\eta_{T,J_0}| > \delta\big) \\ &\leq P\big(\xi^*_{T,J_0} \leq \xi + \delta\big) + 2\varepsilon \leq P(\xi_T < \xi + \delta) + 3\varepsilon \\ &\leq P(\xi_T < \xi) + 4\varepsilon. \end{aligned}$$

Similarly,

$$\begin{aligned} P(\omega_T \leq \xi) &= P(\xi_{T,J_0} + \eta_{T,J_0} \leq \xi) \geq P(\xi_{T,J_0} \leq \xi - \delta) - P\big(|\eta_{T,J_0}| \geq \delta\big) \\ &\geq P\big(\xi^*_{T,J_0} \leq \xi - \delta\big) - 2\varepsilon \geq P(\xi_T \leq \xi - \delta) - 3\varepsilon \\ &\geq P(\xi_T \leq \xi) - 4\varepsilon. \end{aligned}$$

Since the above two inequalities hold for all $\varepsilon > 0$, we must have $P(\omega_T < \xi) = P(\xi_T < \xi) + o(1)$ as $T \to \infty$. *Q.E.D.*

B.2. *Proofs of Other Results*

PROOF OF THEOREM 2: Let $\boldsymbol{\lambda}_T$ be the $p \times 1$ vector of Lagrange multipliers for the constrained GMM estimation. The first order conditions for $\hat{\theta}_{T,R}$ are

$$\text{(B.1)} \qquad \frac{\partial g'_T(\hat{\theta}_{T,R})}{\partial \theta} W_T^{-1}(\tilde{\theta}_T) g_T(\hat{\theta}_{T,R}) + R'\boldsymbol{\lambda}_T = 0 \quad \text{and} \quad R\hat{\theta}_{T,R} = r.$$

Linearizing the first set of conditions and using Assumption 4, we have the system of equations

$$\begin{pmatrix} \tilde{\Psi} & R' \\ R & \mathbf{0}_{p\times p} \end{pmatrix} \begin{pmatrix} \sqrt{T}(\hat{\theta}_{T,R} - \theta_0) \\ \sqrt{T}\boldsymbol{\lambda}_T \end{pmatrix} = \begin{pmatrix} -G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T} g_T(\theta_0) \\ \mathbf{0}_{p\times 1} \end{pmatrix} + o_p(1),$$

where $\tilde{\Psi} := \tilde{\Psi}(\tilde{\theta}_T) = G'W_T^{-1}(\tilde{\theta}_T)G$. From this, we get

$$\text{(B.2)} \qquad \begin{aligned} &\sqrt{T}(\hat{\theta}_{T,R} - \theta_0) \\ &\quad = -\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T} g_T(\theta_0) \\ &\qquad - \tilde{\Psi}^{-1}R'\big\{R\tilde{\Psi}^{-1}R'\big\}^{-1} R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T} g_T(\theta_0) + o_p(1) \end{aligned}$$

and

$$\text{(B.3)} \qquad \sqrt{T}\boldsymbol{\lambda}_T = -\big\{R\tilde{\Psi}^{-1}R'\big\}^{-1} R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T} g_T(\theta_0) + o_p(1).$$

Combining (5) with (B.2), we have

$$\begin{aligned}\text{(B.4)}\qquad & \sqrt{T}(\hat{\theta}_{T,R}-\hat{\theta}_T) \\ & = -\tilde{\Psi}^{-1}R'\{R\tilde{\Psi}^{-1}R'\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T(\theta_0)+o_p(1),\end{aligned}$$

which implies that $\sqrt{T}(\hat{\theta}_{T,R}-\hat{\theta}_T)=O_p(1)$. So

$$g_T(\hat{\theta}_{T,R})=g_T(\hat{\theta}_T)+G_T(\hat{\theta}_T)(\hat{\theta}_{T,R}-\hat{\theta}_T)+o_p(1/\sqrt{T}).$$

Plugging this into the definition of $\mathbb{D}_T$, we obtain

$$\begin{aligned}\text{(B.5)}\qquad \mathbb{D}_T = {} & T(\hat{\theta}_{T,R}-\hat{\theta}_T)'G_T'(\hat{\theta}_T)W_T^{-1}(\breve{\theta}_T)G_T(\hat{\theta}_T)(\hat{\theta}_{T,R}-\hat{\theta}_T)/p \\ & -2Tg_T'(\hat{\theta}_T)W_T^{-1}(\breve{\theta}_T)G(\hat{\theta}_T)(\hat{\theta}_{T,R}-\hat{\theta}_T)/p+o_p(1).\end{aligned}$$

Using the first order conditions for $\hat{\theta}_T$: $g_T'(\hat{\theta}_T)W_T^{-1}(\tilde{\theta}_T)G_T(\hat{\theta}_T)=0$ and Lemma 1(a), (b), we obtain

$$\text{(B.6)}\qquad \mathbb{D}_T=T(\hat{\theta}_{T,R}-\hat{\theta}_T)'\tilde{\Psi}(\hat{\theta}_{T,R}-\hat{\theta}_T)/p+o_p(1).$$

Plugging (B.4) into (B.6) and simplifying the resulting expression, we have

$$\begin{aligned}\text{(B.7)}\qquad \mathbb{D}_T &= \big[R'\{R\tilde{\Psi}^{-1}R'\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T(\theta_0)\big]' \\ &\quad \times\tilde{\Psi}^{-1}\big[R'\{R\tilde{\Psi}^{-1}R'\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T(\theta_0)\big]/p+o_p(1) \\ &= \big[R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T(\theta_0)\big]'\big[R\tilde{\Psi}^{-1}R'\big]^{-1} \\ &\quad \times\big[R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T(\theta_0)\big]/p+o_p(1) \\ &= \big[\sqrt{T}R(\hat{\theta}_T-\theta_0)\big]'\big[R\tilde{\Psi}^{-1}R'\big]^{-1}\big[\sqrt{T}R(\hat{\theta}_T-\theta_0)\big]/p+o_p(1) \\ &= \mathbb{W}_T+o_p(1).\end{aligned}$$

Next, we prove the second result in the theorem. In view of the first order conditions in (B.1) and equation (B.3), we have

$$\begin{aligned}\sqrt{T}\Delta_T(\hat{\theta}_{T,R}) &= \frac{\partial g_T'(\hat{\theta}_{T,R})}{\partial\theta}W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T(\hat{\theta}_{T,R})+o_p(1) \\ &= -\sqrt{T}R'\boldsymbol{\lambda}_T+o_p(1) \\ &= R'\{R\tilde{\Psi}^{-1}R'\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T(\theta_0)+o_p(1) \\ &= \tilde{\Psi}\sqrt{T}(\hat{\theta}_T-\hat{\theta}_{T,R})+o_p(1),\end{aligned}$$

and so

$$\mathbb{S}_T = T(\hat{\theta}_T - \hat{\theta}_{T,R})'\tilde{\Psi}(\hat{\theta}_T - \hat{\theta}_{T,R})/p + o_p(1)$$
$$= \mathbb{D}_T + o_p(1) = \mathbb{W}_T + o_p(1). \qquad Q.E.D.$$

PROOF OF THEOREM 3: *Part* (a). Let

$$H = \left(\frac{B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)}{\|B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\|}, \bar{H} \right)$$

be an orthonormal matrix. Then

$$F_\infty \overset{d}{=} \left\|B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right\|^2 \left[\frac{B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)}{\|B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\|} \right]' H$$
$$\times H'D_{pp}^{-1}HH' \left[\frac{B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)}{\|B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\|} \right] \Big/ p$$
$$= \left\|B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right\|^2 e_p'\left[H'D_{pp}^{-1}H\right]e_p/p,$$

where $e_p = (1, 0, 0, \ldots, 0)' \in \mathbb{R}^p$. But $H'D_{pp}^{-1}H$ has the same distribution as D_{pp}^{-1} and D_{pp} is independent of $B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)$. So

$$\text{(B.8)} \qquad F_\infty \overset{d}{=} \frac{\|B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\|^2/p}{[e_p'D_{pp}^{-1}e_p]^{-1}}.$$

That is, F_∞ is equal in distribution to a ratio of two independent random variables.

It is easy to see that

$$\left[e_p'D_{pp}^{-1}e_p\right]^{-1} \overset{d}{=} \left[e_{p+q}' \left[\int_0^1 \int_0^1 Q_h^*(r,s)\, dB_{p+q}(r)\, dB_{p+q}'(s) \right]^{-1} e_{p+q} \right]^{-1}$$
$$\overset{d}{=} \frac{1}{K}\chi^2_{K-p-q+1},$$

where $e_{p+q} = (1, 0, \ldots, 0)' \in \mathbb{R}^{p+q}$. With this, we can now represent F_∞ as

$$\text{(B.9)} \qquad F_\infty \overset{d}{=} \frac{\chi_p^2(\Delta^2)/p}{\chi^2_{K-p-q+1}/K}$$

and so

$$\kappa^{-1}F_\infty \overset{d}{=} \frac{\chi_p^2(\Delta^2)/p}{\chi^2_{K-p-q+1}/(K-p-q+1)} \overset{d}{=} \mathcal{F}_{p,K-p-q+1}(\Delta^2). \tag{B.10}$$

Part (b). Since the numerator and the denominator in (B.10) are independent, $\kappa^{-1}F_\infty$ is distributed as a noncentral F distribution, conditional on Δ^2. More specifically, we have

$$P(\kappa^{-1}F_\infty < z) = P(\mathcal{F}_{p,K-p-q+1}(\Delta^2) < z) = EF_{p,K-p-q+1}(z, \Delta^2),$$

where $F_{p,K-p-q+1}(z,\lambda)$ is the CDF of the noncentral F distribution with degrees of freedom $(p, K-p-q+1)$ and noncentrality parameter λ, and $\mathcal{F}_{p,K-p-q+1}(\lambda)$ is a random variable with CDF $F_{p,K-p-q+1}(z,\lambda)$.

We proceed to compute the mean of Δ^2. Let

$$\xi_j = \int_0^1 \Phi_j(r)\,dB_p(r) \sim \text{i.i.d. } N(0, I_p) \quad \text{and}$$

$$\eta_j = \int_0^1 \Phi_j(r)\,dB_q(r) \sim \text{i.i.d. } N(0, I_q).$$

Note that $\{\xi_j\}$ are independent of $\{\eta_j\}$. We can represent C_{pq} and C_{qq} as

$$C_{pq} = K^{-1}\sum_{j=1}^K \xi_j\eta_j' \quad \text{and} \quad C_{qq} = K^{-1}\sum_{j=1}^K \eta_j\eta_j'. \tag{B.11}$$

So

$$\begin{aligned} E\Delta^2 &= EB_q(1)'C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}B_q(1) = E\,\mathrm{tr}(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}) \\ &= E\,\mathrm{tr}\left[\left(\frac{1}{K}\sum_{j=1}^K \eta_j\eta_j'\right)^{-1}\left(\frac{1}{K}\sum_{j=1}^K \eta_j\xi_j'\right) \right. \\ &\qquad \left. \times \left(\frac{1}{K}\sum_{j=1}^K \xi_j\eta_j'\right)\left(\frac{1}{K}\sum_{j=1}^K \eta_j\eta_j'\right)^{-1}\right] \\ &= p\,\mathrm{tr}\,E(\Pi), \end{aligned}$$

where $\Pi := (\sum_{j=1}^K \eta_j\eta_j')^{-1}$ follows an inverse Wishart distribution and

$$E(\Pi) = \frac{I_q}{K-q-1}$$

for K large enough. Therefore, $E\Delta^2 = \frac{pq}{K-q-1} = \delta^2$.

Next we compute the variance of Δ^2. It follows from the law of total variance that

$$\mathrm{var}(\Delta^2) = E\big[\mathrm{var}(\Delta^2 | C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1})\big] + \mathrm{var}\big(\big[\mathrm{tr}(C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1})\big]\big).$$

Note that $B_q(1)$ is independent of $C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1}$. So conditional on $C_{qq}^{-1} C'_{pq} \times C_{pq} C_{qq}^{-1}$, Δ^2 is a quadratic form in standard normals. Hence, the conditional variance of Δ^2 is

$$\mathrm{var}(\Delta^2 | C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1}) = 2\,\mathrm{tr}(C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1} C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1}).$$

Using the representation in (B.11), we have

$$\begin{aligned}
&E\,\mathrm{tr}(C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1} C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1}) \\
&\quad = E\,\mathrm{tr}\left(\frac{1}{K}\sum_{j=1}^{K} \eta_j \xi'_j\right)\left(\frac{1}{K}\sum_{j=1}^{K} \xi_j \eta'_j\right) C_{qq}^{-2} \\
&\qquad \times \left(\frac{1}{K}\sum_{j=1}^{K} \eta_j \xi'_j\right)\left(\frac{1}{K}\sum_{j=1}^{K} \xi_j \eta'_j\right) C_{qq}^{-2} \\
&\quad = E\,\mathrm{tr}\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{i_1=1}^{K} \eta_{j_1}(\xi'_{j_1}\xi_{i_1})\eta'_{i_1} C_{qq}^{-2} \sum_{j_2=1}^{K}\sum_{i_2=1}^{K} \eta_{j_2}(\xi'_{j_2}\xi_{i_2})\eta'_{i_2} C_{qq}^{-2}\right] \\
&\quad = E\,\mathrm{tr}\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{i_1=1}^{K}\sum_{j_2=1}^{K}\sum_{i_2=1}^{K} \eta_{j_1}\eta'_{i_1} C_{qq}^{-2} \eta_{j_2}\eta'_{i_2} C_{qq}^{-2} E(\xi'_{j_1}\xi_{i_1})(\xi'_{j_2}\xi_{i_2})\right].
\end{aligned}$$

Since

$$E(\xi'_{j_1}\xi_{i_1})(\xi'_{j_2}\xi_{i_2}) = \begin{cases} p^2, & j_1 = i_1, j_2 = i_2, \text{ and } j_1 \neq j_2, \\ p, & j_1 = j_2, i_1 = i_2, \text{ and } j_1 \neq i_1, \\ p, & j_1 = i_2, i_1 = j_2, \text{ and } j_1 \neq i_1, \\ p^2 + 2p, & j_1 = j_2 = i_1 = i_2, \\ 0, & \text{otherwise}, \end{cases}$$

we have

$$\begin{aligned}
&E\,\mathrm{tr}(C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1} C_{qq}^{-1} C'_{pq} C_{pq} C_{qq}^{-1}) \\
&\quad = E\,\mathrm{tr}\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{j_2=1}^{K} \eta_{j_1}\eta'_{j_1} C_{qq}^{-2} \eta_{j_2}\eta'_{j_2} C_{qq}^{-2}\right] p^2 \\
&\qquad + E\,\mathrm{tr}\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{i_1=1}^{K} \eta_{j_1}(\eta'_{i_1} C_{qq}^{-2} \eta_{j_1})\eta'_{i_1} C_{qq}^{-2}\right] p
\end{aligned}$$

$$
\begin{aligned}
&+ E\,\mathrm{tr}\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{i_1=1}^{K}\eta_{j_1}\left(\eta_{i_1}' C_{qq}^{-2}\eta_{i_1}\right)\eta_{j_1}' C_{qq}^{-2}\right]p \\
&= E\,\mathrm{tr}\left(\sum_{\ell_1=1}^{K}\eta_{\ell_1}\eta_{\ell_1}'\right)^{-2}\left(p^2+p\right) \\
&\quad + E\,\mathrm{tr}\left[\frac{1}{K^2}\sum_{j_1=1}^{K}\eta_{j_1}\eta_{j_1}' C_{qq}^{-2}\right]\mathrm{tr}\left[\frac{1}{K^2}\sum_{i_1=1}^{K}\eta_{i_1}' C_{qq}^{-2}\eta_{i_1}\right]p \\
&= E\left[\mathrm{tr}\left(\Pi^2\right)\right]\left(p^2+p\right) + E\left[\mathrm{tr}(\Pi)\right]^2 p \\
&= \left(\sum_{i=1}^{q}\sum_{j=1}^{q} E\Pi_{ij}^2\right)\left(p^2+p\right) + E\left(\sum_{i=1}^{q}\Pi_{ii}\right)^2 p.
\end{aligned}
$$

It follows from Theorem 5.2.2 of Press (2005, p. 119, using the notation here) that

$$
E\Pi_{ij}^2 = \frac{(K-q+1)\delta_{ij} + (K-q-1)}{(K-q)(K-q-1)^2(K-q-3)} + \frac{\delta_{ij}}{[K-q-1]^2} = O\left(\frac{1}{K^2}\right),
$$

and for $i \neq j$,

$$
\begin{aligned}
E\Pi_{ii}\Pi_{jj} &= \frac{2}{(K-q)(K-q-1)^2(K-q-3)} + \frac{1}{[K-q-1]^2} \\
&= O\left(\frac{1}{K^2}\right).
\end{aligned}
$$

Hence,

$$
E\,\mathrm{tr}\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right) = O\left(\frac{1}{K^2}\right). \tag{B.12}
$$

Next

$$
\begin{aligned}
&\mathrm{var}\left(\left[\mathrm{tr}\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)\right]\right) \\
&\quad \leq E\left[\mathrm{tr}\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)\mathrm{tr}\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)\right] \\
&\quad = E\,\mathrm{tr}\left[\left(\frac{1}{K}\sum_{j=1}^{K}\eta_j\xi_j'\right)\left(\frac{1}{K}\sum_{j=1}^{K}\xi_j\eta_j'\right)C_{qq}^{-2}\right] \\
&\qquad \times \mathrm{tr}\left[\left(\frac{1}{K}\sum_{j=1}^{K}\eta_j\xi_j'\right)\left(\frac{1}{K}\sum_{j=1}^{K}\xi_j\eta_j'\right)C_{qq}^{-2}\right]
\end{aligned}
$$

$$= E\,\mathrm{tr}\left[\frac{1}{K}\sum_{i_1=1}^{K}\sum_{j_1=1}^{K}\eta_{i_1}\left(\xi_{i_1}'\xi_{j_1}\right)\eta_{j_1}'C_{qq}^{-2}\right]$$

$$\times\,\mathrm{tr}\left[\frac{1}{K}\sum_{i_2=1}^{K}\sum_{j_2=1}^{K}\eta_{i_2}\left(\xi_{i_2}'\xi_{j_2}\right)\eta_{j_2}'C_{qq}^{-2}\right]$$

$$= E\frac{1}{K^2}\sum_{i_1=1}^{K}\sum_{j_1=1}^{K}\sum_{i_2=1}^{K}\sum_{j_2=1}^{K}\mathrm{tr}\left(\eta_{i_1}\eta_{j_1}'C_{qq}^{-2}\right)$$

$$\times\,\mathrm{tr}\left(\eta_{i_2}\eta_{j_2}'C_{qq}^{-2}\right)E\left[\left(\xi_{i_1}'\xi_{j_1}\right)\left(\xi_{i_2}'\xi_{j_2}\right)\right]$$

$$= E\left[\mathrm{tr}(\Pi)\right]^2p^2 + E\frac{1}{K^2}\sum_{i_1=1}^{K}\sum_{j_1=1}^{K}\mathrm{tr}\left[C_{qq}^{-2}\eta_{j_1}\eta_{j_1}'C_{qq}^{-2}\eta_{i_1}\eta_{i_1}'\right]2p$$

$$= E\left[\mathrm{tr}(\Pi)\right]^2p^2 + E\,\mathrm{tr}\left[\Pi^2\right]2p.$$

Using the same formulae from Press (2005), we can show that the last term is of $O(K^{-2})$. This, combined with (B.12), leads to $\mathrm{var}(\Delta^2) = O(K^{-2})$.

Taking a Taylor expansion and using the mean and variance of Δ^2, we have

$$\text{(B.13)}\quad P\left(\kappa^{-1}F_\infty < z\right)$$

$$= EF_{p,K-p-q+1}\left(z, \Delta^2\right)$$

$$= EF_{p,K-p-q+1}\left(z, \delta^2\right) + E\frac{\partial F_{p,K-p-q+1}(z,\delta^2)}{\partial\lambda}\left(\Delta^2-\delta^2\right)$$

$$+ E\frac{\partial^2 F_{p,K-p-q+1}(z,\tilde{\Delta}^2)}{\partial\lambda^2}\left(\Delta^2-\delta^2\right)^2$$

$$= EF_{p,K-p-q+1}\left(z, \delta^2\right) + E\frac{\partial^2 F_{p,K-p-q+1}(z,\tilde{\Delta}^2)}{\partial\lambda^2}\left(\Delta^2-\delta^2\right)^2$$

for some $\tilde{\Delta}^2$ between Δ^2 and δ^2. By definition,

$$F_{p,K-p-q+1}(z,\lambda) = P\left(\frac{\chi_p^2(\lambda)}{p}\left[\frac{\chi_{K-p-q+1}^2}{K-p-q+1}\right]^{-1} < z\right)$$

$$= E\mathcal{G}_p\left(pz\left[\frac{\chi_{K-p-q+1}^2}{K-p-q+1}\right], \lambda\right),$$

where $\mathcal{G}_p(z,\lambda)$ is the CDF of the noncentral chi-squared distribution $\chi_p^2(\lambda)$ with noncentrality parameter λ. In view of the relationship $\mathcal{G}_p(z,\lambda) =$

$\exp(-\frac{\lambda}{2}) \times \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} \mathcal{G}_{p+2j}(z)$, we have

$$
\begin{aligned}
&\frac{\partial F_{p,K-p-q+1}(z,\lambda)}{\partial \lambda} \\
&\quad = \sum_{j=0}^{\infty} \frac{\partial}{\partial \lambda}\left[\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^j}{j!}\right] E\mathcal{G}_{p+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right) \\
&\quad = -\frac{1}{2}\exp\left(-\frac{\lambda}{2}\right)\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{j!} E\mathcal{G}_{p+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right) \\
&\qquad + \frac{1}{2}\exp\left(-\frac{\lambda}{2}\right)\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{j!} E\mathcal{G}_{p+2+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right) \\
&\quad = \frac{1}{2}\sum_{j=0}^{\infty}\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^j}{j!} E\left[\mathcal{G}_{p+2+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right)\right. \\
&\qquad \left. - \mathcal{G}_{p+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right)\right]
\end{aligned}
$$

and so

$$
\begin{aligned}
&\left|\frac{\partial^2 F_{p,K-p-q+1}(z,\lambda)}{\partial \lambda^2}\right| \\
&\quad \le \frac{1}{2}\sum_{j=0}^{\infty}\left|\frac{\partial}{\partial\lambda}\left[\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^j}{j!}\right]\right| \\
&\quad \le \frac{1}{4}\sum_{j=0}^{\infty}\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^j}{j!} + \frac{1}{4}\sum_{j=1}^{\infty}\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^{(j-1)}}{(j-1)!} \\
&\quad \le \frac{1}{4} + \frac{1}{4} = \frac{1}{2}
\end{aligned}
$$

for all z and λ. Combining the boundedness of $\partial^2 F_{p,K-p-q+1}(z,\lambda)/\partial\lambda^2$ with (B.13) yields

$$
\begin{aligned}
P\left(\kappa^{-1}F_\infty < z\right) &= F_{p,K-p-q+1}\left(z,\delta^2\right) + O\left[\operatorname{var}\left(\Delta^2\right)\right] \\
&= F_{p,K-p-q+1}\left(z,\delta^2\right) + O\left(\frac{1}{K^2}\right) \\
&= F_{p,K-p-q+1}\left(z,\delta^2\right) + o\left(\frac{1}{K}\right).
\end{aligned}
$$

Part (c). It follows from part (b) that

$$
\begin{aligned}
\text{(B.14)} \quad P(pF_\infty < z) &= E\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}, \delta^2\right) + o\left(\frac{1}{K}\right) \\
&= E\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}, 0\right) + E\frac{\partial}{\partial\lambda}\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}, 0\right)\delta^2 + o\left(\frac{1}{K}\right) \\
&\quad + E\left[\frac{\partial^2}{\partial\lambda^2}\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}, \bar{\delta}^2\right)\right]\delta^4,
\end{aligned}
$$

where $\bar{\delta}^2$ is between 0 and δ^2. As in the proof of part (b), we can show that $|\frac{\partial^2}{\partial\lambda^2}\mathcal{G}_p(z, \lambda)| \leq 1$. As a result,

$$
\text{(B.15)} \quad E\left[\frac{\partial^2}{\partial\lambda^2}\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}, \bar{\delta}^2\right)\right]\delta^4 = O\left(\frac{1}{K^2}\right) = o\left(\frac{1}{K}\right).
$$

Consequently,

$$
\begin{aligned}
P(pF_\infty < z) = E\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}, 0\right) \\
+ E\frac{\partial}{\partial\lambda}\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}, 0\right)\delta^2 + o\left(\frac{1}{K}\right).
\end{aligned}
$$

By direct calculations, it is easy to show that

$$
\text{(B.16)} \quad \frac{\partial}{\partial\lambda}\mathcal{G}_p(z, 0) = -\frac{1}{2}\left[\mathcal{G}_p(z) - \mathcal{G}_{p+2}(z)\right] = -\frac{1}{p}\frac{z^{p/2}e^{-z/2}}{2^{p/2}\Gamma\left(\frac{p}{2}\right)} = -\frac{1}{p}\mathcal{G}'_p(z)z.
$$

Therefore,

$$
\begin{aligned}
P(pF_\infty < z) &= E\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}\right) \\
&\quad - \frac{1}{p}E\mathcal{G}'_p\left(\frac{z\chi^2_{K-p-q+1}}{K}\right)\frac{z\chi^2_{K-p-q+1}}{K}\delta^2 + o\left(\frac{1}{K}\right) \\
&= \mathcal{G}_p(z) + \mathcal{G}'_p(z)zE\left(\frac{\chi^2_{K-p-q+1}}{K} - 1\right)
\end{aligned}
$$

$$+\frac{1}{2}\mathcal{G}_p''(z)z^2\operatorname{var}\left(\frac{\chi^2_{K-p-q+1}}{K}\right)-\frac{\delta^2}{p}\mathcal{G}_p'(z)z+o\left(\frac{1}{K}\right)$$

$$=\mathcal{G}_p(z)+\mathcal{G}_p'(z)z\frac{-p-q+1}{K}+\mathcal{G}_p''(z)z^2\frac{K-p-q+1}{K^2}$$

$$-\frac{q}{K-q-1}\mathcal{G}_p'(z)z$$

$$=\mathcal{G}_p(z)-\mathcal{G}_p'(z)z\left(\frac{p+2q-1}{K}\right)+\mathcal{G}_p''(z)z^2\frac{1}{K}+o\left(\frac{1}{K}\right). \qquad \text{Q.E.D.}$$

PROOF OF THEOREM 4: *Part* (a). Using the same argument for proving Theorem 3(a), we have

$$t_\infty \overset{d}{=} \frac{B_1(1)-C_{1q}C_{qq}^{-1}B_q(1)}{\sqrt{\chi^2_{K-q}/K}} \tag{B.17}$$

and so

$$\frac{t_\infty}{\sqrt{\kappa}} \overset{d}{=} \frac{B_1(1)-C_{1q}C_{qq}^{-1}B_q(1)}{\sqrt{\chi^2_{K-q}/(K-q)}} \overset{d}{=} t_{K-q}(\Delta).$$

Part (b). Since the distribution of t_∞ is symmetric about 0, we have for any $z\in\mathbb{R}^+$,

$$P\left(\frac{t_\infty}{\sqrt{\kappa}}<z\right)$$

$$=\frac{1}{2}+\frac{1}{2}P\left(|t_\infty/\sqrt{\kappa}|<|z|\right)=\frac{1}{2}+\frac{1}{2}P\left(t_\infty^2/\kappa<z^2\right)$$

$$=\frac{1}{2}+\frac{1}{2}F_{1,K-q}\left(z^2,\delta^2\right)+o\left(\frac{1}{K}\right),$$

where the last equality follows from Theorem 3(b). When $z\in\mathbb{R}^-$, we have

$$P\left(\frac{t_\infty}{\sqrt{\kappa}}<z\right)$$

$$=\frac{1}{2}-\frac{1}{2}P\left(|t_\infty/\sqrt{\kappa}|<|z|\right)=\frac{1}{2}-\frac{1}{2}P\left(t_\infty^2/\kappa<z^2\right)$$

$$=\frac{1}{2}-\frac{1}{2}F_{1,K-q}\left(z^2,\delta^2\right)+o\left(\frac{1}{K}\right).$$

Therefore,

$$P\left(\frac{t_\infty}{\sqrt{\kappa}} < z\right) = \frac{1}{2} + \frac{1}{2}\operatorname{sgn}(z) F_{1,K-q}\left(z^2, \delta^2\right) + o\left(\frac{1}{K}\right).$$

Part (c). Using Theorem 3(c) and the symmetry of the distribution of t_∞ about 0, we have for any $z \in \mathbb{R}^+$,

$$\begin{aligned} P(t_\infty < z) &= \frac{1}{2} + \frac{1}{2} P\left(|t_\infty| < |z|\right) = \frac{1}{2} + \frac{1}{2} P\left(t_\infty^2 < z^2\right) \\ &= \frac{1}{2} + \frac{1}{2}\mathcal{G}\left(z^2\right) - \mathcal{G}'\left(z^2\right) z^2 \left(\frac{q}{K}\right) + \frac{1}{2}\mathcal{G}''\left(z^2\right) z^4 \frac{1}{K} + o\left(\frac{1}{K}\right). \end{aligned}$$

Using the relationships that

$$\frac{1}{2} + \frac{1}{2}\mathcal{G}\left(z^2\right) = \Phi(z)$$

and

$$-\mathcal{G}'\left(z^2\right) z^2 \left(\frac{q}{K}\right) + \frac{1}{2}\mathcal{G}''\left(z^2\right) z^4 \frac{1}{K} = -\frac{1}{4K}\boldsymbol{\phi}(z)\left[z^3 + z(4q+1)\right],$$

we have

$$P(t_\infty < z) = \Phi(z) - \frac{1}{4K} z \boldsymbol{\phi}(z)\left[z^2 + (4q+1)\right] + o\left(\frac{1}{K}\right).$$

Similarly, when $z \in \mathbb{R}^-$, we have

$$P(t_\infty < z) = \Phi(z) + \frac{1}{4K} z \boldsymbol{\phi}(z)\left[z^2 + (4q+1)\right] + o\left(\frac{1}{K}\right).$$

Therefore,

$$P(t_\infty < z) = \Phi(z) - \frac{1}{4K} |z| \boldsymbol{\phi}(z)\left[z^2 + (4q+1)\right] + o\left(\frac{1}{K}\right). \qquad \textit{Q.E.D.}$$

Before proving Theorem 5, we present a technical lemma. Part (i) of the lemma is proved in Sun (2014). Parts (ii) and (iii) of the lemma are proved in Sun, Phillips, and Jin (2011).

Define $g_0 = \lim_{x\to 0}[1 - k(x)]/x^{q_0}$, q_0 is the Parzen exponent of the kernel function, $c_1 = \int_{-\infty}^{\infty} k(x)\,dx$, and $c_2 = \int_{-\infty}^{\infty} k^2(x)\,dx$. Recall the definitions of μ_1 and μ_2 in (8).

LEMMA B.1:

(i) *For the conventional kernel HAR variance estimators*, *we have*, *as* $h \to \infty$,

(a) $\mu_1 = 1 - bc_1 + O(b^2)$,

(b) $\mu_2 = bc_2 + O(b^2)$.

(ii) *For the sharp kernel HAR variance estimator*, *we have*, *as* $h \to \infty$,

(a) $\mu_1 = 1 - \frac{2}{\rho+2}$,

(b) $\mu_2 = \frac{1}{\rho+1} + O(\frac{1}{\rho^2})$.

(iii) *For the steep kernel HAR variance estimators*, *we have*, *as* $h \to \infty$,

(a) $\mu_1 = 1 - (\frac{\pi}{\rho g_0})^{1/2} + O(\frac{1}{\rho})$,

(b) $\mu_2 = (\frac{\pi}{2\rho g_0})^{1/2} + O(\frac{1}{\rho})$.

PROOF OF THEOREM 5: *Part* (a). Recall that

$$pF_\infty = \big[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\big]' D_{pp}^{-1}\big[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\big].$$

Conditional on (C_{pq}, C_{qq}, C_{pp}), $B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)$ is normal with mean zero and variance $I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'$. Let $\mathcal{L}$ be the lower triangular matrix such that $\mathcal{L}\mathcal{L}'$ is the Choleski decomposition of $I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'$. Then the conditional distribution of $\zeta := \mathcal{L}^{-1}[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)]$ is $N(0, I_p)$. Since the conditional distribution does not depend on (C_{pq}, C_{qq}, C_{pp}), we can conclude that ζ is independent of (C_{pq}, C_{qq}, C_{pp}). So we can write

$$pF_\infty \stackrel{d}{=} \zeta' A \zeta,$$

where $A = \mathcal{L}' D_{pp}^{-1} \mathcal{L}$. Given that A is a function of (C_{pq}, C_{qq}, C_{pp}), we know that ζ and A are independent. As a result, $\zeta' A\zeta \stackrel{d}{=} \zeta'(OAO')\zeta$ for any orthonormal matrix O that is independent of ζ.

Let $H = (\zeta/\|\zeta\|, \tilde{H})$ be an orthonormal matrix with first column $\zeta/\|\zeta\|$. We choose H to be independent of A. This is possible as ζ and A are independent. Then

$$\begin{aligned} pF_\infty &\stackrel{d}{=} \|\zeta\|^2 \frac{\zeta'}{\|\zeta\|}\big(OAO'\big)\frac{\zeta}{\|\zeta\|} = \|\zeta\|^2\left[\frac{\zeta'}{\|\zeta\|}H\right]H'\big(OAO'\big)H\left[H'\frac{\zeta}{\|\zeta\|}\right] \\ &= \|\zeta\|^2 e_p'\big(H'OAO'H\big)e_p, \end{aligned}$$

where $e_p = (1, 0, \dots, 0)'$ is the first basis vector in $\mathbb{R}^p$. Since $\|\zeta\|^2$, H, and A are mutually independent from each other, we can write

$$pF_\infty \stackrel{d}{=} \|\zeta\|^2 e_p'\big(\mathcal{H}'O'AO\mathcal{H}\big)e_p$$

for any orthonormal matrix $\mathcal{H}$ that is independent of both ζ and A. Letting $O = \mathcal{H}'$, we obtain

$$pF_\infty \overset{d}{=} \|\zeta\|^2 \left(e_p' A e_p\right) = \frac{\|\zeta\|^2}{[e_p' A e_p]^{-1}} \overset{d}{=} \frac{\|\zeta\|^2}{[e_p' \mathcal{L}' D_{pp}^{-1} \mathcal{L} e_p]^{-1}}.$$

Since $\mathcal{L}e_p$ is the first column of $\mathcal{L}$, we have, using the definition of the Choleski decomposition,

$$\mathcal{L}e_p = \frac{[I_p + C_{pq} C_{qq}^{-1} C_{qq}^{-1} C_{pq}'] e_p}{\sqrt{e_p' [I_p + C_{pq} C_{qq}^{-1} C_{qq}^{-1} C_{pq}'] e_p}}.$$

As a result,

$$pF_\infty \overset{d}{=} \frac{\|\zeta\|^2}{\eta^2},$$

where

$$\begin{aligned}\eta^2 &= \left[e_p' \mathcal{L}' D_{pp}^{-1} \mathcal{L} e_p\right]^{-1} \\ &= \frac{e_p' [I_p + C_{pq} C_{qq}^{-1} C_{qq}^{-1} C_{pq}'] e_p}{e_p' [I_p + C_{pq} C_{qq}^{-1} C_{qq}^{-1} C_{pq}'] D_{pp}^{-1} [I_p + C_{pq} C_{qq}^{-1} C_{qq}^{-1} C_{pq}'] e_p}.\end{aligned}$$

Part (b). It is easy to show that

$$EC_{qq} = \mu_1 I_q \quad \text{and} \quad \operatorname{var}\left[\operatorname{vec}(C_{qq})\right] = \mu_2 (I_{qq} + \mathbb{K}_{qq}),$$

where μ_1 and μ_2 are defined in (8), I_{qq} is the $q^2 \times q^2$ identity matrix, and $\mathbb{K}_{qq}$ is the $q^2 \times q^2$ commutation matrix. So $C_{qq} = \mu_1 I_q + o_p(1)$ and $C_{qq}^{-1} = \mu_1^{-1} I_q + o_p(1)$ as $h \to \infty$. Similarly, $C_{pp} = \mu_1 I_p + o_p(1)$ and $C_{pp}^{-1} = \mu_1^{-1} I_p + o_p(1)$ as $h \to \infty$. In addition, using the same argument, we can show that $C_{pq} = o_p(1)$. Therefore,

$$\begin{aligned}\eta^2 &= \frac{e_p' [I_p + C_{pq} C_{pq}' / \mu_1^2] e_p}{e_p' [I_p + C_{pq} C_{pq}' / \mu_1^2] [I_p - C_{pq} C_{pq}' / \mu_1^2]^{-1} [I_p + C_{pq} C_{pq}' / \mu_1^2] e_p} \\ &\quad \times \left(1 + o_p(1)\right) \\ &= 1 + o_p(1).\end{aligned}$$

That is, $\eta^2 \to^p 1$ as $h \to \infty$.

We proceed to prove the distributional expansion. The (i, j)th elements $C_{pp}(i, j)$, $C_{pq}(i, j)$, and $C_{qq}(i, j)$ of C_{pp}, C_{pq}, and C_{qq} are equal to either

$\int_0^1 \int_0^1 Q_h^*(r,s)\,dB(r)\,dB(s)$ or $\int_0^1 \int_0^1 Q_h^*(r,s)\,dB(r)\,d\tilde{B}(s)$, where $B(\cdot)$ and $\tilde{B}(\cdot)$ are independent standard Brownian motion processes. By direct calculations, we have, for any $\varsigma \in (0, 3/8)$,

$$
\begin{aligned}
&P\bigl(\bigl|C_{ef}(i,j) - EC_{ef}(i,j)\bigr| > \mu_2^{\varsigma}\bigr) \\
&\quad \le \frac{E|C_{ef}(i,j) - EC_{ef}(i,j)|^8}{\mu_2^{8\varsigma}} = O\left(\frac{\mu_2^4}{\mu_2^{8\varsigma}}\right) = o(\mu_2),
\end{aligned}
$$

where $e, f = p$ or q. Define the event $\mathcal{E}$ as

$$
\mathcal{E} = \bigl\{\omega : \bigl|C_{ef}(i,j) - EC_{ef}(i,j)\bigr| \le \mu_2^{\varsigma} \text{ for all } i, j \text{ and all } e \text{ and } f\bigr\}.
$$

Then the complement $\mathcal{E}^c$ of $\mathcal{E}$ satisfies $P(\mathcal{E}^c) = o(\mu_2)$ as $h \to \infty$. Let $\tilde{\mathcal{E}}$ be another event. Then

$$
\begin{aligned}
P(\tilde{\mathcal{E}}) &= P(\tilde{\mathcal{E}} \cap \mathcal{E}) + P\bigl(\tilde{\mathcal{E}} \cap \mathcal{E}^c\bigr) = P(\tilde{\mathcal{E}} \cap \mathcal{E}) + o(\mu_2) \\
&= P(\tilde{\mathcal{E}}|\mathcal{E})P(\mathcal{E}) + o(\mu_2) = P(\tilde{\mathcal{E}}|\mathcal{E})\bigl(1 - o(\mu_2)\bigr) + o(\mu_2) \\
&= P(\tilde{\mathcal{E}}|\mathcal{E}) + o(\mu_2).
\end{aligned}
$$

That is, up to an error of $o(\mu_2)$, $P(\tilde{\mathcal{E}})$, $P(\tilde{\mathcal{E}} \cap \mathcal{E})$, and $P(\tilde{\mathcal{E}}|\mathcal{E})$ are asymptotically equivalent. So for the purpose of proving the theorem, it is innocuous to condition on $\mathcal{E}$ or to remove the conditioning, if needed.

Now conditioning $\mathcal{E}$, the numerator of η^2 satisfies

$$
\begin{aligned}
&e_p'\bigl[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\bigr]e_p \\
&\quad = 1 + \frac{1}{\mu_1^2} e_p' C_{pq}\left[I_q + \frac{C_{qq} - EC_{qq}}{\mu_1}\right]^{-1}\left[I_q + \frac{C_{qq} - EC_{qq}}{\mu_1}\right]^{-1} C_{pq}' e_p \\
&\quad = 1 + \frac{1}{\mu_1^2} e_p' C_{pq}(I_q - M_{qq})(I_q - M_{qq})C_{pq}' e_p \\
&\quad = 1 + \frac{1}{\mu_1^2} e_p' C_{pq}\tilde{M}_{qq}C_{pq}' e_p,
\end{aligned}
$$

where M_{qq} is a matrix with elements $M_{qq}(i,j)$ satisfying $|M_{qq}(i,j)| = O(\mu_2^{\varsigma})$ conditional on $\mathcal{E}$, and $\tilde{M}_{qq}$ is a matrix satisfying $|\tilde{M}_{qq}(i,j) - 1\{i = j\}| = O(\mu_2^{\varsigma})$ conditional on $\mathcal{E}$.

Let

$$
C_{p+q,p+q} = \begin{pmatrix} C_{pp} & C_{pq} \\ C_{pq}' & C_{qq} \end{pmatrix}
$$

and let $e_{q+q,p}$ be the matrix consisting of the first p columns of the identity matrix I_{p+q}. To evaluate the denominator of η^2, we note that

$$\begin{aligned}
D_{pp}^{-1} &= \frac{1}{\mu_1} e'_{q+q,p}\left(I_{p+q} + \frac{C_{p+q,p+q} - EC_{p+q,p+q}}{\mu_1}\right)^{-1} e_{q+q,p} \\
&= \frac{1}{\mu_1} e'_{q+q,p}\left(I_{p+q} - \frac{C_{p+q,p+q} - EC_{p+q,p+q}}{\mu_1}\right) e_{q+q,p} \\
&\quad + e'_{q+q,p}\left(\frac{[C_{p+q,p+q} - EC_{p+q,p+q}][C_{p+q,p+q} - EC_{p+q,p+q}]}{\mu_1^2}\right) \\
&\quad \times e_{q+q,p} + M_{pp} \\
&= \frac{1}{\mu_1}\left(I_p - \frac{C_{pp} - EC_{pp}}{\mu_1}\right. \\
&\quad \left. + \frac{[C_{pp} - EC_{pp}][C_{pp} - EC_{pp}] + C_{pq}C'_{pq}}{\mu_1^2}\right) + M_{pp},
\end{aligned}$$

where M_{pp} is a matrix with elements $M_{pp}(i, j)$ satisfying $|M_{pp}(i, j)| = O(\mu_2^{3\varsigma}) = o(\mu_2)$ for $\varsigma > 1/3$ conditional on $\mathcal{E}$.

For the purpose of proving our result, M_{pp} can be ignored as its presence generates an approximation error of $o(\mu_2)$, which is the same as the order of the approximation error given in the theorem. More specifically, let

$$\tilde{C}_{pp} = \frac{C_{pq}\tilde{M}_{qq}C'_{pq}}{\mu_1^2},$$

$$\tilde{D}_{pp}^{-} = \left(I_p - \frac{C_{pp} - EC_{pp}}{\mu_1} + \frac{[C_{pp} - EC_{pp}][C_{pp} - EC_{pp}] + C_{pq}C'_{pq}}{\mu_1^2}\right)$$

and

$$\tilde{\eta}^2 := \tilde{\eta}^2\big(\tilde{C}_{pp}, \tilde{D}_{pp}^{-}\big) = \frac{\mu_1(1 + e'_p\tilde{C}_{pp}e_p)}{e'_p(I_p + \tilde{C}_{pp})\tilde{D}_{pp}^{-}(I_p + \tilde{C}_{pp})e_p}.$$

Then

$$P(pF_\infty < z) = E\mathcal{G}_p\big(\eta^2 z\big) = E\mathcal{G}_p\big(\tilde{\eta}^2 z\big) + o(\mu_2).$$

Note that for any $q \times q$ matrix L_{qq}, we have

$$\begin{aligned}
\text{(B.18)} \quad EC_{pq}L_{qq}C'_{pq} = E\int_0^1\int_0^1 & Q_h^*(r_1, s_1)Q_h^*(r_2, s_2)\, dB_p(r_1)\, dB_q(s_1)' \\
& \times L_{qq}\, dB_q(s_2)\, dB_p(r_1)'
\end{aligned}$$

$$= E\int_0^1\int_0^1 Q_h^*(r_1,s_1)Q_h^*(r_2,s_2)\,dB_p(r_1)\,dB_p(r_1)'$$
$$\times \operatorname{tr}\bigl(dB_q(s_2)\,dB_q(s_1)'L_{qq}\bigr)$$
$$= \operatorname{tr}(L_{qq})\int_0^1\int_0^1 \bigl[Q_h^*(r,s)\bigr]^2\,dr\,drI_p = \operatorname{tr}(L_{qq})\mu_2 I_p.$$

Taking an expansion of $\tilde{\eta}^2(\tilde{C}_{pp},\tilde{D}_{pp}^-)$ around $\tilde{C}_{pp} = q\mu_2/\mu_1^2 I_p$ and $\tilde{D}_{pp}^- = I_p$, we obtain

$$\tilde{\eta}^2 = \eta_0^2 + \mathrm{err},$$

where err is the approximation error and

$$\eta_0^2 = \mu_1 - \frac{2q\mu_2}{\mu_1} + e_p'[C_{pp} - EC_{pp}]e_p$$
$$- \frac{e_p'[C_{pp} - EC_{pp}][C_{pp} - EC_{pp}]e_p}{\mu_1} + \frac{\{e_p'[C_{pp} - EC_{pp}]e_p\}^2}{\mu_1}.$$

We keep enough terms in η_0^2 so that $E\mathcal{G}_p(\tilde{\eta}^2 z) = E\mathcal{G}_p(\eta_0^2 z) + o(\mu_2)$.

Now we write

$$P(pF_\infty < z) = E\mathcal{G}_p\bigl(\eta_0^2 z\bigr) + o(\mu_2)$$
$$= \mathcal{G}_p(z) + \mathcal{G}_p'(z)z\bigl(E\eta_0^2 - 1\bigr)$$
$$+ \frac{1}{2}\mathcal{G}_p''(z)z^2 E\bigl(\eta_0^2 - 1\bigr)^2 + o(\mu_2).$$

In view of

$$E\eta_0^2 - 1 = (\mu_1 - 1) - \frac{2q\mu_2}{\mu_1} - \frac{1}{\mu_1}Ee_p'[C_{pp} - EC_{pp}][C_{pp} - EC_{pp}]e_p$$
$$+ E\frac{\{e_p'[C_{pp} - EC_{pp}]e_p\}^2}{\mu_1}$$
$$= (\mu_1 - 1) - \frac{2q\mu_2}{\mu_1} - \frac{\mu_2}{\mu_1}(p+1) + \frac{2\mu_2}{\mu_1}$$
$$= (\mu_1 - 1) - \frac{2q\mu_2}{\mu_1} - \frac{\mu_2}{\mu_1}(p-1)$$

and

$$E\bigl(\eta_0^2 - 1\bigr)^2$$
$$= E\biggl[\mu_1 - 1 - \frac{2q\mu_2}{\mu_1} + e_p'[C_{pp} - EC_{pp}]e_p\biggr]^2 + o(\mu_2)$$

$$= E\{e_p'[C_{pp} - EC_{pp}]e_p\}^2 + (\mu_1 - 1)^2 + o(\mu_2)$$
$$= 2\mu_2 + (\mu_1 - 1)^2 + o(\mu_2),$$

we have

$$\text{(B.19)} \quad P(pF_\infty < z)$$
$$= \mathcal{G}_p(z) + \mathcal{G}_p'(z)z\left[(\mu_1 - 1) - \frac{2q\mu_2}{\mu_1} - \frac{\mu_2}{\mu_1}(p-1)\right]$$
$$+ \mathcal{G}_p''(z)z^2\mu_2 + o(\mu_2)$$
$$= \mathcal{G}_p(z) + \mathcal{G}_p'(z)z\left[(\mu_1 - 1) - \frac{\mu_2}{\mu_1}(p + 2q - 1)\right]$$
$$+ \mathcal{G}_p''(z)z^2\mu_2 + o(\mu_2). \qquad \textit{Q.E.D.}$$

PROOF OF THEOREM 6: We prove part (a) only, as the proof for part (b) is similar. Using the same arguments and the notation as in the proof of Theorem 1, we have

$$R\big[G'W_\infty^{-1}G\big]^{-1}G'W_\infty^{-1}\Lambda B_m(1) + \delta_0$$
$$\stackrel{d}{=} RVA^{-1}\left(I_d, \quad -C_{12}C_{22}^{-1}\right)B_m(1) + \delta_0$$

and

$$R\big[G'W_\infty^{-1}G\big]^{-1}R' \stackrel{d}{=} RVA^{-1}\big(C^{11}\big)^{-1}\big(A'\big)^{-1}V'R'.$$

So

$$F_{\infty,\delta_0} \stackrel{d}{=} \left[RVA^{-1}\left(I_d, \quad -C_{12}C_{22}^{-1}\right)B_m(1) + \delta_0\right]'$$
$$\times \left[RVA^{-1}\big(C^{11}\big)^{-1}\big(A'\big)^{-1}V'R'\right]^{-1}$$
$$\times \left[RVA^{-1}\left(I_d, \quad -C_{12}C_{22}^{-1}\right)B_m(1) + \delta_0\right]/p.$$

Let $B_m(1) = [B_d'(1), B_q'(1)]'$ and $RVA^{-1} = \tilde{U}\tilde{\Sigma}\tilde{V}'$ be a SVD of RVA^{-1}, where $\tilde{\Sigma} = (\tilde{A}, \quad \tilde{O})$. Then

$$F_{\infty,\delta_0} \stackrel{d}{=} \left\{\tilde{U}\tilde{\Sigma}\tilde{V}'\big[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\big] + \delta_0\right\}'\left(\tilde{U}\tilde{\Sigma}\tilde{V}'\big(C^{11}\big)^{-1}\tilde{V}\tilde{\Sigma}'\tilde{U}'\right)^{-1}$$
$$\times \left\{\tilde{U}\tilde{\Sigma}\tilde{V}'\big[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\big] + \delta_0\right\}/p$$
$$= \left\{\tilde{\Sigma}\tilde{V}'\big[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\big] + \tilde{U}'\delta_0\right\}'\left(\tilde{\Sigma}\tilde{V}'C_{11}^{-1}\tilde{V}\tilde{\Sigma}'\right)^{-1}$$
$$\times \left\{\tilde{\Sigma}\tilde{V}'\big[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\big] + \tilde{U}'\delta_0\right\}/p.$$

The above distribution does not depend on the orthonormal matrix $\tilde{V}$. So

$$\begin{aligned}F_{\infty,\delta_0} &\overset{d}{=} \{\tilde{\Sigma}[B_d(1) - C_{12}C_{22}^{-1}B_q(1)] + \tilde{U}'\delta_0\}'(\tilde{\Sigma}C_{11}^{-1}\tilde{\Sigma}')^{-1} \\ &\quad \times \{\tilde{\Sigma}[B_d(1) - C_{12}C_{22}^{-1}B_q(1)] + \tilde{U}'\delta_0\}/p \\ &= \{\tilde{A}[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)] + \tilde{U}'\delta_0\}'(\tilde{A}D_{pp}\tilde{A})^{-1} \\ &\quad \times \{\tilde{A}[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)] + \tilde{U}'\delta_0\} \\ &= [B_p(1) - C_{pq}C_{qq}^{-1}B_q(1) + \tilde{A}^{-1}\tilde{U}'\delta_0]' \\ &\quad \times D_{pp}^{-1}[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1) + \tilde{A}^{-1}\tilde{U}'\delta_0].\end{aligned}$$

Let $H = (\frac{\tilde{A}^{-1}U'\delta_0}{\|\tilde{A}^{-1}U'\delta_0\|}, \tilde{H})$ be a $p \times p$ orthonormal matrix. Then

$$\begin{aligned}F_{\infty,\delta_0} &\overset{d}{=} [H'B_p(1) - H'C_{pq}C_{qq}^{-1}B_q(1) + H'\tilde{A}^{-1}\tilde{U}'\delta_0]' \\ &\quad \times H'D_{pp}^{-1}H[H'B_p(1) - H'C_{pq}C_{qq}^{-1}B_q(1) + H'\tilde{A}^{-1}\tilde{U}'\delta_0] \\ &= [H'B_p(1) - H'C_{pq}C_{qq}^{-1}B_q(1) + e_p\|\tilde{A}^{-1}\tilde{U}'\delta_0\|]' \\ &\quad \times H'D_{pp}^{-1}H[H'B_p(1) - H'C_{pq}C_{qq}^{-1}B_q(1) + e_p\|\tilde{A}^{-1}\tilde{U}'\delta_0\|].\end{aligned}$$

But the joint distribution of $(H'B_p(1), H'C_{pq}, H'D_{pp}^{-1}H)$ is invariant to H. Hence, we can write

$$\begin{aligned}F_{\infty,\delta_0} &\overset{d}{=} [B_p(1) - C_{pq}C_{qq}^{-1}B_q(1) + e_p\|\tilde{A}^{-1}\tilde{U}'\delta_0\|]' \\ &\quad \times D_{pp}^{-1}[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1) + e_p\|\tilde{A}^{-1}\tilde{U}'\delta_0\|]/p.\end{aligned}$$

That is, the distribution of F_{∞,δ_0} depends on δ_0 only through $\|\tilde{A}^{-1}\tilde{U}'\delta_0\|$. Note that $\tilde{U}\tilde{A}\tilde{A}'\tilde{U}' = \tilde{U}\tilde{\Sigma}\tilde{V}'(\tilde{U}\tilde{\Sigma}\tilde{V}')' = RVA^{-1}(RVA^{-1})'$ and

$$\begin{aligned}\|\tilde{A}^{-1}\tilde{U}'\delta_0\|^2 &= \delta_0'\tilde{U}(\tilde{A}^{-1})'\tilde{A}^{-1}\tilde{U}'\delta_0 = \delta_0'[\tilde{U}\tilde{A}\tilde{A}'\tilde{U}']^{-1}\delta_0 \\ &= \delta_0'[RVA^{-1}(A^{-1})'V'R']^{-1}\delta_0 \\ &= \delta_0'[R(VA'AV')^{-1}R']^{-1}\delta_0 \\ &= \delta_0'\{R[(\Lambda^{-1}G)'\Lambda^{-1}G]^{-1}R'\}^{-1}\delta_0 \\ &= \delta_0'\{R[G'\Omega^{-1}G]^{-1}R'\}^{-1}\delta_0 \\ &= \|\mathcal{V}^{-1/2}\delta_0\|^2 = \|\check{\delta}\|^2,\end{aligned}$$

so we have

$$F_{\infty,\delta_0} \overset{d}{=} \big[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1) + \check{\delta}\big]' \times D_{pp}^{-1}\big[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1) + \check{\delta}\big]/p = F_\infty\big(\|\check{\delta}\|^2\big).$$

Q.E.D.

APPENDIX C: PRACTICAL GUIDANCE AND EMPIRICAL APPLICATION

C.1. *Practical Guidance*

C.1.1. *GMM Estimation and Optimal Weighting Matrix Estimation*

Let $v_t \in \mathbb{R}^{d_v}$ be a vector of observations at time t for $t = 1, \dots, T$. In the GMM framework, we specify a vector of moment conditions,

$$Ef(v_t, \theta_0) = 0, \quad t = 1, 2, \dots, T, \tag{C.1}$$

where $\theta_0 \in \mathbb{R}^d$ is the unknown parameter vector of interest and $f(v_t, \cdot)$ is an $m \times 1$ vector of continuously differentiable functions with $m \geq d$. To achieve identification, we assume that on the parameter space Θ, $Ef(v_t, \theta) = 0$ if and only if $\theta = \theta_0$. The model is possibly overidentified with the degree of overidentification $q = m - d$.

Let

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} f(v_t, \theta).$$

Then the GMM estimator of θ_0 is given by

$$\hat{\theta}_{\text{GMM}} = \arg\min_{\theta\in\Theta} g_T(\theta)' W_T^{-1} g_T(\theta),$$

where W_T is a positive definite weighting matrix.

To obtain an initial first step estimator, we choose a simple weighting matrix W_o that does not depend on model parameters, leading to

$$\tilde{\theta}_T = \arg\min_{\theta\in\Theta} g_T(\theta)' W_o^{-1} g_T(\theta).$$

W_o may depend on the sample size T, in which case the dependence has been suppressed. As an example, we may set $W_o = I_m$ in the general GMM setting. In the IV regression, we may set $W_o = Z'Z/T$, where Z is the data matrix for the instruments.

According to Hansen (1982), the optimal weighting matrix is the long run variance (LRV) matrix of the moment process $\{f(v_t, \theta_0)\}$. Since θ_0 is not known, the moment process $\{f(v_t,\theta_0)\}$ is not directly observable but it can be estimated by $\{\tilde{u}_t := f(v_t,\tilde{\theta}_T)\}$. On the basis of $\{\tilde{u}_t\}$, we estimate the optimal weighting matrix by

$$W_T(\tilde{\theta}_T;h)=\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_h\left(\frac{t}{T},\frac{s}{T}\right)\left(\tilde{u}_t-\frac{1}{T}\sum_{\tau=1}^{T}\tilde{u}_\tau\right)\left(\tilde{u}_s-\frac{1}{T}\sum_{\tau=1}^{T}\tilde{u}_\tau\right)', \tag{C.2}$$

where $Q_h(r,s)$ is a symmetric weighting function that depends on the smoothing parameter h. There are several choices of Q_h in the literature. We list two here.

- Conventional kernel estimators: $Q_h(r,s)=k((r-s)/b)$ with $b=1/h$, where $k(\cdot)$ is the kernel function.
- Orthonormal series (OS) estimators: $Q_h(r,s)=K^{-1}\sum_{j=1}^{K}\phi_j(r)\phi_j(s)$ with $K=h$, where $\{\phi_j(r)\}$ are orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1\phi_j(r)\,dr=0$.

In the case of kernel estimation, we can take $k(\cdot)$ to be the Bartlett kernel, the Parzen kernel, or the QS kernel, the three commonly used positive definite kernels. In this case, bT is usually referred to as the truncation lag or bandwidth. Andrews (1991) shows that the AMSE-optimal smoothing parameter h is[1]

$$\text{Bartlett kernel: } h_T^*=0.8736\big[\alpha(q_0)\big]^{-1/(2q_0+1)}T^{2q_0/(2q_0+1)}\quad\text{for}\quad q_0=1,$$

$$\text{Parzen kernel: } h_T^*=0.3757\big[\alpha(q_0)\big]^{-1/(2q_0+1)}T^{2q_0/(2q_0+1)}\quad\text{for}\quad q_0=2,$$

$$\text{QS kernel: } h_T^*=0.7564\big[\alpha(q_0)\big]^{-1/(2q_0+1)}T^{2q_0/(2q_0+1)}\quad\text{for}\quad q_0=2,$$

where

$$\alpha(q_0)=\frac{2\operatorname{vec}(B_{q_0})'\operatorname{vec}(B_{q_0})}{\operatorname{tr}[(I_{m^2}+\mathbb{K}_{mm})(\Omega\otimes\Omega)]}$$

and

$$B_{q_0}=\sum_{j=-\infty}^{\infty}|j|^{q_0}Ef(v_t,\theta_0)f(v_{t-j},\theta_0)'\quad\text{and}$$

$$\Omega=\sum_{j=-\infty}^{\infty}Ef(v_t,\theta_0)f(v_{t-j},\theta_0)'.$$

In the above expression, I_{m^2} is the $m^2\times m^2$ identity matrix and $\mathbb{K}_{mm}$ is the $m^2\times m^2$ commutation matrix.

[1] Andrews (1991) gives the AMSE-optimal smoothing parameter in terms of the "truncation lag" S_T. In our notation, $h_T^*=T/S_T^*$.

In the case of OS estimation, we can assume that h is even and take $\phi_j(\cdot)$ to be $\phi_{2j-1}(x) = \sqrt{2}\cos 2j\pi x$, $\phi_{2j}(x) = \sqrt{2}\sin 2j\pi x$, $j = 1, \ldots, h/2$. It follows from Phillips (2005) that the AMSE-optimal and even h is given by

$$h_T^* = 2 \times \left\lceil 0.3567\left[\alpha(q_0)\right]^{-1/(2q_0+1)} T^{2q_0/(2q_0+1)} \right\rceil \quad \text{for} \quad q_0 = 2,$$

where $\lceil\cdot\rceil$ is the ceiling function.

To obtain a data-driven h^*, we employ the standard plug-in procedure. The VAR(1) plug-in procedure employs a VAR(1) as the parametric model to approximate the dynamics in $\{f(v_t, \theta_0)\}$. It involves the following steps:

(i) Fit a VAR(1) by ordinary least squares to the estimated process $\{\tilde{u}_t := f(v_t, \tilde{\theta}_T)\}$ to obtain an estimate $\tilde{A}$ of the autoregressive matrix and an estimate $\tilde{\Sigma}$ of the error variance matrix.

(ii) Use the formulae on page 835 of Andrews (1991) to obtain estimates $\tilde{B}_{q_0}$ and $\tilde{\Omega}$.[2]

(iii) Plug $\tilde{B}_{q_0}$ and $\tilde{\Omega}$ into the definition of h_T^* to obtain the data-driven $\tilde{h}_T$.

In the plug-in procedure, we can also use other approximating parametric models such as univariate AR(1) models, higher order AR models, and moving average models (univariate or multivariate). For more details, see Andrews (1991).

Our inference procedure below treats $\tilde{h}_T$ as if it is deterministic. To minimize the undue influence from the randomness of $\tilde{h}_T$, we could discretize the set of possible scaling constants in h_T^*, replacing $\tilde{h}_T$ with the closest value, $\tilde{h}_T^\dagger$, in some finite set. The estimation uncertainty in $\tilde{h}_T^\dagger$ is small enough that it will not affect our asymptotic results. This remedy is more of theoretical interest. To maintain the automated nature of our testing procedures, we recommend skipping the remedy and treating $\tilde{h}_T$ as a deterministic sequence in practice.

C.1.2. *Two-Step Test Statistics*

With the variance estimator $W_T(\tilde{\theta}_T; \tilde{h}_T)$, the two-step GMM estimator is

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} g_T(\theta)' W_T(\tilde{\theta}_T; \tilde{h}_T) g_T(\theta).$$

Our two-step inference is based on the above two-step estimator $\hat{\theta}_T$.

Suppose we want to test the linear null hypothesis $H_0 : R\theta_0 = r$ against $H_0 : R\theta_0 \neq r$, where R is a $p \times d$ matrix with full row rank.[3] We consider three

[2]In terms of the notation here, $\tilde{B}_{q_0} = 2\pi \hat{f}^{(q_0)}$ and $\tilde{\Omega} = 2\pi \hat{f}$, where $\hat{f}^{(q_0)}$ and $\hat{f}$ are given on page 835 of Andrews (1991) with $\hat{A}$ replaced by $\tilde{A}$ and $\hat{\Sigma}$ replaced by $\tilde{\Sigma}$.

[3]For the Wald and t types of tests, nonlinear restrictions can be transformed into linear ones using the delta method.

types of test statistics. The first type is the conventional Wald statistic. The (normalized) Wald statistic is

$$\begin{aligned}\mathbb{W}_T &:= \mathbb{W}_T(\hat{\theta}_T) \\ &= T(R\hat{\theta}_T - r)'\big\{R\big[G_T(\hat{\theta}_T)'W_T^{-1}(\tilde{\theta}_T; \tilde{h}_T)G_T(\hat{\theta}_T)\big]^{-1}R'\big\}^{-1} \\ &\quad \times (R\hat{\theta}_T - r)/p,\end{aligned}$$

where $G_T(\theta) = \frac{\partial g_T(\theta)}{\partial \theta'}$. When $p = 1$ and for one-sided alternative hypotheses, we can construct the t statistic

$$t_T := t_T(\hat{\theta}_T) = \frac{\sqrt{T}(R\hat{\theta}_T - r)}{\{R[G_T(\hat{\theta}_T)'W_T^{-1}(\tilde{\theta}_T; \tilde{h}_T)G_T(\hat{\theta}_T)]^{-1}R'\}^{1/2}}.$$

The second type of test statistic is based on the likelihood ratio principle. Let $\hat{\theta}_{T,R}$ be the restricted second-step GMM estimator:[4]

$$\hat{\theta}_{T,R} = \arg\min_{\theta \in \Theta} g_T(\theta)'W_T^{-1}(\tilde{\theta}_T; \tilde{h}_T)g_T(\theta) \quad \text{s.t.} \quad R\theta = r.$$

The likelihood ratio principle suggests the GMM distance statistic (or GMM criterion function statistic) given by

$$\begin{aligned}\mathbb{D}_T = \big[&Tg_T(\hat{\theta}_T)'W_T^{-1}(\tilde{\theta}_T; \tilde{h}_T)g_T(\hat{\theta}_T) \\ &- Tg_T(\hat{\theta}_{T,R})'W_T^{-1}(\tilde{\theta}_T; \tilde{h}_T)g_T(\hat{\theta}_{T,R})\big]/p.\end{aligned}$$

The third type of test statistic is the GMM counterpart of the score statistic or Lagrange multiplier statistic. It is based on the score or gradient of the GMM criterion function, that is, $\Delta_T(\theta) = G_T'(\theta)W_T^{-1}(\tilde{\theta}_T, \tilde{h}_T)g_T(\theta)$. The test statistic is given by

$$\mathbb{S}_T = T\big[\Delta_T(\hat{\theta}_{T,R})\big]'\big[G_T'(\hat{\theta}_{T,R})W_T^{-1}(\tilde{\theta}_T, \tilde{h}_T)G_T(\hat{\theta}_{T,R})\big]^{-1}\Delta_T(\hat{\theta}_{T,R})/p.$$

C.1.3. *Approximating Distributions*

It is shown that $\mathbb{W}_T$, $\mathbb{D}_T$, and $\mathbb{S}_T$ have the same limiting distribution. To describe the approximating distributions for the test statistics $\mathbb{W}_T$, $\mathbb{D}_T$, $\mathbb{S}_T$, and t_T, we let $e_t := (e'_{t,p}, e'_{t,d-p}, e'_{t,q})' \sim$ i.i.d. $N(0, I_m)$. The subscripts p, $d-p$, and q on e indicate not only the dimensions of the random vectors, but also distin-

[4]The constraints imposed are the restrictions under the null hypothesis. Linearization is not necessary if the null restrictions are nonlinear.

guish them so that, for example, $e_{t,p}$ is different and independent from $e_{t,q}$ for all values of p and q. Denote

$$C_{p,T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{t,p}, \quad C_{q,T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{t,q}$$

and

$$\text{(C.3)} \quad \begin{aligned} C_{pp,T} &= \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} Q_{\tilde{h}_T}\left(\frac{t}{T}, \frac{\tau}{T}\right) \tilde{e}_{t,p} \tilde{e}'_{\tau,p}, \\ C_{pq,T} &= \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} Q_{\tilde{h}_T}\left(\frac{t}{T}, \frac{\tau}{T}\right) \tilde{e}_{t,p} \tilde{e}'_{\tau,q}, \\ C_{qq,T} &= \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} Q_{\tilde{h}_T}\left(\frac{t}{T}, \frac{\tau}{T}\right) \tilde{e}_{t,q} \tilde{e}'_{\tau,q}, \\ D_{pp,T} &= C_{pp,T} - C_{pq,T} C_{qq,T}^{-1} C'_{pq,T}, \end{aligned}$$

where $\tilde{e}_{i,j} = e_{i,j} - \frac{1}{T} \sum_{s=1}^{T} e_{s,j}$ for $i = t, \tau$ and $j = p, q$.

Define

$$\begin{aligned} F_{eT} &:= \left[C_{p,T} - C_{pq,T} C_{qq,T}^{-1} C_{q,T}\right]' D_{pp,T}^{-1} \left[C_{p,T} - C_{pq,T} C_{qq,T}^{-1} C_{q,T}\right] / p; \\ t_{eT} &:= \left[C_{p,T} - C_{pq,T} C_{qq,T}^{-1} C_{q,T}\right] / \sqrt{D_{pp,T}} \quad \text{for} \quad p = 1. \end{aligned}$$

Then the distributions of $\mathbb{W}_T$, $\mathbb{D}_T$, and $\mathbb{S}_T$ can be approximated by that of F_{eT} and the distribution of t_T can be approximated by that of t_{eT}. The distributions of F_{eT} and t_{eT} can be easily simulated, as they involve only T i.i.d. standard normal vectors.

In the special case of OS variance estimation, we can approximate the distributions of $\mathbb{W}_T$, $\mathbb{D}_T$, and $\mathbb{S}_T$ by a noncentral F distribution. More specifically, let

$$\tilde{\delta}_T^2 = \frac{pq}{\tilde{h}_T - q - 1}$$

and let $\mathcal{F}^{1-\alpha}_{p,\tilde{h}_T - p - q + 1}(\tilde{\delta}_T^2)$ be the $(1-\alpha)$ quantile of the noncentral F distribution $F_{p,\tilde{h}_T - p - q + 1}(\tilde{\delta}_T^2)$ with degrees of freedom $(p, \tilde{h}_T - p - q + 1)$ and noncentrality parameter $\tilde{\delta}_T^2$. Then we can use

$$\text{(C.4)} \quad \frac{\tilde{h}_T - p - q + 1}{\tilde{h}_T} \mathcal{F}^{1-\alpha}_{p,\tilde{h}_T - p - q + 1}\left(\tilde{\delta}_T^2\right)$$

as the α-level critical value for the tests based on the statistics $\mathbb{W}_T$, $\mathbb{D}_T$, and $\mathbb{S}_T$.

Similarly, in the case of OS variance estimation, the $(1-\alpha)$ quantile of the finite sample distribution of t_T can be approximated by

$$\text{(C.5)}\qquad \begin{aligned} &\sqrt{\frac{(\tilde{h}_T - q + 1)}{\tilde{h}_T}\mathcal{F}^{1-2\alpha}_{1,\tilde{h}_T - q}\big(\tilde{\delta}_T^2\big)}, \quad \text{if} \quad \alpha < 0.5, \\ &-\sqrt{\frac{(\tilde{h}_T - q + 1)}{\tilde{h}_T}\mathcal{F}^{2\alpha-1}_{1,\tilde{h}_T - q}\big(\tilde{\delta}_T^2\big)}, \quad \text{if} \quad \alpha \geq 0.5. \end{aligned}$$

For a two-sided t test, the $(1-\alpha)$ quantile of $|t_T|$ can be approximated by

$$\text{(C.6)}\qquad \sqrt{\frac{(\tilde{h}_T - q + 1)}{\tilde{h}_T}\mathcal{F}^{1-\alpha}_{1,K-q}\big(\tilde{\delta}_T^2\big)}.$$

In essence, we construct the test statistics in the usual way. The difference between the newly proposed testing procedure and existing ones lies in the critical value used. In the case of kernel HAR variance estimation, we simulate F_{eT} and t_{eT} and use their quantiles as the critical values. In the case of OS HAR variance estimation, we can either use the simulated critical values from F_{eT}/t_{eT} or use noncentral F critical values given in (C.4)–(C.6).

C.2. *Empirical Application*

To illustrate the fixed-smoothing approximations, we consider the log-normal stochastic volatility model of the form

$$\begin{aligned} &r_t = \sigma_t Z_t, \\ &\log \sigma_t^2 = \omega + \beta\big(\log \sigma_{t-1}^2 - \omega\big) + \sigma_u u_t, \end{aligned}$$

where r_t is the rate of return and (Z_t, u_t) is i.i.d. $N(0, I_2)$. The first equation specifies the distribution of the return as heteroscedastic normal. The second equation specifies the dynamics of the log volatility as an AR(1). The parameter vector is $\theta = (\omega, \beta, \sigma_u)$. We impose the restriction that $\beta \in (0, 1)$, which is an empirically relevant range. The model and the parameter restriction are the same as those considered by Andersen and Sorensen (1996), which gives a detailed discussion on the motivation of the stochastic volatility models and the GMM approach. For more discussions, see Ghysels, Harvey, and Renault (1996) and references therein.

We employ the GMM to estimate the log-normal stochastic volatility model. The data are weekly returns of Standard & Poor's 500 stocks, which are constructed by compounding daily returns with dividends from a Center for Research in Security Prices index file. We consider both value-weighted returns (vwretd) and equal-weighted returns (ewretd). The weekly returns range from the first week of 2001 to the last week of 2012 with sample size $T = 627$. We use weekly data so as to minimize problems associated with daily data such as asynchronous trading and bid–ask bounce. This is consistent with Jacquier, Polson, and Rossi (1994).

The GMM approach relies on functions of the time series $\{r_t\}$ to identify the parameters of the model. For the log-normal stochastic volatility model, we can obtain the moment conditions

$$\begin{aligned} &E|r_t|^\ell = \mathfrak{c}_\ell E(\sigma_t^\ell) \quad \text{for} \quad \ell = 1, 2, 3, 4 \\ &\quad \text{with} \quad (\mathfrak{c}_1, \mathfrak{c}_2, \mathfrak{c}_3, \mathfrak{c}_4) = (\sqrt{2/\pi}, 1, 2\sqrt{2/\pi}, 3), \\ &E|r_t r_{t-j}| = 2\pi^{-1} E(\sigma_t \sigma_{t-j}) \quad \text{and} \\ &E r_t^2 r_{t-j}^2 = E(\sigma_t^2 \sigma_{t-j}^2) \quad \text{for} \quad j = 1, 2, \ldots, \end{aligned}$$

where

$$\begin{aligned} &E(\sigma_t^\ell) = \exp\left[\frac{\omega\ell}{2} + \frac{\ell^2\sigma_u^2}{8(1-\beta^2)}\right], \\ &E(\sigma_t^{\ell_1}\sigma_{t-j}^{\ell_2}) = E(\sigma_t^{\ell_1}) E(\sigma_t^{\ell_2}) \exp\left[\frac{\sigma_u^2 \ell_1 \ell_2 \beta^j}{4(1-\beta^2)}\right]. \end{aligned}$$

Higher order moments can be computed, but we choose to focus on a subset of lower order moments. Andersen and Sorensen (1996) point out that it is generally not optimal to include too many moment conditions when the sample is limited. On the other hand, it is not advisable to include just as many moment conditions as the number of parameters. When $T = 500$ and $\theta = (-7.36, 0.90, 0.363)$, which is an empirically relevant parameter vector, Table 1 in Andersen and Sorensen (1996) shows that it is MSE-optimal to employ nine moment conditions. For this reason, we employ two sets of nine moment conditions given in the Appendix of Andersen and Sorensen (1996). The baseline set of the nine moment conditions are $Ef_i(r_t, \theta) = 0$ for $i = 1, \ldots, 9$ with

$$\begin{aligned} &f_i(r_t, \theta) = |r_t|^i - \mathfrak{c}_\ell E(\sigma_t^\ell), \quad i = 1, \ldots, 4, \\ &f_5(r_t, \theta) = |r_t r_{t-1}| - 2\pi^{-1} E(\sigma_t \sigma_{t-1}), \\ &f_6(r_t, \theta) = |r_t r_{t-3}| - 2\pi^{-1} E(\sigma_t \sigma_{t-3}), \\ &f_7(r_t, \theta) = |r_t r_{t-5}| - 2\pi^{-1} E(\sigma_t \sigma_{t-5}), \\ &f_8(r_t, \theta) = r_t^2 r_{t-2}^2 - E(\sigma_t^2 \sigma_{t-2}^2), \\ &f_9(r_t, \theta) = r_t^2 r_{t-4}^2 - E(\sigma_t^2 \sigma_{t-4}^2). \end{aligned} \tag{C.7}$$

The alternative set of the nine moment conditions are $Ef_i(r_t, \theta) = 0$ for $i = 1, \dots, 9$ with

$$f_i(r_t, \theta) = |r_t|^i - \mathfrak{c}_\ell E(\sigma_t^\ell), \quad i = 1, \dots, 4,$$
$$f_5(r_t, \theta) = |r_t r_{t-2}| - 2\pi^{-1} E(\sigma_t \sigma_{t-2}),$$
$$f_6(r_t, \theta) = |r_t r_{t-4}| - 2\pi^{-1} E(\sigma_t \sigma_{t-4}),$$
$$f_7(r_t, \theta) = |r_t r_{t-6}| - 2\pi^{-1} E(\sigma_t \sigma_{t-6}),$$
$$f_8(r_t, \theta) = r_t^2 r_{t-1}^2 - E(\sigma_t^2 \sigma_{t-1}^2),$$
$$f_9(r_t, \theta) = r_t^2 r_{t-3}^2 - E(\sigma_t^2 \sigma_{t-3}^2).$$

In each case $m = 9$ and $d = 3$, and so the degree of overidentification is $q = m - d = 6$.

We focus on constructing 90% and 95% confidence intervals (CIs) for β. Given the high nonlinearity of the moment conditions, we invert the GMM distance statistic to obtain the CIs. The CIs so obtained are invariant to the reparametrization of model parameters. For example, a 95% confidence interval is the set of β values in $(0, 1)$ that the $\mathbb{D}_T$ test does not reject at the 5% level. We search over the grid from 0.01 to 0.99 with increments of 0.01 to invert the $\mathbb{D}_T$ test. As in the simulation study, we employ three different critical values: $\chi_p^{1-\alpha}/p$, $\mathcal{F}_\infty^{1-\alpha}[0]$, and $\mathcal{F}_\infty^{1-\alpha}[q]$, which correspond to three different asymptotic approximations. For the series LRV estimator, $\mathcal{F}_\infty^{1-\alpha}[0]$ is a critical value from the corresponding F approximation. Before presenting the empirical results, we notice that there is no "hole" in the CIs we obtain—the values of β that are not rejected form an arithmetic sequence. Given this, we report only the minimum and maximum values of β over the grid that are not rejected. It turns out that the maximum value is always equal to the upper bound 0.99. It thus suffices to report the minimum value, which we take as the lower limit of the CI.

Table C.I presents the lower limits for different 95% CIs together with the smoothing parameters used. The smoothing parameters are selected in the same ways as in the simulation study. Since the selected K values are relatively large and the selected b values are relatively small, the CIs based on the χ^2 approximation and the $F_\infty[0]$ approximation are close to each other. However, there is still a noticeable difference between the CIs based on the χ^2 approximation and the nonstandard $F_\infty[q]$ approximation. This is especially true for the case with baseline moment conditions, the Bartlett kernel, and equally-weighted returns. Taking into account the randomness of the estimated weighting matrix leads to wider CIs. The log volatility may not be as persistent as previously thought.

TABLE C.I

LOWER LIMITS OF DIFFERENT 95% CIS FOR β AND DATA-DRIVEN SMOOTHING PARAMETERS IN THE LOG-NORMAL STOCHASTIC VOLATILITY MODEL[a]

	Equally-Weighted Return		Value-Weighted Return	
	Baseline	Alternative	Baseline	Alternative
Series χ^2	0.74	0.58	0.78	0.62
Series $F_\infty[0]$	0.73	0.57	0.78	0.62
Series $F_\infty[q]$	0.70	0.53	0.76	0.57
K	(140)	(138)	(140)	(140)
Bartlett χ^2	0.56	0.52	0.78	0.60
Bartlett $F_\infty[0]$	0.54	0.52	0.77	0.60
Bartlett $F_\infty[q]$	0.35	0.42	0.75	0.53
b	(0.0155)	(0.0157)	(0.155)	(0.0155)
Parzen χ^2	0.72	0.52	0.78	0.50
Parzen $F_\infty[0]$	0.73	0.52	0.78	0.50
Parzen $F_\infty[q]$	0.69	0.47	0.77	0.43
b	(0.0136)	(0.0139)	(0.0136)	(0.0136)
QS χ^2	0.74	0.52	0.78	0.52
QS $F_\infty[0]$	0.74	0.52	0.78	0.52
QS $F_\infty[q]$	0.71	0.48	0.77	0.46
b	(0.0068)	(0.0069)	(0.0067)	(0.0067)

[a] Baseline: CI using the baseline nine moment conditions given in (C.7). Alternative: CI using the alternative nine moment conditions given in (C.8). Series χ^2: the CI constructed based on the OS HAR variance estimator and using χ_1^2 as the reference distribution. Other row titles are similarly defined.

REFERENCES

ANDERSEN, T. G., AND B. E. SORENSEN (1996): "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business & Economic Statistics*, 14 (3), 328–352. [25, 26]

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [21,22]

GHYSELS, E., A. C. HARVEY, AND E. RENAULT (1996): "Stochastic Volatility," in *Statistical Methods in Finance*, ed. by G. S. Maddala and C. R. Rao. Amsterdam: North-Holland, 119–191. [25]

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [20]

JACQUIER, E., N. G. POLSON, AND P. E. ROSSI (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 12 (4), 371–389. [26]

PHILLIPS, P. C. B. (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142. [22]

PRESS, J. (2005): *Applied Multivariate Analysis* (Second Ed.). New York: Dover. [7,8]

SUN, Y. (2014): "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference," *Journal of Econometrics*, 178 (3), 659–677. [12]

SUN, Y., P. C. B. PHILLIPS, AND S. JIN (2011): "Power Maximization and Size Control in Heteroscedasticity and Autocorrelation Robust Tests With Exponentiated Kernels," *Econometric Theory*, 27 (6), 1320–1368. [12]

Dept. of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, U.S.A.; yisun@ucsd.edu.

Manuscript received June, 2013; final revision received July, 2014.